RESEARCH STATEMENT

Bohdan Khomtchouk, Ph.D. (bohdan@uchicago.edu)

Dear Committee:

As a fan of Formula One (F1) motorsport, one of my favorite quotes comes from the F1 racing legend, Ayrton Senna, who in the 20th century said:

> *"Because in a split second, it's gone."*

Limited by the technology of his time, Ayrton Senna was killed at the age of 34 after his car crashed into a concrete barrier while he was leading the 1994 San Marino Grand Prix at the Autodromo Enzo e Dino Ferrari in Italy. Today, such accidents are virtually unheard of, courtesy of cutting-edge technology that shields drivers in even the most spectacular of crashes.

Like Ayrton Senna, my father (and my uncle who subsequently replaced him as my father figure) both died in a split second – from stroke and heart attack, respectively, in their early 40s. Yet, we currently do not have the technology to forecast or prevent unexpected tragedies like this. These personal experiences ultimately had a profound impact on my professional career development and today the goal of my lab at UChicago Medicine is to build a world-class research program in cardioinformatics [1] by leveraging bioinformatics and computational biology approaches to illuminate novel cardio-(vascular/renal/metabolic) biology and enable new therapeutics.

Background and Current Work

Similar to F1, the field of medicine has made equally important technological strides in the last 30 years – constantly pushing the boundaries of engineering, information technology and data science to extend the frontiers of their respective fields and address unmet needs – and it's my belief that the next generation of F1 drivers share more in common with the next generation of precision medicine researchers than most people initially realize. Racing against the clock, cardiovascular disease (CVD) is currently the leading cause of death worldwide and, by 2030, almost half of the global adult population is projected to have a CVD diagnosis of some kind (hypertension, coronary artery disease, myocardial infarction, stroke, etc.). By 2050, the aging population over the age of 65 will triple [2], creating an impending rise in future demand for healthcare innovation within the CVD market sector, since the most important determinant of cardiovascular health is a person's age [3], which is why age is such a critical component of CVD etiology.

In 2017-19, as an American Heart Association (AHA) Postdoctoral Fellow at Stanford University, I set out on a bold mission to pioneer a new field of science called cardioinformatics, with the goal of fighting heart disease (and its associated renal and metabolic comorbidities) with computation by creating better drugs. I was inspired by the momentum of a *Journal of the American Heart Association* publication I had co-authored in graduate school at the Center for Therapeutic Innovation of the University of Miami Miller School of Medicine [4], where I served as the lead bioinformatician on an international study whose computational results were experimentally validated over the course of several years. As a PhD candidate and National Defense Science & Engineering (NDSEG) Fellow in the Human Genetics & Genomics PhD program at the time, I recognized early on that the bioinformatics and computational genomics footprint in cardio-(vascular/renal/metabolic) disease was modest relative to cancer, neurodegenerative diseases or other causes of death, and so my hypothesis was that expanding this computational footprint with improved data science techniques will ultimately improve healthcare in the context of precision cardiology, as has been the case for precision oncology. A few years later, when I joined the faculty at the University of Chicago in 2019, our department chair of cardiology at UChicago Medicine commented on my lab's research mission in a press release [5]:

> *"The field of cardiovascular disease has always been data and technology driven, but when it comes to taking advantage of the vast amount of information that already exists and developing a systematic approach for analyzing it, the field has fallen short. The emergence of computational biology and bioinformatics in the cardiovascular field will provide insights and potential therapeutic avenues for the treatment of cardiovascular patients."*

This year, my lab's research program was highlighted in a press release entitled: "Databases of the Heart: A computational biologist takes on the world's number one killer"[6].

Building next-generation technology for CVD drug discovery

Inspired by data commons and data ecosystems for genomic and clinical data in the field of oncology (e.g., cBioPortal (cbioportal.org) and NCI Genomic Data Commons (GDC) (gdc.cancer.gov)), I developed a similar knowledge portal infrastructure framework in the field of cardiology. With a focus on cardiovascular disease (CVD) phenotypes, HeartBioPortal (heartbioportal.com) brings together all CVD-relevant gene expression, genetic association, alternative splicing, and ancestry data and bioinformatically preprocesses it into easy-to-interpret figures and charts. Previously published in *Circulation: Genomic and Precision Medicine* in 2019 and *Database (Oxford)* in 2020 [2, 7], this platform technology is now being used at Dock Therapeutics, Inc. as a digital drug discovery dashboard serving big pharma clients such as Novo Nordisk and large biotechs such as NewAmsterdam Pharma B.V.

In general, I am interested not only in bridging the bench-to-bedside divide, but also the informatics-to-medicine divide that still exists in biomedical data science. Broadly speaking, my research can be classified into three major categories:

1. Independent tools/methods development [databases, software packages, and other resources that enable the broader research community (e.g., HeartBioPortal [2, 7], Shinyheatmap [8], HeatmapGenerator [9], geneXtendeR [10], MicroScope [11], NERO [12, 13], etc.)].
2. Bioinformatics and AI/machine learning support in a collaborative team science setting that focuses on the biology (e.g., publications in *Journal of the American Heart Association* [4], *Molecular Psychiatry* [14], *Addiction Biology* [15], *npj Biofilms and Microbiomes* [16], etc.).
3. Secondary analysis of existing genetic and associated clinical datasets (e.g., publications in *G3: Genes, Genomes, Genetics* [17], *Atherosclerosis* [18], including a press release on a recently funded K12 grant on HIV-associated cardiovascular disease [20]).

As a junior faculty member and early career investigator[1], most of my work has greatly benefited from interaction with a number of colleagues and my former academic advisors (graduate and postdoctoral). Even single-author manuscripts [21, 22] could not have been written without the generous input of multiple department colleagues, who often internally reviewed and commented upon my work. I describe my work below, with a particular focus on publications, preprints, and recently funded grants that highlight my 5-10 year research plan.

Cardioinformatics approaches to illuminate novel cardio-(vascular/renal/metabolic) biology and enable computationally-derived therapeutics

The Khomtchouk Lab at the University of Chicago Medicine is developing a research program that concerns the emerging field of cardioinformatics, working at the nexus of bioinformatics and precision cardiology, with the goal of integrating computational biology and AI/machine learning-based approaches to illuminate cardio-(vascular/renal/metabolic) biology and enable new therapeutics. Our lab members are pioneering the field of cardioinformatics with the overarching mission to fight heart disease with computation by creating better drugs through genetics-driven novel target discovery and repurposing efforts. To validate our approach in the markets, we launched a successful biotech company called Dock Therapeutics [23], spun out of my lab in April 2021, currently in business with Novo Nordisk and NewAmsterdam Pharma and in contract discussions with key stakeholders at Pfizer, Bristol-Myers Squibb, Novartis, Merck, and Boehringer Ingelheim.

Here is a list of several translational research questions that will guide my lab's 5-10 year research mission:

1. *How do we leverage single-cell data to construct more precise/tailored cardiovascular medicine therapies in diverse human populations?*

   Following up on our *Database (Oxford)* publication in 2020 entitled "HeartBioPortal2.0: new developments and updates for genetic ancestry and cardiometabolic quantitative traits in diverse human populations" [2], we recently published a follow-up paper in *Atherosclerosis* entitled: "Enhanced single-cell transcriptomics workflow reveals coronary artery disease cellular cross-talk and candidate drug targets" [18]. Similar to HeartBioPortal, we developed an open source web application, PlaqView (`http://plaqview.com`), to allow users to easily interact with existing single-cell datasets relevant to atherosclerosis and coronary artery disease and run their own analyses. Using this resource, we showed how to use single-cell RNA-seq analysis of human coronary arteries to reveal vascular smooth muscle cell (VSMC) transitions and candidate drug targets. We developed an enhanced, user-friendly, and reproducible single-cell RNA-seq workflow that has multiple key features: (1) automated cell type assignment, (2) pseudotemporal analysis of gene expression, (3) cell-cell communication network inference, and (4) druggability analysis. We applied this workflow to uncover key mechanisms and candidate drug targets for coronary artery disease. Specifically, we identified distinct derivations of chrondrocyte-like and fibroblast-like cells from smooth muscle cells. We then highlighted several key ligand-receptor interactions with potential for drug development or repurposing to treat atherosclerotic disease. These results shed new light on the protective and maladaptive smooth muscle cell phenotype transitions and provide testable hypotheses for experimental perturbation studies. A press release was subsequently issued [19].

2. *How do we leverage large-scale genetic association data to construct more precise/tailored cardiorenal medicine therapies?*

   My 5-10 year research plan in cardioinformatics focuses not only on bioinformatics approaches to cardiovascular medicine, but also cardiorenal and cardiometabolic disease phenotypes. My lab currently has a preprint under review [24] at *JACC: Basic to Translational Science* entitled "Epidermal Growth Factor Receptor Inhibition Prevents Caveolin-1-dependent Calcifying Extracellular Vesicle Biogenesis", where the focus is on chronic kidney disease (CKD). Vascular calcification is the leading predictor of cardiovascular morbidity, and widespread medial calcification is especially prevalent in individuals with CKD. Computational analyses led by my team (primarily myself and an undergraduate research assistant in my lab) showed a correlation between serum concentrations of epidermal growth factor receptor (EGFR) and the presence of coronary artery calcium. Specifically, our cardioinformatics analyses of the MESA and Framingham cohorts showed for the first time that individuals with single-nucleotide polymorphisms (SNPs) associated with increased serum concentrations of EGFR had elevated coronary artery calcium scores. Together with our wet-lab collaborator (Dr. Joshua Hutcheson at Florida International University), we tested and confirmed this

[1]Note that my ESI status officially expired in May 2022 upon receipt of my first NIH R01 grant from NIDDK on cardiometabolic disease (diabetes), comprising $> 2M$ in total funding over a 4-year period. See CV for details.

computational finding experimentally using both *in vitro* cultures of human coronary artery smooth muscle cells and a murine model of chronic kidney disease, essentially tying together *in silico* with *in vitro* and *in vivo* work.

EGFR also interacts with caveolin-1, which is a known contributor to vascular calcification. In our study, we showed that inhibiting EGFR kinase activity prevents CKD-induced vascular calcification in mice and mineralization of human coronary artery smooth muscle cells in culture. We were surprised to find no effect on osteogenic phenotypes of cells in tissue nor in culture. Our experimental data indicated that inhibition of calcification by EGFR inhibition occurs due to altered caveolin-1 trafficking, which prevents the formation of calcifying extracellular vesicles (EVs). Calcifying EVs serve as the nucleating foci for vascular calcification, so preventing their formation prevents mineralization. The EGFR inhibition had the opposite effect in bone, increasing mineral density!

Taken together, our *in silico* data analyses of the MESA and Framingham cohorts ultimately led us to our *in vitro* results, which supported the *in vivo* findings that EGFR inhibition reduced the release of procalcific CAV1-positive EVs. Given that EGFR inhibitors exhibit clinical safety and efficacy in other pathologies such as cancer, the current data suggest that EGFR may be an ideal target to prevent pathological vascular calcification. This cardioinformatics workflow highlights the importance of bridging not only the bench-to-bedside, but also the informatics-to-medicine divide that still exists in modern precision cardiology research to enable computationally-derived therapeutics, as we have done in this manuscript. We're currently working to better understand the mechanisms through which EGFR and caveolin-1 interact in calcification and believe that the novel association between EGFR signaling and vascular calcification and the resultant implications in improved cardiovascular health in CKD warrants further consideration for drug development (e.g., in clinical indications of high unmet need such as end-stage renal disease).

3. *How do we drug the cytoskeleton (and the underlying extracellular matrix that supports its cellular architecture) to enable an entirely new class of cardiovascular therapeutics?*

   We have a recently published manuscript at *Expert Opinion on Drug Discovery* entitled "Targeting the Cytoskeleton and Extracellular Matrix in Cardiovascular Disease Drug Discovery" [25] where we demonstrate how the cardiac cytoskeleton and extracellular matrix (ECM) prove to be a tractable therapeutic target in cardiovascular disease drug discovery (e.g., by therapeutically altering plaque morphology). Our study builds off of previous publications that indicate relationships between ECM proteins and disruptions in the cellular architecture. These previously undiscovered targets and undervalued assets present new research and development opportunities for novel target discovery and repurposing initiatives, both of which are covered extensively in our published manuscript. For example, we describe how observed changes in cellular architecture during cardiovascular health-to-disease transitions are accompanied by cytoskeleton/ECM dysregulation. Therefore, a cardioinformatics approach to analyzing cytoskeletal/ECM-related genes may potentially lead to the discovery of novel therapeutic modalities by facilitating systematic target identification and prioritization. Since the mechanisms behind several currently existing therapeutics support the idea of targeting the cytoskeleton/ECM at different cardiovascular disease development stages, we believe that our next few years of work in this space may spotlight an entirely new therapeutic modality in cardiovascular medicine.

4. *How can wearable devices and sensors enable better cardiometabolic health and digital therapeutics?*

   I am currently writing an R01 proposal and AHA Career Development Award based on the hypothesis that year-long patient recordings from multiple wearable devices, such as those that record continuous real-time measurements of glucose, heart rate, blood pressure, and sleep/wake cycles, can both retrospectively impute and prospectively predict a spectrum of cardiometabolic outcomes, such as arrythmia or type 2 diabetes mellitus (T2DM) and, likely, diseases outside of the immediate cardiometabolic spectrum, especially when integrated with other data modalities (e.g. diagnoses, medications, procedures, family history, and clinical notes from a patient's electronic health record, as well as socioeconomic status). The focus of my proposal has been on the adult population across socioeconomic and ethnic groups on the Southside of Chicago, which is comprised mostly of African Americans and Hispanics, who are most exposed to the risk of developing metabolic and cardiovascular disease outcomes. This unique patient population will be an asset when developing new approaches to integrating wearable devices into telemedicine in the time of the COVID-19 pandemic. Particular attention will be given to recruiting women and underrepresented minorities.

5. *How can cardioinformatics approaches enable novel target discovery in various subtypes of monogenic diabetes?*

   In addition to the cardiovascular and cardiorenal research opportunities discussed above, I recently received my first funded R01 grant (1R01DK132090-01) in cardiometabolism from the NIDDK, on which I am an MPI with Dr. Fumihiko Urano at Washington University School of Medicine. Our goal is to computationally explore the molecular genetics of hereditary endoplasmic reticulum (ER) diabetes to establish functional studies of gene variants affecting ER homeostasis, design treatments targeting common molecular pathways altered in ER stressed β cells, and identify other ER genes involved in β cell dysfunction and death. In this project, we will characterize *WFS1* and *CISD2*, *EIF2AK3*, *IER3IP1*, and *INS* variants using functional assays and cardioinformatics approaches to test novel treatments targeting the common molecular pathways altered in β cells expressing pathogenic variants of *WFS1* and *CISD2*, *EIF2AK3*, *IER3IP1*, and *INS* genes. Successful completion of this study will lead to the establishment of new precision medicine therapeutics approaches for hereditary ER diabetes.

To continue the F1 metaphor further, upon which this research statement is built, I would like to join a crew of faculty members that help me launch the next-generation of computationally-derived precision therapeutics for cardio-(vascular/renal/metabolic) disease

and continue to help me pioneer the field of cardioinformatics. Therefore, I am applying to your institution to assess mutual fit and meet with your research teams to discuss a tenure-track faculty position at the rank of Assistant Professor and the establishment of my future lab on the road to tenure.

On a personal note, I want to thank every team member of my lab — past, present, (and future) — for working extremely hard in their relentless pursuit of fighting heart disease with computation. It's through their pioneering efforts that we are building the field of cardioinformatics. Thank you for joining me.

#WeRaceAsOne,

Bohdan Khomtchouk, Ph.D.
Instructor, Department of Medicine
Section of Computational Biomedicine and Biomedical Data Science
Master of Science in Biomedical Informatics (MScBMI) Program
Institute for Genomics and Systems Biology
University of Chicago
900 East 57th Street, Knapp Center for Biomedical Discovery
Office suite #10150
Chicago, IL 60637

Assistant:
Ms. Temi Okubadejo
(773) 702-6685
tokubade@bsd.uchicago.edu

References

[1] **Khomtchouk BB***, Tran DT, Vand KA, Might M, Gozani O, Assimes TL: "Cardioinformatics: the nexus of bioinformatics and precision cardiology", *Briefings in Bioinformatics*, 2020, 21(6): 2031-2051. doi: 10.1093/bib/bbz119 (* indicates corresponding authorship)

[2] **Khomtchouk BB***, Nelson CS, Vand KA, Palmisano S, Grossman RL: "HeartBioPortal2.0: new developments and updates for genetic ancestry and cardiometabolic quantitative traits in diverse human populations", *Database*, 2020. (* indicates corresponding authorship) https://doi.org/10.1093/database/baaa115

[3] North BJ, Sinclair DA: "The intersection between aging and cardiovascular disease", *Circulation Research*, 2012. 110, 1097–1108.

[4] Gidlöf O, Johnstone AL, Bader K, **Khomtchouk BB**, O'Reilly JJ, Celik S, Van Booven DJ, Wahlestedt C, Metzler B, Erlinge D: "Ischemic Preconditioning Confers Epigenetic Repression of Mtor and Induction of Autophagy Through G9a-Dependent H3K9 Dimethylation", *Journal of the American Heart Association*, 2016. 5(12): e004076

[5] Richards S: "Research sets sights on heart disease discoveries through precision cardiology and bioinformatics." *University of Chicago Medicine Press.* https://www.uchicagomedicine.org/forefront/heart-and-vascular-articles/2019/october/research-sets-sights-on-heart-disease-discoveries-through-precision-cardiology-and-bioinformatics

[6] Baker P: "Databases of the Heart: A computational biologist takes on the world's number one killer." *University of Chicago Professional Education.* https://professional.uchicago.edu/stories/master-science-biomedical-informatics/databases-heart

[7] **Khomtchouk BB***, Vand KA, Koehler WC, Tran DT, Middlebrook K, Sudhakaran S, Nelson CS, Gozani O, Assimes TL: "HeartBioPortal: An Internet-of-Omics for Human Cardiovascular Disease Data", *Circulation: Genomic and Precision Medicine*, 2019, 12(4):e002426. (* indicates corresponding authorship)

[8] **Khomtchouk BB***, Hennessy JR, Wahlestedt C: "Shinyheatmap: ultra fast low memory heatmap web interface for big data genomics", 2017, *PLoS One* 12(5): e0176334 (* indicates corresponding authorship)

[9] **Khomtchouk BB***, Van Booven DJ, Wahlestedt C: "HeatmapGenerator: high performance RNA-seq and microarray visualization software suite to examine differential gene expression levels using an R and C++ hybrid computational pipeline", *Source Code for Biology and Medicine*, 2014, 9(30) (* indicates corresponding authorship)

[10] **Khomtchouk BB***, Koehler WC, Van Booven DJ, Wahlestedt C: "Optimized functional annotation of ChIP-seq data", 2019, *F1000Research* 8(612) (* indicates corresponding authorship)

[11] **Khomtchouk BB***, Hennessy JR, Wahlestedt C: "MicroScope: ChIP-seq and RNA-seq software analysis suite for gene expression heatmaps", 2016, *BMC Bioinformatics* 17(390) (* indicates corresponding authorship)

[12] Wang K, Stevens R, Alachram H, Li Y, Soldatova L, King R, Ananiadou S, Schoene AM, Li M, Christopoulou F, Ambite JL, Garg S, Hermjakob U, Marcu D, Sheng E, Beißbarth T, Wingender E, Galstyan A, Gao X, Chambers B, Pan W, **Khomtchouk BB***, Evans JA, Rzhetsky A: "NERO: A Biomedical Named-entity (Recognition) Ontology with a Large, Annotated Corpus Reveals Meaningful Associations Through Text Embedding", *npj Systems Biology and Applications*, 2021 7(38). (* indicates co-corresponding/co-senior authorship, responsible for software package development, data analysis and interpretation) https://www.nature.com/articles/s41540-021-00200-x

[13] Khomtchouk BB: "NERO-nlp: A biomedical Named Entity (Recognition) Ontology." *The Python Package Index (PyPI).* https://pypi.org/project/NERO-nlp/

[14] Barbier E, Johnstone A, **Khomtchouk BB**, Tapocik J, Pitcairn C, Rehman F, Augier E, Borich A, Schank J, Rienas C, Van Booven DJ, Sun H, Nätt D, Wahlestedt C, Heilig M: "Dependence-induced increased of alcohol self-administration and compulsive drinking mediated by the histone methyltransferase PRDM2", 2017, *Molecular Psychiatry* 22: 1746-1758.

[15] Johnstone A, Andrade N, Barbier E, **Khomtchouk BB**, Rienas C, Lowe K, Van Booven DJ, Domi E, Esanov R, Vilca S, Tapocik J, Rodriguez K, Maryanski D, Keogh MC, Meinhardt M, Sommer W, Heilig M, Zeier Z, Wahlestedt C: "Dysregulation of the histone demethylase KDM6B in alcohol dependence is associated with epigenetic regulation of inflammatory signaling pathways", 2019, *Addiction Biology* e12816.

[16] Khomtchouk KM, Joseph LI, **Khomtchouk BB**, Kouhi A, Massa S, Xia A, Koliesnik I, Pletzer D, Bollyky PL, Santa Maria PL: "Treatment with a neutrophil elastase inhibitor and ofloxacin reduces *P. aeruginosa* burden in a mouse model of chronic suppurative otitis media", 2021, *npj Biofilms and Microbiomes* 7(31). https://www.nature.com/articles/s41522-021-00200-z

[17] **Khomtchouk BB***, Nonner W: "Gaussian-distributed codon frequencies of genomes", 2019, *G3: Genes, Genomes, Genetics* 9(5): 1449-1456 (* indicates corresponding authorship).

[18] Ma WF, Hodonsky CJ, Turner AW, Wong D, Song Y, Mosquera JV, Ligay AV, Slenders L, Gancayco C, Pan H, Barrientos NB, Mai D, Alencar GF, Owsiany K, Owens G, Reilly MP, Li M, Pasterkamp G, Mokry M, van der Laan SW, **Khomtchouk BB***, Miller CL*: "Enhanced single-cell transcriptomics workflow reveals coronary artery disease cellular cross-talk and candidate drug targets", 2021, *Atherosclerosis* 340: 12-22. (* indicates co-corresponding/co-senior authorship) https://www.atherosclerosis-journal.com/article/S0021-9150(21)01464-7/fulltext

[19] Richards S: "New computational workflow for analyzing gene expression in single cells reveals potential drug targets to treat atherosclerosis." *University of Chicago Medicine Press.* https://www.uchicagomedicine.org/forefront/research-and-discoveries-articles/new-computational-workflow-for-analyzing-gene-expression-in-single-cells-reveals-potential-drug-targets-to-treat-atherosclerosis

[20] `https://www.thirdcoastcfar.org/2021/10/31/dr-bohdan-khomtchouk-selected-for-k12-award-on-cardiovascular-disease-computational-ge`

[21] **Khomtchouk BB***: "Codon usage bias levels predict taxonomic identity and genetic composition", 2021, *bioRxiv.* (* indicates corresponding authorship) https://doi.org/10.1101/2020.10.26.356295

[22] **Khomtchouk BB***: "WebAssembly enables low latency interoperable augmented and virtual reality software", 2021, *arXiv.* (* indicates corresponding authorship) https://arxiv.org/abs/2110.07128

[23] `https://docktherapeutics.com/`

[24] Nik AB, Ng HH, Sun P, Iacoviello F, Shearing PR, Bertazzo S, Mero D, **Khomtchouk BB***, Hutcheson JD: "Epidermal Growth Factor Receptor Inhibition Prevents Caveolin-1-dependent Calcifying Extracellular Vesicle Biogenesis", 2022, *bioRxiv*. (* indicates co-corresponding/co-senior authorship, responsible for study design, data analysis and interpretation) https://doi.org/10.1101/2021.11.08.467799

[25] **Khomtchouk BB***, Lee YS, Khan ML, Sun P, Mero D, Davidson MH: "Targeting the Cytoskeleton and Extracellular Matrix in Cardiovascular Disease Drug Discovery", 2022, *Expert Opinion on Drug Discovery*. 17(5): 443-460. (* indicates corresponding authorship) https://doi.org/10.1080/17460441.2022.2047645