ORNL/TM-2020/1756

Oak Ridge National Laboratory

Joe Lake
Program Manager

9/1/2020

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE US DEPARTMENT OF ENERGY

ORNL/TM-2020/1756

National Institutes of Health, National Center for Advancing Translational Sciences Clinical and Translational Science Award

COOPERATIVE RESEARCH AND DEVELOPMENT AGREEMENT WITH GEORGETOWN UNIVERSITY REPORT

Joseph Lake

Date Published:
9/1/2020

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee 37831-6283
managed by
UT-BATTELLE, LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

Table of Contents

1. ABSTRACT

Oak Ridge National Laboratory (ORNL) is participating with Georgetown University (GU) as a subrecipient in response to the National Institutes of Health (NIH), National Center for Advancing Translational Sciences (NCATS) Clinical and Translational Science Award (CTSA) (U54 Clinical Trial Optional) funding opportunity announcement. This Cooperative Research and Development Agreement is put in to place to facilitate the development and implementation of clinical interventions that demonstrably improve human health is currently a complex, recursive, and inefficient process that leads to delays of years or decades before discoveries in biomedical research result in health benefits for patients and communities. NCATS conducts and supports research in the science of translation, to discover the mechanistic and operational principles of the intervention development and dissemination process, thereby providing the scientific foundation for improvements in translational efficiency that will accelerate the realization of interventions that improve human health. Under NCATS' leadership, the CTSA Program supports a national network of medical research institutions called hubs. GU is the lead institution in one of the NIH hubs that was created as a result of a previous NIH CTSA.

The missions of the GU have historically included the advancement of health through research in the clinical and biomedical sciences, the education of future leaders in medical and nursing practice and academia, and the provision of compassionate and scientifically competent patient care and service to the Washington, DC community and the nation. GU is the lead institution for the Georgetown-Howard Universities Center for Clinical and Translational Science (GHUCCTS), a multi-institutional partnership of medical research institutions forged from a desire to promote clinical research and translational science. Through multiple collaborations among these institutions, GHUCCTS is transforming clinical research and translational science in order to bring new scientific advances to health care.

Oak Ridge National Laboratory is the Department of Energy's (DOE) largest science and energy laboratory. Managed since April 2000 by a partnership of the University of Tennessee and Battelle, ORNL was established in 1943 as a part of the secret Manhattan Project to pioneer a method for producing and separating plutonium.

During the 1950s and 1960s, ORNL became an international center for the study of nuclear energy and related research in the physical and life sciences. With the creation of DOE in the 1970s, ORNL's mission broadened to include a variety of energy technologies and strategies. Today the laboratory supports the nation with a peacetime science and technology mission that is just as important as, but very different from, its role during the Manhattan Project.

ORNL is home to the world's premier center for high performance supercomputing to enable scientific discovery. ORNL has extensive expertise in various areas of computer science that are uniquely situated to support GU. Additionally, ORNL's leading computational user facilities present a unique opportunity to leverage the largest scale machines for open science in support of the stated mission of the NCATS CTSA. ORNL's partnership with GU will offer unparalleled opportunity in data analytics, deep-learning, artificial intelligence, and urban dynamics.

2. STATEMENT OF OBJECTIVES

2.1 PURPOSE AND BACKGROUND

The purpose of this **Multi-Task** Cooperative Research and Development Agreement (CRADA) between UT-Battelle, LLC (the "Contractor") and Georgetown University (the "Participant") is to establish a collaboration that will develop insights into the systems genetics underlying individual susceptibilities to diseases, such as cancer, metabolic and neurological disorders. The Contractor will bring to bear the power of neutron scattering and other technologies to develop better early markers of disease. The Contractor and the Participant will be jointly referred to as the "Parties."

The Parties will dedicate their **combined** resources to improving discovery translation and, thus, drug discovery and development. Drug discovery and development are key parts of the life sciences pipeline and a systems-base approach to medicine. By elucidating the underlying genetically encoded vulnerabilities and environmental factors that conspire to cause disease, researchers and physicians can identify molecules that underlie the risk and utilize knowledge of their functions for drug discovery.

The **respective** leadership of the Parties has agreed that this shared ambition should culminate with the execution of this **Multi-Task** type CRADA covering the Parties collaborative research and development activities and making possible the seamless exchange of ideas and information associated with the individual tasks assigned.

2.2 SCOPE OF WORK

The scope of work for this CRADA was collaborative research in the following thematic areas:

1. Structural Biology
2. Biomarkers
3. Computational Biology
4. Systems Genetics
5. **Drug Discovery**
6. **Low Dose Radiation**

The research under this Multi-Task CRADA shall be directed under the general following general task areas and, throughout the duration of this CRADA, the work statement will be modified to match the needs of the Parties and the direction of the research.

2.3 GENERAL TASKS

2.3.1 General Task 1 – Structural Biology:

Multi-domain protein structures provide a deep insight into function and underlying mechanisms. Though the number of available protein structures is growing rapidly, solving large protein structures is still challenging. If the structures of the component domains are known, systematic docking or multimeric threading may be tied to predicting protein-protein interaction sites. However, such approaches require enormous computational capacity for a genome-wide application. This is more so for proteins that interact in multiple orientations on homologous protein surfaces. Thus, cataloguing all the known interfaces for protein-protein

interactions may provide an alternative basis for uncovering connections and inter-connection between signal transduction pathways. The high powered computational techniques of the Contractor would be an enormous resource for these research activities of the Parties.

2.3.2 **General Task 2 - Biomarkers:**

Generally in the area of therapeutics it is a goal to detect onset of a disease at the earliest stage. Such detections are especially important in cancer because of rapid progression and the metastatic nature of the disease. Therefore, having a reliable biomarker detection tool is highly desirable. To this end, development of a chip-based biomarker detection tool can provide basis for collaboration between the Participant's clinicians and the Contractor's researchers.

2.3.3 General Task 3 – Computational Biology:

One approach to understanding the complex processes encoded by the human genome is to assign its proteins/enzymes to biochemical pathways that regulate biochemical transformations. Pathways and interaction between pathways place genes in their larger biological context, and enable causal inferences about the likely effects of mutations, drug interventions and changes in gene regulation. Thus pathway mapping is the first step toward quantitative modeling of metabolism and systems interaction. Such assignment of genes to pathways also permits a validation of the human genome annotation.

With that understanding, it would important to assign and predict genes and gene products through signal transduction pathways. Pathway prediction is a significant process for uncovering interconnectivities at systems level, which then can be employed for prophylactic and therapeutic-based disease interventions. This is a fertile area for collaborations between the Parties.

2.3.4 General Task 4 – Systems Genetics:

The main uses of the systems genetics approach in quantitative genomics are: refinement of the identified Quantitative Trait Locus (QTL), candidate gene and sialophorin (SNP) discovery, understanding gene-environment and gene-gene interactions, detection of candidate regulator genes/expressive Quantitative Trait Locus (eQTL), discriminating multiple QTL/eQTL, and detection of pleiotropic QTL/eQTL, in addition to its use in reconstructing regulatory networks. The potential uses in animal breeding are direct selection on heritable gene expression measures, termed "expression assisted selection," and genetical genomic selection of both QTL and eQTL based on breeding values of the respective genes, termed "expression-assisted evaluation."

The Collaborative Cross program at the Contractor's facility is a revolutionary program which can enable a vast array of the Participant's research initiatives to expand their utilization of systems genetics to improve the understanding of complex disease and its interactions within the mammalian system.

2.3.5 General Task 5 – Drug Discovery:

The potential of having high end crystallography and compound refinement with the crystal structure is of major value and would provide the needed front end contribution for rapid drug discovery with multiple applications, including in the areas of B-cell lymphoma and breast cancer, two key areas of interest.

Since breast and prostate are two major focuses in cancer research by the Participant, it would be ideal to start modeling protein-protein targets for crystallization. Modeling that intersect would provide a platform for new breast and prostate cancer therapeutics. The Parties could then expand to other prominent cancers such as lymphoma.

2.3.6 General Task 6 – Low Dose Radiation:

While representatives of the Participant and the Contractor agree that low dose **radiation** had little future interest, intermediate radiation dose levels, with clinical relevance were more interesting. The Contractor's strength lies in the availability of novel mouse models. **Contractor's** Dr. Voy has developed genetically defined mosaic mice for investigations of radiation effects on immunity. The Participant has data from Dr. Mira Jung that the Contractor's proprietary Histone Deacetylase (HDAC) inhibitors protect mice from bone marrow toxicity after exposure of mice to lethal doses of radiation. The Participant **may wish** to collaborate by providing Dr. Voy with this first generation drug, from its portfolio for testing for radiation protection in Dr. Voy's mouse model. Subsequently, the Participant may be able to work on second generation drugs by bringing Milton Brown into the project.

3. TECHNICAL DISCUSSION OF WORK PERFORMED

3.1 PROJECT: HIGH PERFORMANCE COMPUTING FOR RATIONAL DRUG DISCOVERY AND DESIGN

Structure-based, rational drug design is a key component of modern pharmaceutical research. Knowledge of the 3-dimensional structure of a protein allows the virtual (computer-based) docking of pharmaceuticals in the protein's structure to identify those compounds that are likely to bind strongly to the target, and hence likely to be molecular effectors exhibiting protein activation or inhibition properties. This saves a considerable amount of experimental effort and cost since the "best fit" compounds are prioritized first for synthesis and experimental validation. The amount of candidate compounds that can be screened against a given protein target depends on the available computer power, size of the library and the efficiency of the computer programs.

ORNL's unique High-Performance Computing (HPC) capabilities allow such a discovery strategy to be done on an unprecedented scale. The strength of the HPC capabilities, the world's most powerful computer, has been leveraged by the development of a parallel implementation of well-validated docking algorithms by the researchers. ORNL's computational technology allows the docking of up to 10 million of compounds a day on Jaguar in a fast-primary screening mode (where chemicals are treated as rigid), and to up to 200,000 compounds/day in a more accurate docking mode (where chemicals are treated as flexible). It is proposed to use these powerful computing tools to dock the Participant in-house unique virtual compounds library (called Hoyas Library) against three-dimensional structures of the Histone deacetylases (HDAC)-4, HDAC-6 and HDAC-8 that are involved in prostate cancer and radiation sensitization. The Hoyas Library consists of about 80 million compounds (both commercial & novel virtual compounds) and is compiled from more than 300 chemical companies and literature and web resources. Virtual docking on such a scale has never been performed so far. The large Hoyas Library will greatly expand the number and the uniqueness of potential ligands against these enzyme targets, and the ORNL HPC capabilities will provide new chemical entities that have never been considered before in drug discovery.

The chemicals predicted to exhibit the best "predicted" binding properties will be communicated to the Participant for experimental validation. Compounds that are validated as HDAC inhibitors will be evaluated for effects on human cancer cell growth.

Docking of ~80 million compounds against the protein structure snapshots: The Contractor will apply a multiple stage docking strategy that funnels down the number of compounds through more and more accurate (but slower) docking protocols set upon mutual consultation between participant and the contractor. This includes docking, data management, organization and communication. Participant member can participate in the high-performance docking in the Jaguar.

3.2 ELUCIDATING THE STRUCTURE OF DISORDERED C-TERMINAL B-CATENIN BY INTEGRATING NEUTRON SCATTERING, ATOMISTIC SIMULATIONS AND STATISTICAL INFERENCE

Intrinsically disordered proteins (IDPs) represent a novel class of proteins that challenge the traditional protein structure-function paradigm; they lack stable tertiary structure under physiological conditions but undergo synergistic folding with substrate-specific conformations when bound. IDP dysfunction is implicated in several diseases including cancers, neurodegenerative and cardiovascular disorders, and diabetes. IDP binding and their aggregation mechanisms can help design self-assembly of supra-

molecular structures. Understanding the biophysical mechanisms of how IDP conformational diversity influences its function is one of the forefront problems in structural biology with impact in health, environmental and energy biosciences.

The disordered C-terminal domain (CTD) of β-catenin (residues 664-781 following the ARM repeat domain) represents one such IDP, which functions as transcriptional co-activator. As a key regulator of the want signaling pathway, β-catenin is activated in many human cancers, including hepatocellular carcinoma and colorectal cancer. β-catenin is a 781-residue protein, with a N-terminal regulatory domain, followed by a central region of 12 armadillo repeats (ARM) and a structurally flexible CTD. The CTD of β-catenin is known to be a transcription co-activator and its transcriptional activity is largely responsible for β-catenin's oncogenic effect. Although the structure and function of several of β-catenin ARM interactions have been characterized, similar structural information is not available for the CTD. However, no structural information for this IDP exists and it is currently unknown (1) whether it interacts with the ARM domain and (2) whether it interacts with other binding partners of β-catenin. Absence of the CTD results in β-catenin being transcriptionally inactive and truncation of residues from 723-781 results in a reduction of β-catenin activity by nearly 80%. Given that β-catenin's transcriptional activity is largely responsible for its oncogenic effects, understanding the structural details of whether and how the C-terminal domain folds is important to design protein mutations and/or small molecule inhibitors that can target various types of cancers, including colorectal and hepatocellular cancer. Drs. Stephen Byers and Salim Shah at Georgetown University Medical Center and Lombardi Cancer Center have been interested in understanding the structure and interaction of the CTD of β-catenin.

In this project, we have investigated the structural flexibility of the intrinsically disordered CTD of β-catenin protein. Since the CTD is highly flexible and is intrinsically disordered, its full conformational range is not observable by any one structure determination technique. Therefore, we envision developing integrated experimental and computational techniques to elicit high-resolution structural details of intrinsically disordered protein ensembles. Specifically, we will (1) develop methods to prepare selectively deuterated C-terminal β-catenin and utilize SANS contrast variation techniques to refine high-resolution conformational ensembles; (2) build parallel ensemble simulation strategies on heterogeneous computer architectures to generate millisecond timescale atomistic simulations; and (3) design Bayesian inference techniques for statistical characterization of IDP conformational ensembles. The proposed work will improve our understanding of the structure and function of the β-catenin CTD, especially in its ability to activate transcription and guide the design of novel small-molecules to inhibit β-catenin function in oncogenesis. The integrated framework proposed here would have broad applications for SANS analysis of flexible and order-disorder phenomena in polymer and materials science research.

This project has focused extensively on developing the methods and tools needed to investigate intrinsically disordered proteins (IDPs) such as β-Catenin. In particular, this activity has resulted in the development of novel experimental and computational techniques that probe disordered regions within proteins. Since the C-terminal domain of β-Catenin consists of over 150 amino-acid residues and is disordered, no one particular structure determination technique can provide unique insights into the functionality of this region. Therefore, to probe the disordered C-terminal domain, the investigators developed an integrated workflow that iteratively probed the conformational flexibility of this disordered region using neutron scattering experiments and long time-scale computational simulations. Activities have specifically resulted in the following deliverables:

- Development of new algorithms to integrate low dimensional/resolution experimental observables with computational simulations. This resulted in the publication of two papers that validated the algorithms on smaller IDP systems.

- Novel computational capabilities to simulate long time-scale conformational dynamics of IDPs and in particular, β-Catenin. Further, it also developed novel computational approaches to carry out efficient molecular docking simulations accelerated on graphics processing units (GPUs).
- A novel neutron scattering and computational approach to probe IDP ensembles and function.

The team has recently been successful in putting together the required simulations of full-length human β-Catenin and its complex with Transcription factor 4 (TCF4) in full-solvent conditions. Based on some of the accelerated simulations carried out, they have identified a list of amino acid residues that are important for enabling the C-terminal end of beta catenin to fold on to its arm domain. Mutants of these constructs are being generated at GU and will be tested at the ORNL Spallation Neutron Source facility. A workshop on disordered proteins has been organized in Telluride (Jul 2017) with planned participation by >20 eminent scientists in this area. Two papers (one accepted, and one in preparation) and a full R01 submission are expected to result from this collaborative team science project. This project has also provided the preliminary data for an exciting partnership between the Department of Energy and the National Cancers Institute of NIH involving three pilot projects centered on cancer. Specifically, the algorithms developed in this project will be used for novel neutron scattering experiments to probe the structural and conformational dynamics of the oncogene RAS in the context of its activation process.

3.3 PROJECT: DEVELOPING COMPUTER-AIDED SUPPORT ALGORITHMS TO MITIGATE THE DISRUPTIVENESS OF INTERRUPTION IN RADIOLOGY

The overarching goal of this project is to leverage eye tracking to develop a process model of how radiologists handle interruptions during the search process and to use this model to develop algorithms that can facilitate resumption and reduce the likelihood of perceptual errors. Novel methods will be developed to mitigate the disruptiveness of interruptions and thus increase patient safety in radiology.

Ten years ago the Agency for Healthcare Research and Quality (AHRQ) published a report on healthcare working conditions that highlighted the patient safety risks associated with task interruptions[1]. Interruptions are particularly detrimental in radiology; radiologists must examine numerous images and make complex interpretations that require focused attention resulting in significant cognitive demands.[2,3] In radiology, interruptions not only increase the time required to examine an image, which reduces efficiency, interruptions may also increase the likelihood of perceptual or diagnostic errors resulting in serious threats to patient safety.

Interruptions during the search process have been shown to eradicate search memory, so it follows that the interruption of radiologists while they are interpreting images and searching for abnormalities may be particularly risky.[4] To understand the cognitive processes underlying visual search researchers have examined radiologist's eye movements. Eye tracking technology provides detailed information on where radiologists are looking and provides insight into the cognition of the radiologist. While eye tracking methods have been used to understand how radiologists search images, little is known about how interruptions influence the search process.

The overarching goal of this proposal is to leverage eye tracking to develop a process model of how radiologists handle interruptions during the search process and to use this model to develop algorithms that can facilitate resumption and reduce the likelihood of perceptual errors. In a joint effort between GU and ORNL, experts in cognitive psychology, medical imaging, and computer science have formed a collaboration to achieve these goals.

While little is known about how radiologists resume an interrupted search and the types of errors that might result from interruptions, research scientists at the GU have extensive experience studying interruptions and examining both the memory and perception processes during the cycle of search,

interruptions, and resumption. In addition to developing models of these processes the GU team has also worked to develop mechanisms to facilitate resumption which have reduced time costs and lowered error rates. The BSEC team has extensive experience working with radiologists to develop cognitive decision support mechanisms and have developed several different computer-based support algorithms to facilitate detection. We will leverage the previous work of the GU and ORNL to create a synergistic collaboration. Novel methods will be developed to mitigate the disruptiveness of interruptions and thus increase patient safety in radiology.

At the conclusion of this study, we will have developed a process model of how radiologists handle interruptions during the process of interpreting imaging studies and will have developed preliminary support algorithms to facilitate task resumption. This initial collaborative effort will serve to establish a close relationship between GU and ORNL. The collaborative work will provide the foundation for larger follow-on proposals to further develop and refine computer aided support mechanisms and to test the effectiveness of these mechanisms in a real-world setting.

A common task for radiologists is to search images and recognize clinically relevant areas of interest.[13] Because visual search is a core component of what radiologists do on a daily basis, numerous researchers have examined the cognitive processes underlying search and the types of errors that might occur. Borrowing from a long history of visual search studies in cognitive psychology, researchers have identified two pathways for visual search in radiology: a selective and a non-selective pathway. The selective pathway is the process by which individual targets are examined and attention is directed to those targets for object recognition. The non-selective pathway relies on more global information and does not require specific object selection.[13–15]

During the visual search and recognition process several different types of diagnostic errors have been described.[16] *Search errors* occur when the radiologist fails to fixate on the abnormality. *Recognition errors* occur when the radiologist briefly fixates on the abnormality, but fails to recognize that the fixated item is actually abnormal. *Decision making errors* occur when radiologists fixate on the abnormality for an extended period of time but fail to indicate that the abnormality exists. Both search errors and recognition errors are often called perceptual errors.[17] With a greater understanding of the visual search and recognition processes as well as the types of errors that are committed during inspection of the image, several different clinical decision support mechanisms have been developed to facilitate radiologist's performance. However, the clinical decision support mechanisms are generally not contextually relevant and little work has leveraged the power of eye tracking technology to understand when support is needed.

While numerous researchers have focused on the visual search processes of radiologists none of these studies have examined how radiologists manage interruptions and the effects interruptions can have on visual search and recognition. The Agency for Healthcare Research and Quality (AHRQ) has highlighted the potential patient safety risks associated with task interruptions and several studies examining the workflow of radiologists have confirmed that interruptions are frequent and are a serious hazard.[18,19] For example, in a study examining radiologists interpreting body CT scans, 26% of interpretations were interrupted by referring clinicians and 39% of interpretations were interrupted by radiology related activities such as phone calls to the reading room or scan reviews.[3] Given the nature of radiology and the pressing need for unpredictable emergency cases to be examined in a timely manner it is difficult to eliminate interruptions from the workflow of radiologists. Therefore, it is critical to mitigate the effects of these interruptions, and this is the fundamental goal of our proposal.

Interruptions not only increase the stress and frustration levels of the radiologist, they also increase the time it takes to interpret the image. Some studies suggest a two-fold increase in image interpretation time when interrupted. Importantly, studies examining the effect of interruptions on performance in medicine, as well as other domains, have shown that interruptions significantly increase the likelihood of error and can be as catastrophic as increasing error rates by 10 fold. Cognitive psychology studies that have specifically examined the effect of interruptions on visual search tasks have shown that interruptions

longer than one minute can eradicate search memory. The average length of interruptions in radiology is approximately two minutes. Thus, the interruption of radiologists during the visual search and recognition process may pose a very serious threat to patient safety by increasing the likelihood of perceptual errors. When radiologists suspend their current search task, address the interruption, and subsequently attempt to resume the interrupted search task, they are likely to be more prone to perceptual errors; critical areas of the image may be neglected resulting in the missed identification of abnormalities. Yet, many studies demonstrate the persistently high rate of interruptions for radiologists. Thus, new technologies are necessary to mitigate the effects of these interruptions.

The goal of the proposed research is to leverage eye tracking technology to examine how radiologists resume an interrupted search, to understand the types of errors that may occur following interruption, and to develop preliminary algorithms that can facilitate the resumption process. This collaborative effort between Georgetown University and Oak Ridge National Laboratories Biomedical Sciences and Engineering Center (BSEC) will leverage the strengths of our multidisciplinary team, with expertise in cognitive psychology, medical imaging, and computer science, and our previous research, to develop methods to facilitate radiologist performance. This research effort will lead to advanced eye movement algorithms, which can give insight into radiologists' cognitive state, to provide contextually relevant information to mitigate the disruptiveness of interruptions. Using real-time eye tracking in radiology is a completely novel methodology and has the potential to significantly advance the performance of radiologists.

What the team found in the analysis of the results:

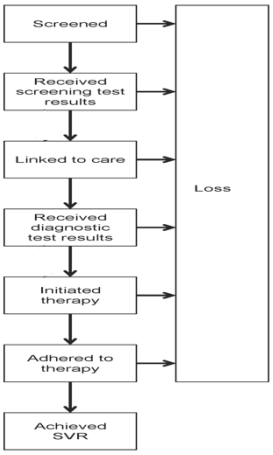
- There was a significant effect of interruptions on visual behavior of the residents group as captured using average reading time ($p=0.01$), number of fixations ($p=0.01$), duration of fixation ($p<0.01$), and inter-fixation distance ($p<0.01$);
- Analysis of Variance (ANOVA) results show no higher order effects between interruptions and other factors except individual differences in average duration of fixation ($p=0.03$).

The team at this time is currently determining whether this model can be utilized to develop algorithms that can facilitate resumption and reduce the likelihood of perceptual errors while reading imaging studies.

3.4 PROJECT: INTEGRATED DATA, ANALYTICS, MODELING AND SIMULATION (IDAMS): ELUCIDATING THE PATHWAY TO ELIMINATE HEPATITIS C VIRAL INFECTION.

Recent Center for Disease Control (CDC) studies show that incident viral hepatitis C (HCV) infections are rising significantly in adults under 30, and newly diagnosed prevalent cases are abundant in 'baby boomers' (i.e. born 1945-1965 CDC and United States Preventative Services Task Force (USPSTF) recommendations). Worryingly, given the long-term prognosis of HCV, more than 50% of those infected are not even aware of their infection. The "silent epidemic" is no longer "silent", it is now just epidemic. Outreach to the HCV community is difficult because of social and behavioral barriers including societal stigma. The HCV Cascade of Care (CoC) is complex because of significant behavioral, institutional, cost of therapy and process dependencies, as well as the presumption that HCV will remain a chronic disease. Yet in the translational research field HCV treatment is undergoing a paradigm shift and new tolerable and efficacious medications are available. Both shorter and HCV-pan-genotypic regimens are under investigation and have received breakthrough therapy designation with the Food and Drug Administration (FDA); *elimination and eradication* of HCV, nationally and globally, is for the first time a real possibility. HCV may become an episodic infection, rather than a chronic one, and episodic infections are realistic targets for elimination. And yet the hurdles enmeshed in the HCV CoC are a major barrier to elimination. Despite treatment progress, a cure for HCV as it currently stands is inefficient and complex. Implementation and creation of new best practices with critical supporting policy changes are essential.

We propose an innovative tripartite data analytics, modeling, and simulation approach to *Elucidating the Pathway to Eliminate Hepatitis C Viral Infection.*



Figue 1. Cascade of Care for HCV

The rapid advent and refinement of direct acting antiviral (DAA) therapies has made a cure (or individual sustained virologic response—SVR) for chronic hepatitis C (HEP-C) viral infection (HCV) a realistic outcome for more than 90% of HCV patients. It is now realistic to think of HCV as a target for elimination in existing chronic, and undiagnosed populations. Recently, two distinct DAA, interferon-free HCV treatment regimens were approved for genotype 1 patients: ledipasvir, sofosbuvir (Harvoni) and ombitasvir, paritaprevir/ritonavir, dasabuvir (Viekira). Both have efficacy rates greater than 90% SVR, are tablets, very tolerable, and taken for anywhere from 8 – 24 weeks. However, these therapies are expensive; a 12-week treatment of ledipasvir/sofosbuvir is retail-priced at $94,500. Yet despite this cost these treatments are cost effective in the long term when one considers cost offsets by reductions in future medical costs from complications such as cirrhosis or liver cancer [4] and increased workforce productivity.

The challenge. Despite the high efficacy, successful therapy completion rates remain low from time of screening [5,6,7]: fewer than 10% are treated and cured. The culprit is the typical losses throughout the CoC for HCV (Figure 1). The CoC is 12 to 18 months long and represents a complex and fragmented set of medical delivery, therapeutic cost, administrative, behavioral, social, and community interfaces and processes; this complexity leads to constant subject attrition. Although incremental improvement in the CoC may be possible, e.g. improved case management and patient navigation, the level of improvement needed for HCV population elimination is insurmountable using current approaches. We believe that using *in silico* IDAMS modeling to guide principled refinements of the CoC and test interventions is highly innovative (the innovation even increases when the insilico and of clinical investigation are coupled simultaneously in the future). The Titan supercomputer at Oak Ridge National Laboratory - the Contractor's location - allows us to perform vastly complex iterations for every stage of the CoC. We will model all outcomes against Georgetown-Howard Universities Center for Clinical and Translational Science - the Participant's location - data. This population and systems-level model will guide and accelerate refinement of the CoC as more data are accumulated and added to the models.

HCV delivery relies on inherently complex and fragmented social systems that must deliver health care services to tens of thousands of patients across multiple non-overlapping and dynamic healthcare domains. An effective and useful HCV delivery system must therefore involve local and state governments, payers, multiple care and testing providers, healthcare and wellness facilities, pharmaceutical providers, and a complex social and behavioral environment. At best, these persons and entities can have only limited awareness of confounding social determinants and health situations. Thus there is an enormous 'coordination gap' when addressing population-based and patient-specific social, behavioral and health situations such as co-morbidities (including drug & alcohol addiction and mental health issues), inadequate access to transportation, and impoverishment. Indeed, simply coordinating testing, care referral, and healthcare delivery for a costly HCV DAA regimen among these groups is challenge enough. Yet the issue is compounded further by current public expectations for patient-centered care tailored to the needs of individuals and populations. The challenge is to understand these interventions within the context of the vast complexity of the healthcare delivery system and patient population. For example, which interventions improve care? In which populations? What are their intended consequences? How will we address and adapt to unintended consequences? What segments of the patient population are at greatest risk of falling out in the CoC? This proposal is innovative as we will provide proof of principle that a systems-level modeling approach is a feasible and scalable method to

address this complexity of human-level and entity-level interactions faced in delivering sustainable increases in HCV cure.

The goal of this project is to model and simulate the cascade of care involving the treatment of patients with Hepatitis-C to determine factors influencing positively and negatively the end treatment goals. The project utilizes the rich clinical experience of GU in combination with ORNL's big data analytics and counterfactual simulation of IDAMS-Hepatitis C viral infection to predict outcomes with computer simulated trials.

3.5 PROJECT: REAL-TIME ADAPTIVE MOTION CORRECTION FOR PROTON THERAPY

Proton Therapy is a relatively new method for effectively treating tumors associated with numerous forms of cancer. Treating prostate cancer is has been one of the initial treatment use cases for proton therapy. For men, the percentage is very high that at some point in their lives they will contract this disease. Thus the patient population that this treatment can reach is potentially huge, and this is just one of the clinical conditions proton therapy can address. The project team understands the importance of clinical oversight and involvement and includes Dr. Dritschilo MD, and Dr. Collins MD both from Georgetown University Medical Center. Dr. Dritschilo is a translational scientist with clinical experience, and a radiation oncologist with laboratory experience as a radiation biologist. He specializes in the treatment of breast and prostate cancer and holds a series of patents in radiosurgery and stereotactic methods. Dr. Collins is an attending physician in radiation oncology with expertise in prostate cancer and also utilizes and performs research using stereotactic methods. These researchers comprises a powerful team well equipped to address the significant challenges this project faces.

Due to the nature of proton beams, proton therapy has an advantage over other ionizing radiation treatment methods as it can reduce the radiation exposure patients receive during treatment. However, how to optimally apply the proton beam dose for tumors undulating due to patient respiration is still an active area of research and development. The current clinical best in care for proton therapy today is to perform a gated X-ray CT scan[1] of the patient prior to administering proton therapy from which the corresponding treatment plans for each gate interval are produced and subsequently applied during the administration of proton therapy. Here we realize that the gate interval amplitude and phase may vary on the order of 10 to 20% each, and the result can/will be a mismatch between the pre-treatment developed treatment plan for the corresponding gate interval and the current position of the patient tumor. Today this situation is unavoidable to varying degrees and can contribute to under dosing of tumor tissue as well as overdosing healthy tissue – both cases are undesirable and can lead to complications for patient recovery.

The objectives of this project are to first develop a proof of concept adaptive motion correction and treatment planning software for proton therapy and then in the second phase of the project enhance these software modules to run in (near) real-time at a rate sufficient to support performing proton therapy. Utilizing and evaluating the products produced here in the clinic are out of scope for this project. Rather, de-identified patient data obtained separately will be provided to the project by Provision, University of Tennessee Medical Center, and Georgetown Medical Center. These data will provide the input necessary for developing and testing algorithms produced via this project. Thus software will be developed with input from clinical experts however as proof of concept software, it will not be utilized in the clinic or with any patients or animals.

[1] Gated CT or 4D CT is comprised of a number of individual CT scans obtained at various phases of the respiratory cycle. http://cancer.ucsd.edu/care-centers/radiation-oncology/procedures/Pages/respiratory-gating.aspx

Computation science and computation performance are two major obstacles that performing advanced motion correction for proton therapy face and are briefly illustrated now. Due to the requirement to minimize patient radiation dose, it is not feasible to continuously perform X-Ray CT scans during treatment. For this reason, it is necessary to utilize the gated X-Ray CT data set acquired prior to performing therapy and to develop methods and non-linear models for transforming this static volumetric data set in real-time corresponding to the measured patient motion into a new, "warped" CT data volume. Performing this transformation in real-time is a significant computational challenge. Following each volume transformation, it will be necessary to re-identify the tumor boundaries utilizing segmentation and tracking algorithms, again, both computationally expensive and needing to be done in real-time. The transformed, segmented data are then input to treatment planning software which utilizes nuclear physics computations to convert the X-Ray CT data into proton beam position and beam power settings. Producing these treatment plans in real-time is also computationally expensive. Tight coupling of high performance computational resources will be necessary to achieve these project goals as it would be necessary to produce an updated treatment plan in 50 ms to 100 ms or better. The flow diagram for producing the treatment plans in real-time is shown below in figure 1. Following this process, the motion information will be used to demonstrate proof of concept custom adapted treatment plans intended for steering proton therapy in real-time to help better optimize proton beam dose delivery to tumor tissue.

The development of this motion correction system can potentially have application in other areas of research performed at the Department of Energy (DOE) user facilities. Typically, statically positioned samples are probed however motion correction could provide data with better fidelity and accuracy in cases where a sample is in motion or is changing shape with time. The dynamic nature of the sample under study can be quite important however remain largely unstudied due to the complexity of such experiments. For example, in the case of heating or cooling during an experiment, a sample may morph into another shape and such a phenomenon typically creates artifacts in standard tomography. Similarly, material migration such as fluids through materials or examining the lifecycle of surfactants face similar issues. Though carefully constructing the experiment including coupling the sample morphing time constant with the tomographic acquisition intervals can provide a framework from which motion compensation or correction can be performed. In so doing, new areas of scientific research can be enabled. Similarly, the motion correction, segmentation, and conformal mapping of the CT data produced via this project would be directly applicable to medical imaging with performance improvements helping to increase patient throughput as well as facilitating advanced medical research. DOE high performance computing resources would feature prominently in enabling and performing research involving dynamically changing conditions. Applying computing resources to dynamic processes is also a step in the direction of enabling experiment steering which is seen as a critical capability across a number of scientific research areas.

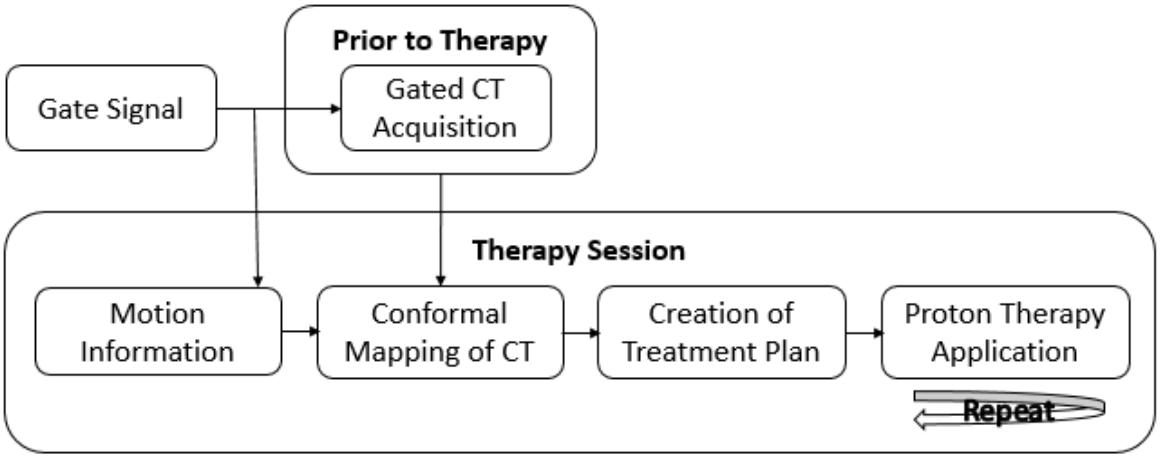


Figure 2. Block diagram illustrating the primary blocks and data flow to support real-time motion correction for proton therapy. Gated CT data are collected prior to performing therapy and subsequently utilized during therapy with motion information applied via conformal mapping in order to produce a custom treatment plan based upon the transformed X-Ray CT data volume with the tumor automatically segmented in real-time during the current gate interval. The process is repeated for each gate interval until the therapy session is complete.

This project capitalizes on the unique resources of partner. GU/Lombardi Cancer Center built a new proton-beam facility that will become operational in 2017, and Provision Health in Knoxville, Tennessee had just completed their facility and began treating patients in 2015. This allowed the assembly of an experienced multi-site and multidisciplinary team that can accomplish the primary goal of improving proton beam targeting by utilizing the super-computer resources of ORNL. The collaboration will utilize existing clinical data to develop innovative algorithms to better target and deliver proton therapy.

3.6 PROJECT: GENOMIC AND FUNCTIONAL CHARACTERIZATION OF THE MICROBIOTA OF OBESE AFRICAN AMERICANS

In this study, the gut microbiota of an obese group (Body Mass Index (BMI): 35.0–39.9) with those of a lean group (BMI: 18.5–24.9) will be compared. Both groups will be made up of 10 African American subjects who are 18 to 22 years old, female, and generally healthy. Stool samples from the recruited subjects will be collected on day 0 before any intervention. This will inform on the gut microbiota of obese and lean individuals before the experiment. To control for the effect of diet on the gut microbiota, we will provide the same meal plan for the subjects on both groups over a period of 10 days. After the first 7 days, we expect the microbiota of all subjects will be fully shifted to being sustained by the provided balanced diet and the energy intake and overall metabolism of both obese and lean subjects will converge on those based on the same diet. Fecal samples will be collected on days 8, 9 and 10. The integrated metagenomics, metatranscriptomics and metaproteomics analyses of these fecal samples will help to identify the community composition and functional structure of the microbiota in obese and lean subjects.

ORNL extracted total metagenomic DNA from 80 fecal samples using the MoBio PowerFecal DNA Isolation Kit. The 80 fecal samples will consist of 4 fecal samples per subject from 20 subjects. The team will sequence the 80 samples in 10 lanes by multiplexing 8 samples in a lane. After removing reads originated from human sequences, the metagenomic sequencing data will be assembled using two complementary approaches. The team used the Sigma algorithm [Ahn et al, 2015] to identify and reconstruct microbial genomes using reference genomes available in GenBank. A large number of

common fecal microorganisms have been sequenced by the HMP project as reference genomes. Sigma will map the metagenomic reads onto all reference genomes, identify microorganisms present in our samples, and reconstruct their genomes based on their variations from the corresponding reference genomes. Sigma has been scaled up to thousands of nodes on the Titan supercomputer. Such an analysis will help assess the distribution of an organism, or genes within an organism, across samples, and will also allow for the identification of regions within a genome that are distributed in unusual ways suggestive of their role in adapting organisms for particular environments. Sigma will also be used to identify genomic regions with sharply altered coverage, marking strain variations, laterally transferred genes, and/or candidate pathogenicity islands. Because the same population of microorganisms is expected to be present at different relative abundances across the triplicate fecal samples, the team will use the coverage depth variations between the triplicate fecal samples to bin the contigs into microorganism bins and produce genome-resolved metagenomes. The de novo assembly may generate genomes of novel microorganisms not represented by any reference genome in the GenBank. Genomic sequences will be annotated using a gene prediction tool, Prodigal, and a protein function assignment system based on the UniFam database, as described previously [Chai et al, 2014].

The team extracted total proteins from fecal samples using a SDS boiling method [Xiong et al, 2015] and be processed as complex metaproteome samples [Li et al, 2014]. The team then analyzed the 80 metaproteomes in ten 24-hour two-dimensional liquid chromatograph-tandem mass spectrometry runs (2D-LC-MS/MS) on LTQ-Orbitrap mass spectrometers at ORNL. In each 2D-LC-MS/MS run, the Contractor will multiplex 80 samples from post-diet fecal samples of an obese/lean pair using the iTRAQ 8-plex reagents as described [Mosier et al, 2014]. Proteins will be identified by searching the 2D-LC-MS/MS data of each subject using the Sipros algorithm against a matched protein sequence database constructed from their own metagenomes and metatranscriptomes. The protein abundance changes between the obese/lean pairs will be quantified based on the TMT reporter ion intensities using the ProRata algorithm [Wang et al, 2013].

Through comparisons of the obese and lean subjects, these analyses will identify taxonomic groups and metabolic signatures that are associated with or indicative of obesity. The results are expected to yield a holistic genes-to-systems view of the taxonomic and functional underpinnings of the microbiota associated with obesity. Specifically, the Participant and the Contractor will jointly extract the following expected results: (1) Identification of genes that are candidates for biomarker or significant risk factor for obesity. (2) Identification of taxonomic groups and associated metabolic pathways that show differences in obese subjects; (3) Identification of gene content and genome structure variation in sequenced microbial genomes between lean and obese groups; and (4) Draft assemblies of abundant taxonomic groups that may be implicated in the processes that are causative or consequential of obesity.

The extracted Deoxyribonucleic Acid (DNA) samples are presently being sequenced for metagenomics to assess similarities and differences in bacterial genes' functions. Ribonucleic acid (RNA) extracts are presently undergoing Next Generation Sequencing for transcriptomic analysis to assess the dynamics of gene expression in the two groups of subjects (Obese vs. Lean) under different diets. Protein extracts are currently being analyzed at ORNL for metaproteomics analysis.

3.7 PROJECT: USE OF ATRA© TECHNOLOGY TO ENABLE COMPLETE PATIENT PRIVACY WHEN MERGING AND ANALYZING HEALTHCARE DATABASES

The purpose of this project is to validate a new technology ("ATra") that has the capacity to resolve privacy and security issues when merging Electronic Health Record clinical databases. ATra is a technology platform that creates a secure environment when analyzing sensitive data. ATra does not require "owners" of data to share their data with other "owners" in order to perform multi-dimensional analysis. Unlike other methods that seek to anonymize data, ATra can be engineered to remove the human

analyst from the data analysis. ATra is a core component of the AvesTerra distributed, large-scale, analytics environment

ATra therefore has wide applicability to one of most important clinical research approaches of this era – the analysis of "big data" in clinical care. We (Participant and Contractor) are applying ATra technology to Type II diabetes mellitus in merged VA and MedStar EHR (Electronic Health Record) data made available to investigators who are joint appointed at Georgetown University (Participant) and DCVA. In this proof of concept project, will utilize ATra in a T4/5 study to provide preliminary data on certain Type II Diabetes Mellitus parameters and a preliminary analysis if there are differences between mono users of VA healthcare and dual users of VA and the MedStar Health System. On one hand, dual users may have better results because they cleverly utilize both healthcare systems or they may be sicker and need more extensive care. Our plan is to leverage the results of this pilot evaluation in a CTSA (Clinical Translational Science Award) U01 or R21 proposal.

Participant seeks to partner with Contractor's Cyber and Information Security Research Group (CISR) to utilize the CADES (Compute And Data Environment for Sciences) facility utilizing Participant's ATra Black Box platform within Contractor's highly- and provably-secure environment for sharing and cross-correlating PHI (Public Health Information) from two distinct HealthCare entities, MedStar Health and the Washington District of Columbia Veterans Administration (DCVA). All key personnel from these entities are jointly appointed Georgetown faculty and are acting as faculty of Participant. The ATra system developed at Georgetown University provides a unique method for multiple organizations to analyze private information while maintaining a very high level of privacy assurance. At the core of the system is a computing device this is specifically engineered so that its contents can never be revealed to anyone under any circumstances. That is, the device operates as a "Black Box" in which sensitive information from multiple organizations can be input, but never allowed to be output.

This project did not initiate due to regulatory hurdles and timelines.

3.8 PROJECT: HYBRID NANOSTRUCTURED/MICROSTRUCTURED PLATFORMS FOR THE ENRICHMENT OF CIRCULATING TUMOR CELLS

Although great advances have been made in cancer detection and therapy, the disease continues to exact an enormous toll in morbidity and mortality. The majority of cancers in industrialized nations are solid epithelial tumors (i.e. carcinomas) and the main cause of death from these tumors is not the primary cancer, but rather metastases which arise from them (Alix-Panabieres et al, 2008). Spreading of tumor cells from the primary site of the tumor via the blood can be considered a crucial step in the metastasis process, a fact recognized more than a century ago (Ziegschmid et al,2005; Fidler, 2003). Many recent studies (examples: Cristofanilli et al, 2004, Allard et al, 2004, Fehm et al, 2002, Riethdorf et al, 2007, Liu et al, 2009, Ignatiadis, 2011; Reviews: Riethdorf and Pantel, 2010, Lin et al, 2011, Mocellin et al, 2006, Dotan et al, 2009, Mavroudis, 2010, Braun and Naume, 2005) have demonstrated the potential clinical relevance of detection and enumeration of these circulating tumor cells (CTC), both with regard to patient prognosis and staging of therapy. Unfortunately, technological shortfalls still limit the potential for widespread application of CTC measurements and resulting advances in relating such measurements to tumor biology and clinical course (Schmidt, 2010).

CTCs are present in a milieu of a vastly greater number of formed blood elements, and the phenotype of the CTCs can be very heterogeneous. Therefore, unambiguous identification and characterization of CTCs demands analytical techniques of extreme sensitivity and great specificity (Alix-Panabieres et al, 2008). The usually stated requirement for detection is ability to detect 1 CTC in a background of 10^6-10^7 leukocytes, and therefore some form of tumor cell enrichment is generally employed.

Various enrichment strategies have been described, including density gradient centrifugation, isolation by size of epithelial tumor cells (filtration), and, most commonly employed, immunomagnetic enrichment. In this technique, CTCs are removed from the bulk of leukocytes by using magnetic microbeads or colloidal ferro-suspensions which have anti-EpCAM (epithelial cell adhesion molecule) antibodies attached to the magnetic particles. CTCs, by virtue of having EpCAM on their surface, bind to the magnetic particles and can be separated from the bulk of leukocytes by a magnet. Subsequently, immunochemical staining using two antibodies, one specific for epithelial tumor cells (anti-cytokeratin), and one specific for a leukocyte protein (anti-CD45) are used to detect and differentiate CTCs from leukocytes (many articles, for example: Fehm et al, 2002, Moreno et al,2005, Fizazi et al, 2007, Danila et al, 2007, Wang et al, 2000, and Talasaz et al, 2009). This procedure is essentially that best exemplified by the CellSearch device, which has received FDA approval for use in CTC enumeration per 7.5 mL blood in metastatic breast, prostate and colon cancer. A known limitation of this approach is difficulty in isolating *viable* cells, due to agglomeration, shear and other factors associated with magnetic bead separation strategies.

Recently, microfluidic manifolds with either in-line microposts or "herringbone"-based advection mixing features have been described to provide solid-phase antibody-based capture of CTCs (Stott et al, 2010, Nagrath et al, 2007). Microscale dimensions of these manifolds provide laminar flow and thereby impart less shear to captured cells than conventional bead-based methods. The Nagrath CTC-chip employs arrays of microposts as fluidic barriers which essentially define capillary beds to provide high-surface area capture cross section along the length of the fluidic manifold. This capillary-bed approach forces the required flow through an extremely limited path and inherently limits the overall throughput of the device. The Stott, 'herringbone' chip features a coverslip with a staggered herringbone vortex generator, but otherwise an essentially open microfluidic flowpath. In contrast to axial disruptions by solid posts, the vortex generator is designed to disrupt axial cell trajectories resulting at low Reynolds flows, and thereby to increase interactions between CTCs and the antibody-functionalized sidewalls of the device. The open architecture of this approach addresses throughput restrictions of 'post-based' designs, but results in inefficient capture at higher flow rates. As flowrate increases, the mean free path of cell/wall interactions increase, therefore requiring longer flowpaths for efficient cell capture. Capture efficiency of both of these approaches drops dramatically with flow rate, with the 'herringbone' chip achieving only 40% and the CTC-chip achieving <10% capture efficiency at the maximum published flowrate of 0.5 mL/hr flowrates. Even at this flow rate, >14 hrs would be required for the volume of blood currently processed by the commercially available CellSearch device (7 mL)!

We anticipate that more effective capture can be realized using a hybrid of the micropost and herringbone approaches. The strength of the micropost approach is that the posts present disruptions orthogonal to the major axis of flow, resulting in the benefit of near continuous cell/barrier interaction (i.e. short mean free path). However, the capillary-bed basis of the design and limitations in lithography and materials properties requires that the posts be large with respect to the overall flow cross section. For example, the described Nagrath chip employs 100 μm posts with 10 μm flow channels, with a post:channel ratio of 10:1. With this geometry, volumetric throughput is extremely limited. Conversely, the herringbone approach has the advantage of an essentially open flow path, yielding a much higher volumetric flow rate. However, by providing capture only at the periphery of channel walls this advantage is lost at higher flow rates due to increases in the cell mean free path (decreasing probability of cell/surface interactions). A hybrid solution is achieved by the ability to construct arrays of obstructions orthogonal to the axial flowpath, but using post elements with a much smaller cross-section with respect to the channel width. We have developed a unique fabrication approach whereby each post can be <100 nm in diameter, yet sufficiently physically robust to survive flow and cell capture. Spacing arrays of these posts at a staggered 10-20 μm pitch provides a post array chip with a post:channel cross section of 1:50 – 1 :100. This potentially provides a 1000-fold higher volumetric throughput than the Nagrath chip and is nearly on par with the open-path of the herringbone chip with a mean free path for cell/post interactions of only 10 μm, independent of flow rate.

We have evaluated this hybrid nano/micropost approach for capture and harvest of CTCs. This invovled the production of three different designs of microfluidic CTC harvest chips. Two will be based on the conventional Nagrath and Herringbone devices. The third will be based on the incorporation of nanostructured post arrays using a unique class of material, the vertically-aligned carbon nanofiber (VACNF). This material has been developed at ORNL for a variety of microfluidic and cellular interfacing applications including microfiltration systems (Fletcher 2004), gene and drug delivery arrays (McKnight 2003, 2004; Mann 2008; Peckys 2008,2009), and electroanalytical/electrophysiological probing systems (McKnight 2004, 2006; Yu 2007). In this effort, arrays of VACNFs will be incorporated into microfluidic designs in substitution of conventional lithographically defined posts and herringbone features.

As an aid to device design, the blood flow through the microchannel with the presence of microposts will be simulated using ANSYS-CFX. The blood (or plasma) will be modeled as a Newtonian fluid with viscosity and density relevant to the experiments. To model the cells motion and interactions with the microposts, a particle transport model will be used, where the cells will be tracked through the flow in a Lagrangian way. The "stickiness" or interaction between the posts and the cells will be investigated by using restitution coefficient on the posts.

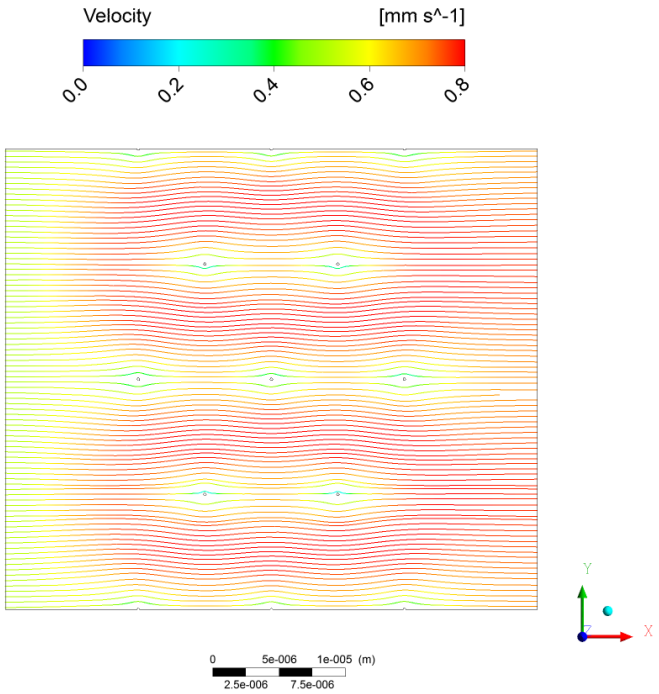


Figure 3. Streamlines for blood flow through a microchannel with 200nm diameter microposts 10 microns apart.

Each of the three platforms will then be functionalized with immobilized anti-EPCAM and evaluated for both 'CTC' capture efficiency and volumetric processing rate using cell populations spiked with known quantities of epCAM-expressing CTCs. As this effort will focus on the isolation and enrichment steps, these initial experiments will be conducted using normal and CTC cells fluorescently labeled (celltracker Green and Orange, respectively) for rapid enumeration. Results from the most promising platform(s) will

subsequently be compared to those obtained using the CellSearch technology on blood samples collected from metastatic breast cancer patients through an institutional protocol at Georgetown University Medical Center.

4. SUBJECT INVENTIONS

No Subject inventions have been associated with this CRADA

5. COMMERCIALIZATION POSSIBILITIES

No commercialization possibilities have been associated with this CRADA

6. PLANS FOR FUTURE COLLABORATION

The current CRADA team of UT-Battelle, LLC. and Georgetown University have submitted and won an award to continue the work under the National Institutes of Health, National Center for Advancing Translational Sciences Clinical and Translational Science Award. The purpose of the continued working relationship between the participants is to further a collaboration that will develop insights into the systems genetics underlying individual susceptibilities to diseases, such as cancer, metabolic and neurological disorders. ORNL will bring to bear the power of high performance computing, and broad scientific expertise to develop better early markers of disease through state-of-the-art applications such as artificial intelligence and accessing and processing unconventional data sources and highly refined geospatial mapping.

ORNL's work will be in the area of environmental determinants of health with the aim to develop, implement and support innovative projects that will identify environmental contributors to health risks in a manner that will inform more targeted future research endeavors. There is strong emerging evidence for a substantial role of environmental factors in the health disparities that disproportionately burden African Americans. Using data from the National Health Interview Survey collected between 1989 and 1994, it was found that 15 to 75 percent of the disparities between Black and White respondents are accounted for by residential context. The survey cited factors contributing to the effects of place on health as including environmental exposures (e.g., toxins), built environment (e.g., availability of safe recreational facilities), and social conditions (e.g., exposure to neighborhood violence and drugs). In a separate study, self-ratings of poor health and impairment were 2 to 3 times higher among those in the highest quartile for neighborhood problems compared to the lowest quartile. These findings were independent of individual risk factors and neighborhood socioeconomic status, which inferred that an important contribution of chronic stress, in addition to the factors cited above, could be attributed to the effect of environment on health. Further elucidation of mechanisms by which threatening environments impact health is needed to both understand and potentially ameliorate health disparities.

Additionally, another joint venture that has occurred stemming from the CTSA partnership is the collaboration with DOE/ORNL/Battelle and the Department of Veteran's Affairs. ORNL was selected to house the Million Veteran Program's data, and this assignment affords our contemplated collaboration and CTSA partnership the possibilities of collaborating and combining our different expertise and resources to perform high-level analytics synergistically on this as well as additional data. These ventures are timely, because they will provide the expertise and methods to contribute to other important national initiatives such as Precision Medicine, the Moon Shot for cancer cures, the CURES Act, and the National Strategic Computing Initiative. These national initiatives are in many ways based on the foundation built upon the CTSA collaborative team science mission.

7. CONCLUSIONS

The purpose of this research and development agreement has been to establish a collaboration between ORNL and GU that will develop insights into the systems genetics underlying individual susceptibilities to diseases, such as cancer, metabolic and neurological disorders. ORNL has brought to bear the power of high-performance computing, and broad scientific expertise to develop better early markers of disease through state-of-the-art applications such as artificial intelligence and accessing and processing unconventional data sources and highly refined geospatial mapping. Additionally, this multi-task research and development agreement has established a collaboration that pioneered Systems Medicine via the resources, technologies and expertise available only at ORNL. These resources include multi-domain protein structures; chip-based biomarker detection tools; computational biology; systems genetics and quantitative genomics; high end crystallography and drug target development.

The fruits of this collaboration have realized advances in "Personalized Medicine" and educated the next generation of physicians in the context of interdisciplinary science. For DOE, the goal was to incorporate input from potential users of the next generation of supercomputers to influence the requirements and overall design of HPC. This effort strengthens the computing mission of both DOE and ORNL by requiring large-scale data analytics that push the boundaries of networking, compute processing, data storage, visualization and analytical tools. Finally. the work was in line with the scientific mission of DOE across multiple domains of biology, genomics, and other.