



Albert Einstein College of Medicine

Montefiore

Department of Cell Biology

1300 Morris Park Avenue, Bronx, NY 10461

ARTHUR I. SKOULTCHI, Ph.D.

Judith and Burton P. Resnick Professor

University Chairman

PHONE: (718) 430-2169 / 2168 (lab)

FAX: (718) 430-8574

EMAIL: arthur.skoultchi@einsteinmed.org

September 29, 2021

Re: Emily Schwenger

Dear Members of the Selection Committee:

I am writing to you on behalf of Emily Schwenger's eligibility for the Google PhD Fellowship. Emily is a full-time Medical Scientist Training Program (MSTP) student and current MD/PhD candidate at the Albert Einstein College of Medicine with an expected graduation of 2025. Her primary mentor is Dr. Ulrich Steidl within the Department of Cell Biology, for which I serve as chair, and she is co-mentored by Dr. Aviv Bergman, Chair of the Department of Systems & Computational Biology.

Please consider this letter confirmation that Emily has successfully completed her graduate coursework as of Spring 2021. I can also confirm that Emily is not supported by any comparable industry award; further, she is not herself a Google employee, nor does she share a household with a Google employee. For these reasons, I can verify that Emily meets all eligibility requirements for application to the Google PhD Fellowship. Please do not hesitate to reach out should you have any questions or concerns.

Sincerely,

A handwritten signature in black ink that reads "Arthur I. Skoultchi".

Arthur I. Skoultchi
Professor and Chair

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Schwenger, Emily

ERA COMMONS USER NAME (credential, e.g., agency login): emilyschwenger

POSITION TITLE: M.D./Ph.D. student

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE <i>(if applicable)</i>	Completion Date MM/YYYY	FIELD OF STUDY
McGill University	B.Sc.	06/2014	Pharmacology

A. Personal Statement

As a self-driven jack-of-all-trades, I try to approach projects with a new and interesting lens. Within my short career as a scientist, I have worked in both academia and industry on numerous different projects, building a repertoire of skills that has culminated in oncology bioinformatics. During my undergraduate degree, I studied transcription elongation factors in yeast in the Tanny lab, resulting in my first publication. In 2015, despite being a first-generation college graduate, I was offered a position as a graduate scientist at the pharmaceutical company AstraZeneca and moved to Boston, Massachusetts. There, I led the crystallography and biophysics work for a neuroscience target validation project in industry, developed a novel oncology population to be used in physiologically-based pharmacokinetic (PBPK) modeling, and refined immune cell signatures for analysis of bulk tumor transcriptome data to better evaluate immuno-oncology treatment outcomes. As a computational scientist, I frequently design pipelines for various variant calling and functional analyses in the lab, including calling of small (single nucleotide variants and indels) and large variants (copy number variants and structural variants), as well as bulk RNA-seq, scRNA-seq and downstream pathway analyses. My long-term research interests include approaching high-dimensional genomic, transcriptomic, and proteomic datasets in new systems-based ways. Having worked in very early stages of drug development (crystallography) to Phase II clinical trials, I have developed a healthy amount of skepticism and reliable sense of what makes a strong drug candidate. Overall, I have become quite comfortable driving my own projects and learning techniques as I go. I am confident that this fellowship will provide me with a powerful platform on which I can work towards becoming an effective and independent biomedical researcher.

1. Mbogning J. et al. Functional interaction of Rpb1 and Spt5 C-terminal domains in co-transcriptional histone modification. *Nucleic Acids Res* **43**, 9766-9775, doi:10.1093/nar/gkv837 (2015).
2. Schwenger, E. et al. Harnessing Meta-analysis to Refine an Oncology Patient Population for Physiology-Based Pharmacokinetic Modeling of Drugs. *Clin Pharmacol Ther* **103**, 271-280, doi:10.1002/cpt.917 (2018).
3. Ueda K, Kumari R, Schwenger E, Wheat JC, Bohorquez O, Narayananagari SR, Taylor SJ, Carvajal LA, Pradhan K, Bartholdy B, Todorova TI, Goto H, Sun D, Chen J, Shan J, Song Y, Montagna C, Xiong S, Lozano G, Verma A, Steidl U. MDMX acts as a pervasive preleukemic-to- acute myeloid leukemia switch mechanism. *39(4)*, 529-547, doi: 10.1016/j.ccr.2021.02.006.
4. Piszcztowski RT, Townley RA, Sundaravel S, Schwenger E, Stein CM, Liu Y, Stanley P, Verma A, Zheng D, Seidel RD, Almo SC, Bülow HE*, Steidl U*. A glycan-based approach to cell characterization and isolation. (submission imminent)

5. Schwenger E, Steidl U. The evolution of subclonal complexity in MDS and AML: harnessing in silico approaches to reassess relationships between genetic events and disease. *Blood Cancer Discov.* **2**(3), 201-215, doi:10.1158/2643-3230.bcd-20-0219 (2021).

B. Positions and Honors

Positions and Employment

- 2012 – 2014 Undergraduate Student Research Assistant, McGill University
2015 – 2017 Graduate Scientist, AstraZeneca
2017 – 2019 M.D. Student, Albert Einstein College of Medicine
2017 – 2019 Student Assistant (Steidl Lab) , Albert Einstein College of Medicine
2019 – present M.D./Ph.D. Student, Albert Einstein College of Medicine

Other Experience and Professional Memberships

- 2017 - 2019 Member, American Medical Women's Association
2017 - 2019 Member, American Medical Association

Honors

- 2014 First Class Honors Pharmacology w/ Distinction, McGill University
2016 Top Poster, American Society for Clinical Pharmacology & Therapeutics
2017-2019 Service and Research Scholarship (\$30,000 annual), Albert Einstein College of Medicine

C. Contributions to Science

I got my research feet wet in an academic lab at McGill University in the Pharmacology & Therapeutics Department. With the guidance and mentorship of Dr. Jason Tanny, I investigated the *S. pombe* transcription elongation factor Spt5 for my undergraduate honors independent research project. RNA Polymerase II (RNAPII)-bound Spt5 is conserved across all three domains of life, and its aberrance can result in transcript elongation-related pathologies like cardiac hypertrophy and c-Myc-driven leukemias. I validated candidate genes found in a genome-wide screen for synthetic sick or lethal (SSL) interactions between variants of Spt5 mutated at its carboxy-terminal domain (CTD) and candidate mutants. My results illustrated compelling evidence for several novel roles for Spt5, including maintenance of boundaries between eu- and heterochromatin and shuttling of nascent transcripts through the nuclear pore complex. Spt5 strongly interacted with histone variant H2A.Z, as well as Swr1 and Swc2, subunits of H2A.Z's chaperone complex. Since all of the aforementioned proteins have established roles in chromatin boundary maintenance, this offers promise as an as-of-yet unknown link between active transcription elongation and chromatin boundary dynamics.

Upon completion of my undergraduate thesis, I opted to stay in Montreal for an additional four months to perform chromatin immunoprecipitation (ChIP) experiments and nucleosome mapping analyses. Venturing to corroborate my putative findings, I designed a series of primer pairs for a ChIP experiment in which DNA regions of interest, located near inverted repeats (IRCs) known to flank heterochromatin at centromeres, were used to illustrate heterochromatin spreading in Spt5 with a truncated CTD. Though further experiments must be performed for conclusive evidence, my work in the Tanny Lab laid the foundation both for a future publication and for a lifelong career in research.

In September of 2015, I began a two-year internship in the Graduate Program at AstraZeneca in Waltham, Massachusetts for the opportunity to contribute to diverse areas in biomedical research with direct applications to human health. As an industry scientist, I took on various technical and leadership responsibilities. Between September 2015 and May 2016, I led a target validation project aiming to characterize the interaction between the enzyme glutamine synthetase (GS) and the GABA_B receptor for future development of an Alzheimer's Disease and/or medial temporal lobe epilepsy therapeutic. With the guidance of Drs. Andrew Ferguson and Nichole O'Connell, I established a robust expression system for GS, determined that residues 890-904 on the GABA_B R2 (GBR2) C-terminus were most essential for binding to GS, and optimized a corresponding crystallization system. I also developed a highly sensitive enzymatic assay for measuring GS ATP turnover under the guidance of Dr. Nancy Su, which is of sufficient calibre for use in a high-throughput screen.

From May 2016 to January 2017, I took on a role as a clinical pharmacologist within AstraZeneca's Early Clinical Development Department. I conducted a meta-analysis comparing healthy volunteer and cancer patient

pharmacokinetic (PK) parameters, as well as analyzed clinical PK for the DNA damage response therapeutic AZD0156. With the guidance of Dr. Karthick Vishwanathan, I evaluated the impact of tumor-associated inflammation on enzyme and transporter expression in order to better inform PBPK extrapolation from healthy volunteers to cancer patients. Harnessing sensitivity analysis of the PBPK platform SimCYP®, I adjusted hepatic and intestinal CYP1A2, CYP2C19, and CYP3A4 abundance. Of the 11 substrates investigated, 7 displayed marked exposure differences >1.25-fold between oncology patients and HV. For caffeine, theophylline, midazolam, simvastatin, omeprazole, and a subset of oncology compounds, reducing CYP1A2, CYP2C19, and CYP3A4 abundances by 20, 33, and 30%, respectively, in a virtual oncology population effectively captured cancer patient concentration-time profiles. This oncology population enhances predictive capability and provides a more relevant characterization of cancer patient PK and was published as a manuscript in Clinical Pharmacology & Therapeutics, the leading journal on clinical pharmacology (2).

My final rotation at AstraZeneca was in the Oncology Bioinformatics Department working in the field of immuno-oncology under the supervision of Dr. Aleksandra Markovets. For my project, I refined immune cell gene signatures in order to more accurately enumerate cellular composition of the tumor microenvironment from bulk transcriptome data. Initially, I refined CD4+ T-cell, CD8+ T-cell, and macrophage signatures via an *a priori* approach, retaining only those genes that exhibit ideal marker coexpression and specificity using a scoring method adapted from the traditional Pearson correlation. I supplemented this approach by evaluating the capacity of mean log expression of these refined marker sets to quantify sample-to-sample relative abundance of immune cell subsets using simulations of known percentages of specific immune cells. Alongside this, I developed a refined leukocyte expression matrix to be used as a training set to deconvolute bulk tumor RNA-seq data, thereby providing absolute quantitation of cell subsets. Preliminary data illustrated that reducing CD4+ T-cell, CD8+ T-cell, and macrophage/monocyte signatures to a lesser number of more robustly expressed genes exhibiting ideal marker behavior provided more precise absolute and relative quantitation for RNA-seq data. Extrapolation of these findings into additional immune cell subtypes, as well as various dilutions of malignant cells, is ongoing.

Between 2017 and 2018, I provided bioinformatics support for a project investigating a mouse model for acute myeloid leukemia (AML) in the lab of Dr. Ulrich Steidl. Specifically, I conducted pathway and overall survival analysis based on RNA-seq data from a FLT3-internal tandem duplication (FLT3-ITD)/HLX knock-in mouse model within the lab in parallel with analysis of publicly available human AML datasets. High HLX expression has been shown to produce a median overall survival in human patients of only 8 months. Since the mechanism by which HLX overexpression contributes to AML pathogenesis is as-of-yet unknown, I aimed to elucidate the key pathways involved in producing functional cooperativity in the FLT3-ITD background. Utilizing gene set enrichment analysis, I detected enrichment in pathways critical to differentiation of hematopoietic cell lineages, oxidative phosphorylation, and inflammation, among many others. These findings have provided the basis for molecular and biochemical studies assessing the binding partners and functional relevance of HLX in the laboratory. Ultimately, the elucidation of mechanisms and pathways mediating the pre-leukemic to leukemic progression in high-risk AML patients, such as those with FLT3-ITD mutations and/or HLX overexpression, may permit the development of enhanced prognostication approaches, risk stratification methods, and targeted therapeutic intervention in the future.

Further, I analyzed a model harboring a heterozygous deletion of an enhancer of the PU.1 transcription factor in the context of a Msh2 mutation, resembling an expedited stem cell aging process. In the presence of the Msh2 knockout, the PU.1-mutated mice developed a full AML phenotype. We hypothesized that the PU.1/Msh2 mice would model acquisition of DNA lesions during the aging process, which leads to the initiation and development of AML in the mouse. My role was to perform quality assessment of sequencing data, genome alignment, detection of somatic mutations, annotation of mutation patterns, and subsequent clonal analysis. Importantly, I developed a pipeline for structural variant discovery using whole genome sequencing (WGS) data, including, but not limited to, large-scale translocations, insertions, and deletions. Utilizing existing databases and literature describing essential driver mutations in AML disease pathogenesis, I detected 10 candidate driver mutations, including, but not limited to, Npm1 transcript ablation, a bidirectional gene fusion of Lgals9 and Ppary, as well as structural variants of several proteins involved in packaging of nascent pre-mRNAs into heterogeneous ribonucleotide particles, transport of mRNAs into the cytoplasm, and protein ubiquitination.

Over the past year, I have designed a pipeline for identification of small variants in primary human samples, copy number variants, as well as scRNA-seq quantification and downstream pseudotime analysis in mouse hematopoietic stem and progenitor cells (HSPC). Pseudotime analysis incorporating Seurat, Scanpy, Phate, and Paga programs has allowed us to identify novel transcriptional regulators involved in differentiation of progenitors beginning at the stem cell level. Further, this platform has enabled us to identify putative

multipotent progenitor (MPP) and hematopoietic stem cell (HSC) subsets that remain to be functionally validated in the lab.

D. Additional Information: Research Support and/or Scholastic Performance

Scholastic Performance

YEAR	COURSE TITLE	GRADE
MCGILL UNIVERSITY		
2010	Introduction to Psychology	A
2010	Intro Physics – Mechanics	A
2010	General Chemistry 1	A
2010	AP Credit: Calculus AB	G
2010	Principles: Organismal Biology	A
2010	Cell and Molecular Biology	A
2010	General Chemistry 2	A-
2010	Intro Physics – Electromagnetism	A
2010	Calculus 2	A
2011	Principles of Statistics 1	A
2011	Physical Chemistry	B
2011	Mammalian Physiology	A-
2011	Intro Organic Chemistry 1	A
2011	Introductory Physiology Lab 1	B
2011	Molecular Biology	A
2011	Intro Organic Chemistry 2	A
2011	Cell Biology & Metabolism	A-
2011	Mammalian Physiology 2	A
2011	Introductory Physiology Lab 2	A
2011	Basic Genetics	A
2012	Cell and Molecular Laboratory	A
2012	Metabolic Biochemistry	A-
2012	Survey of English Literature	B+
2012	Drug Action	A
2012	Molecular Biology of Oncogenes	B+
2012	Macroeconomic Analysis & Applications	A
2012	Research Essay & Rhetoric	A-
2012	Drugs and Disease	A
2012	Principles of Toxicology	A
2012	Intro Organic Chemistry 3	A
2013	Foundations of Programming	A
2013	Microeconomic Analysis & Application	A
2013	Drug Discovery & Development	A-
2013	General Pharmacology 1	A
2013	Pharmacology Research Project	A-
2013	Organic Chemistry: Natural Products	A-
2013	Pharmacology Selected Topics	A
2013	General Pharmacology 2	A-
2013	Pharmacology Research Project	A-
2013	Deviance	A
ALBERT EINSTEIN COLLEGE OF MEDICINE		
2017	Clinical & Developmental Anatomy	P

YEAR	COURSE TITLE	GRADE
2017	Epid Pop Hlth & Ev Based Med 1	P
2017	Bioethics 1	P
2017	Histology & Cell Structure	P
2017	Intro to Clinical Med Year 1	P
2017	Molecular and Cellular Foundations of Med	P
2017	Disease Mechanisms	P
2017	Medical Spanish – Basic II	P
2017	Principles of Pharmacology	P
2017	General & Cardiovascular Physiology	P
2017	Renal Disease	P
2018	Bioethics 2	P
2018	Cardiovascular System	P
2018	Endocrine System	P
2018	Epid Pop Hlth & Ev Based Med 2	P
2018	GI/Liver System	P
2018	Hematology System	P
2018	Intro to Clinical Med Year 2	P
2018	Micro/Infectious Disease	P
2018	Musculoskeletal Disorders	P
2018	Nervous System/Human Behavior	P
2018	Parasitology & Global Medicine	P
2018	Pulmonary System	P
2018	Reproductive Sys & Human Sex	P
2019	Molecular Genetics	P
2019	Intro to Systems Biology	P
2020	Gene Expression: Beyond the Double Helix	P
2020	Computational Biology of Proteins	P
2020	Responsible Conduct of Research	P

USMLE Step 1 score: 256

Albert Einstein College of Medicine medical courses are graded P (pass) or F (fail). Passing is 65% or better.

Curriculum Vitae

Ulrich Steidl

Professor of Cell Biology, and of Medicine (Oncology)

Diane and Arthur B. Belfer Faculty Scholar in Cancer Research

Albert Einstein College of Medicine – Montefiore Health System, New York, U.S.A.

Professional Qualifications:

2000 PhD German Cancer Research Center (DKFZ), Heidelberg, Germany
2000 MD University of Heidelberg Medical School,

Current Positions:

2021- Co-Director, Einstein-Montefiore Blood Cancer Institute
2020- Associate Director for Basic Science, Albert Einstein Cancer Center (AECC)
2013- Associate Chair for Translational Research, Department of Oncology
2013- Scientific Director, Division of Hemato-Oncology
2011- Director, Stem Cell Isolation & XenoTx Facility, Institute for Stem Cell Research and Regen. Medicine

Last three positions:

2013-2016 Associate Professor, Albert Einstein College of Medicine
2008-2013 Assistant Professor, Albert Einstein College of Medicine
2003-2008 Fellow/Instructor in Medicine, Harvard Medical School, Boston, MA

Awards and Honors (selection):

2021 Outstanding Investigator Award, National Cancer Institute (NCI)
2021 President-elect, International Society for Experimental Hematology (ISEH)
2020 Saul R. Korey Award in Translational Medicine and Science
2020 Julius Marmor Mentorship Award for Outstanding Mentoring in Graduate Research
2019 Scholar Achievement Award, Leukemia & Lymphoma Society U.S.A.
2019 Scientific Editor, Blood Cancer Discovery (AACR)
2017 Research Mentoring Award, Einstein Graduate Programs in Biomedical Sciences
2017 Board of Directors, International Society for Experimental Hematology (ISEH)
2017 Senior Editor, Clinical Cancer Research (AACR)
2016 Elected Member, American Society for Clinical Investigation (ASCI)
2015 Elected Member, Interurban Clinical Club (Founder: Sir William Osler)
2014 Research Scholar, Leukemia & Lymphoma Society U.S.A.
2008 Howard Temin Award in Cancer Research, NIH/NCI
2005 Merit Award, American Society of Hematology

Contributions to Science (selection of senior author publications out of a total of 150):

Ueda K, et al., **Cancer Cell**. 2021; 39:529-547. PMID: 33667384

Wheat JC, et al., **Nature**. 2020; 583:431-436. PMID: 32581360

Chen J, et al., **Nat Med**. 2019; 25:103-110. PMID: 30510255

Kao C, et al., **Science Transl Med**. 2018; 10:eaas9563. PMID: 30209246

Mitchell K, et al., **J Exp Med**. 2018; 215:1709-1727. PMID: 29773641

Carvajal L, et al., **Science Transl Med**. 2018; 10:eaao3003. PMID: 29643228

Antony-Debre I, et al., **J Clin Invest**. 2017; 127:4297-4313. PMID: 29083320

Stanley RF, Piszcztowski R, et al., **J Exp Med**. 2017; 214:753-771. PMID: 28232469

Will B, Vogler T, et al., **Nat Med**. 2015; 21:1172-81. PMID: 26343801

Okoye-Okafor UC, et al., **Nat Chem Biol**. 2015; 11:878-86. PMID: 26436839

Schinke C, et al., **Blood**. 2015; 126:1118-1127. PMID: 25810490

Eckel AM, Stanley RF et al., **Blood**. 2015; 125:3144-3152. PMID: 26170031

Bartholdy B, Christopeit M, et al., **J Clin Invest**. 2014;124:1158-67. PMID: 24487588

Will B et al., **Nat Immunol**. 2013; 14:437-45. PMID: 23563689

Kawahara M, Eckel AM, Bartholdy B et al., **Cancer Cell**. 2012; 22:194-208. PMID: 22897850

Will B, et al., **Blood**. 2012; 120: 2076-86. PMID: 22753872

Barreyro L, et al., **Blood**. 2012; 120:1290-1298. PMID: 22723552

Roth M, et al., **Blood**. 2012; 120:386-394. PMID: 22627766

Complete list of publications: http://www.ncbi.nlm.nih.gov/pubmed?term=Steidl_U



Montefiore

October 1, 2021

Science at the heart of medicine

Albert Einstein Cancer Center

Gottesman Institute for Stem Cell and
Regenerative Medicine Research

Ulrich G. Steidl, M.D., Ph.D.

Professor of Cell Biology
Professor of Medicine (Oncology)

Diane and Arthur B. Belfer Faculty Scholar
in Cancer Research

Associate Director for Basic Science,
Albert Einstein Cancer Center

Co-Director, Einstein-Montefiore Blood
Cancer Institute

Associate Chair for Translational Research,
Department of Oncology

Chanin Bldg., Rm.# 601-605
1300 Morris Park Avenue
Bronx, NY 10461

Tel 718.430.3437 Fax 718.430.8574
ulrich.steidl@einsteinmed.org

RE: Google PhD fellowship mentor statement

Dear members of the selection committee,

I am writing in most enthusiastic support of the application of Emily Schwenger, an outstanding M.D./Ph.D. student in my laboratory, for the Google PhD Fellowship.

Emily has worked in my laboratory as a student assistant since 2017 and officially joined as an MSTP student in 2019. She has a very strong academic record, having graduated from McGill University with a 3.85 GPA and distinction, along with two years of industry experience at AstraZeneca. Emily frequently thinks outside of the box, owing to her diverse research experiences in both wet lab and computational biology.

Emily impressed me from the first day she joined our team; she is an extremely goal-oriented, creatively thinking, and hardworking student, quickly grasping novel concepts of leukemia pathogenesis and stem cell biology. Initially, she assisted a postdoc on a project analyzing whole-genome sequencing data obtained from a pre-leukemic to leukemia progression mouse model and quickly developed into the actual intellectual driver of the project. She not only refined and improved the data analysis strategy, but also began to implement changes in the experimental strategy based on a deep biological understanding of the project, despite doing this alongside her regular medical school curriculum.

Emily is currently performing clonal trajectory analysis of purified stem cell populations from longitudinal samples of patients with myelodysplastic syndrome. She has been pushing the boundaries of the project's translational potential by means of incorporating evolutionary and systems biology principles and taken it upon herself to study these and relevant mathematics topics in her spare time.

I have had the great pleasure of mentoring some of the most outstanding MSTP students in our program, and, based on my close interactions with Emily over the past 2.5 years, I would firmly place her in the top 5%. Her creative thinking, hard-working attitude, and outstanding team skills, represent a terrific (but rare) combination that bodes extremely well for her future. In order for healthcare to truly benefit from data science, more future physicians will need to truly embody a multidisciplinary mindset, and Emily is an ideal candidate to do so.

Sincerely,

A handwritten signature in blue ink that reads "Ulrich Steidl".

9/28/21

Dear members of the selection committee,

I am writing to express my advocacy for Emily Schwenger's candidacy in Google's PhD fellowship program.

I have known Emily since 2019 as a student, in an advisory capacity on her graduate committee, and, more recently, in the form of a formal co-mentorship. In each of these roles, I have continued to grow more and more impressed with her intellect, curiosity, and passion to ask the more difficult scientific questions and explore uncharted territory in her research. She is, at her core, deeply unsatisfied with simply following the status quo and continues to reinvent herself and cross disciplinary boundaries, qualities essential for true leadership in academia.

Emily has come to appreciate the necessity of formalizing theory with mathematical abstraction and has taken it upon herself to learn linear algebra, advanced calculus, and advanced statistical learning in her spare time, along with the standard course load and research commitments. In this regard, Emily is atypical: she views seemingly insurmountable challenges as opportunities to push oneself to the limit, thereby discovering new things that one never thought they were capable of. This was evident in systems biology lecture during which I challenged the students to write novel scientific literature, which was met with almost unanimous protest. Emily was one of the only students to embrace the challenge.

As her mentor, I have found that Emily embodies what I believe to be the essential qualities of a scientist. As a first-generation college graduate, she is a true pioneer, having graduated from McGill University with a 3.85 GPA and distinction, despite simultaneously working to pay off her loans and engaging in an honors research project. Prior to entering the MD program at Einstein, she obtained a highly competitive position as a graduate scientist at AstraZeneca, where she worked on projects spanning X-ray crystallography to pharmacokinetic modeling to immunotherapeutic biomarker identification. She was the sole trainee hired from Canada at AstraZeneca and is one of a very select few international students in the Medical Scientist Training Program (MSTP) at Einstein.

Emily is not interested in checking off any pre-defined boxes in pursuit of an established career; rather, she creates her own boxes. With mentorship and support from Google, I have no doubt that she will be an unstoppable force in efforts to not just generate data to improve medical care, but to truly *think* about the meaning and knowledge to be gained, as well as the important bioethical implications.

Sincerely,



Aviv Bergman, Ph.D.
Professor and Founding Chairman

A network-based approach to determine dysregulation of gene modules driving progression of complex myeloid malignancies

Background: Myelodysplastic syndromes (MDS) and acute myeloid leukemias (AML) are a heterogeneous constellation of malignant stem cell (SC)-derived diseases defined by inadequate hematopoiesis, consequent peripheral cytopenias, and uncontrolled proliferation of myeloid blasts. Despite the introduction of targeted, hypomethylating, and immunomodulatory therapies into treatment algorithms, clinical outcomes in older MDS patients have remained largely unchanged for six decades, with a median overall survival of less than two years in higher risk subtypes¹. Delaying progression of MDS to a higher-risk stage and further AML is one of the key challenges in the clinical management of patients with MDS and inclusive quantitative metrics for prognostication and early prediction are lacking. Gene coexpression networks (GCN) provide a theory-driven strategy to tackle both better prognostication and comprehension of MDS/AML pathogenesis.

Aim 1: Perform GCN analysis in bulk MDS and AML datasets and input module eigengenes as latent variables in classification algorithms predictive of clinical parameters.

While aberrations at the level of DNA, i.e. mutations and cytogenetic alterations, are the source of heritability and consequent clonal evolution, the genotype-phenotype map that determines the functional state of the cell is governed by a greater gene regulatory network. The complex connectivity between gene products is ignored by conventional differential expression analyses, which are designed to solely detect differences in mean expression and are the methodology of choice for standard clinical marker identification. **We hypothesize that leukemogenesis is often not the result of up- or down-regulated expression of a select few gene products, but rather the more insidious breakdown of highly interconnected regulatory networks involving thousands of genes.** To this end, we will construct GCNs as a systems-based framework to further elucidate functionally relevant transcriptional modules driving MDS pathogenesis, and, in turn, associate these modules to phenotypic metadata, such as mutations, cytogenetics, peripheral blood counts, and overall survival.

We aim to identify common cryptic transcriptional patterns heralding disease progression across patients with heterogeneous mutational profiles, thereby accounting for previously unexplained variance between patients within identical prognostic categories, as well as commonalities between patients in opposing categories. We will harness publicly available MDS and AML bulk RNA datasets to identify transcriptional modules characteristic of disease subtypes and build network-based classifiers.

Aim 2: Assess module-specific transcriptional coordination in malignant hematopoietic stem and progenitor cells from early to later disease progression at the single-cell level.

We propose that a more complete elucidation of MDS/AML pathogenesis requires assessment of network connectivity in longitudinal samples that include early disease stages, including pre-leukemic syndromes such as MDS. Accordingly, we will analyze longitudinal samples from initial low-risk subtypes of MDS spanning early disease up until progression to secondary AML. Due to heterogeneity of both cell types and clonal populations within the bone marrow, we will leverage single-cell RNA-seq data to evaluate gene expression modularity on a per-cell and per-cluster basis. Upon identification of modules associated with important phenotypes, we will develop and optimize covariance metrics that are representative of functionally relevant gene regulatory dysregulation.

Specifically, we will: (1) characterize single cells based on marker expression, copy number profile, mutations, and pseudotime; (2) perform weighted gene coexpression network analysis (WGCNA); (3) correlate module eigengenes with clinical characteristics as well as those outlined in (1); and (4) cross-validate single-cell derived signatures in the aforementioned bulk datasets. This complexity and connectivity-aware approach has the potential to revolutionize MDS/AML patient management and prognostication strategies.

Background

Myelodysplastic syndromes (MDS) are a group of notoriously heterogeneous disorders defined by ineffective blood cell production due to both quantitative changes—or peripheral cytopenias—and qualitative dysplastic changes affecting red cell, granulocyte, and platelet precursors in the peripheral blood and bone marrow². Contingent on the presence of molecular and cytogenetic aberrations within distinct hematopoietic lineages, clinical manifestations of MDS can include anemia, neutropenia, and thrombocytopenia, presenting as fatigue, infection, and bleeding, respectively^{3,4}. Interestingly, blood cells are frequently “stuck” in the form of subfunctional, poorly differentiated blasts, defined as the committed stage prior to terminal differentiation, which can be observed in both the bone marrow and peripheral blood². Despite the introduction of hypomethylating, immunomodulatory, and iron-chelating agents into treatment algorithms, clinical outcomes in older MDS patients have remained largely unchanged for six decades, with a median overall survival of less than two years in higher risk subtypes¹. Notably, precision options in MDS are usually isolated to uncommon genetic subtypes and have seen extremely limited success⁵.

The significant variation in both clinical presentation and likelihood of transformation to higher risk stages of MDS and subsequent bone marrow failure has necessitated the iterative development of strict criteria by which clinicians can stratify patients into prognostic categories⁶. These criteria—currently the IPSS-R—predominantly take into account clinical parameters (blast percentages and cytopenias), bone marrow cytogenetics, and, more recently, hotspot mutations in a select few “driver” genes. However, parameters incorporated into prediction algorithms such as cytogenetics detectable at the bulk level are often associated with more advanced disease, neglecting detection of earlier, disease-initiating cells⁷. Further, mutations are input into prognostic algorithms based upon a binary presence or absence, neglecting information on the proportion of cells harboring the mutation of interest, as well as any information on the context in which the mutations are occurring. Overall, the current point-based system is perhaps an oversimplification of the highly complex nature of myeloid malignancies, and a strong consensus on precisely how to derive an inclusive quantitative metric for prognostication is lacking.

Historically, translational science has distilled complex diseases down to a handful of biomarkers present at the clonal level, largely due to financial reasons and ease of interpretability. This strategy has been successful for certain myeloid malignancy subtypes such as *BCR-ABL*-driven chronic myeloid leukemia (CML) and *PML-RARA*-driven acute promyelocytic leukemia (APL), in which neomorphic gain-of-function transgenes drive disease in a clear causal manner; however, it has failed to produce effective treatment algorithms for the vast majority of patients suffering from myeloid malignancies and outcomes remain poor^{8,9}. Unfortunately, patients frequently exhibit vast *inter-* and *intra*-patient heterogeneity and do not fit neatly into their respective prognostic box, lacking any singular clear

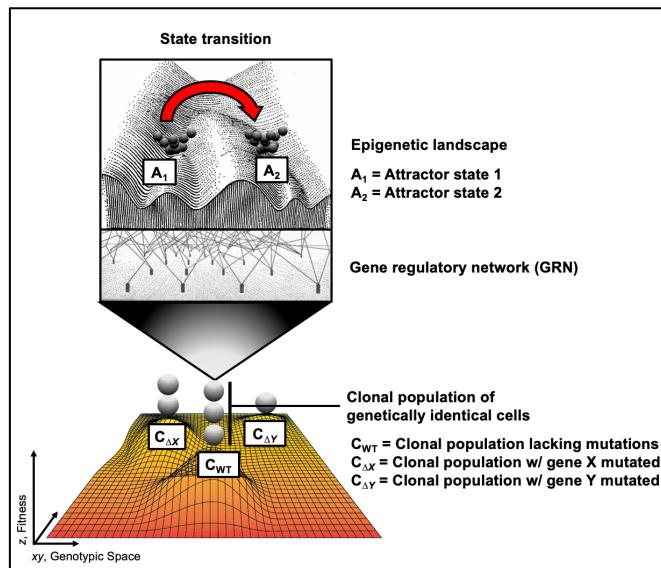


Figure 1: State transitions are governed by the GRN. In order for a cell to change its phenotype, the cell must transition from an original phenotype that is evolutionarily adaptive (and, therefore, very stable, e.g. attractor state 1) to a new stable phenotype (e.g. attractor state 2). This transformation is driven by the gene regulatory network (GRN), or the overarching architecture between all gene products that orchestrates gene expression profiles and, in turn, determines the final cellular “state” or phenotype.

disease driver. In particular, patients suffering from differentiation as their predominant disease feature and often die from bone marrow insufficiency rather than uncontrolled proliferation of a dominant clone. For these patients, a reductionist approach is unlikely to capture the highly nuanced and multifactorial nature of their disease pathogenesis, and innovative approaches are overdue.

To further necessitate the need to revise the somatic mutation model for leukemogenesis, next-generation sequencing studies have brought to light the phenomenon known as clonal hematopoiesis (CH), or the age-related development of subclones within the pool of healthy hematopoietic stem cells (HSCs) in the absence of any disease phenotype. Evidently, mutations such as *DNTM3A*, *TET2*, *ASXL1* and even *TP53*, are

frequently present in individuals greater than 70 years of age and compatible with healthy hematopoiesis^{10–13}. While the existence of CH challenges the orthodoxy of attributing malignant transformation to stereotypical sequences of driver mutations, it also raises new and interesting questions. Borrowing from systems biology, ‘mutations... are components of a greater system, or gene regulatory network (GRN), which orchestrates the gene expression profile, and, in turn, determines the final cellular “state” or phenotype that the cell will adopt’ (**Figure 1**)¹⁴. Importantly, the link between genotype and phenotype is highly dynamic and non-linear in nature: while a mutation in an oncogene may very well have “driving” effects, those effects are subject to a slew of buffering mechanisms both within and surrounding a cell that favor a “wild-type” outcome and vary with existing genetic, non-genetic, and environmental variation’. More inclusive algorithms that account for these subtleties are in order.

Since phenotype is more closely linked to mRNA expression than genotype, a sensible strategy is to leverage transcriptomic data to better understand how mutations ultimately give rise to clinical outcomes. With sequencing costs exponentially declining, generation of high-dimensional datasets has become standard practice in research laboratories on a global scale, and whole transcriptome mRNA expression profiling is clinically feasible. Further, high-resolution technologies like single-cell RNA sequencing (scRNA-seq) are increasingly affordable and outsourced, granting potential for broad standardization and clinical utility. Nevertheless, patient classification remains restricted to a small number of biomarkers typically ascertained by differences in mean expression, with the focus on *which* few genes are up- or downregulated rather than *how* the larger collection of genes interact with one another to produce stable transcriptional states. We argue that complexity must be embraced in order to be outsmarted.

Therefore, we hypothesize that MDS/AML pathogenesis is often driven by a large number of insidious, incremental changes that can be best captured by construction of gene coexpression networks and assessment of transcriptional module coordination.

We aim to identify common cryptic transcriptional patterns heralding disease progression across patients with heterogeneous mutational profiles, thereby accounting for previously unexplained variance between patients within identical prognostic categories, as well as commonalities between patients in opposing categories. To this end, we will harness publicly available MDS and AML bulk RNA datasets to identify transcriptional modules characteristic of disease subtypes and build network-based classifiers (**Figure 2**). Further, we propose that a more complete elucidation of MDS/AML pathogenesis requires assessment of network connectivity in longitudinal samples that include early disease stages, including pre-leukemic syndromes such as MDS. Accordingly, we will analyze longitudinal samples from initial low-risk subtypes of MDS spanning early disease up until progression to secondary AML. Due to heterogeneity of both cell types and clonal populations within the bone marrow, we will leverage scRNA-seq data to evaluate gene expression modularity on a per-cell and per-cluster basis. Upon identification of modules associated with important phenotypes, we will develop and optimize covariance metrics that are representative of functionally relevant gene regulatory dysregulation.

Importantly, our approach is founded in systems biology theory and has the capacity to not only improve prognostic accuracy in an empirical sense, but also contribute to a deeper understanding of the multifactorial interplay of events leading up to disease. While I have used myeloid malignancies as a proof-of-concept due to their well-described heterogeneity and rich phenotypic data accessible via peripheral blood, the scope of this approach spans the full spectrum of complex diseases. Additionally, our strategy has potential for clinical implementation, for example as an open-access calculator or web application, enabling ease of interpretation for patients and physicians alike.

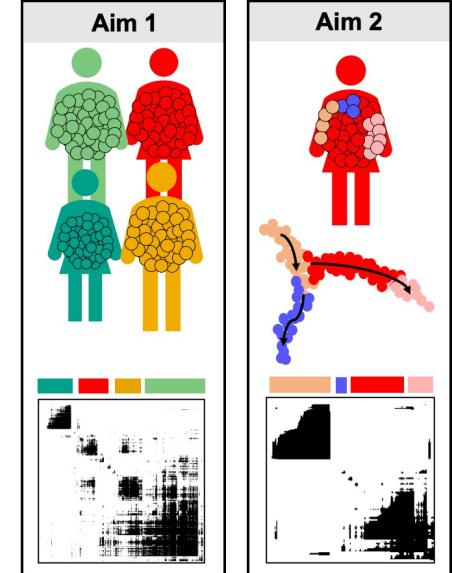


Figure 2: Summary of aims. In Aim 1, we will perform GCN analysis in bulk MDS and AML datasets to resolve GCNs associated with interpatient heterogeneity. In Aim 2, we will assess module-specific transcriptional coordination in malignant hematopoietic stem and progenitor cells from early to later disease progression at the single-cell level.

Research Strategy

Aim 1: Perform GCN analysis in bulk MDS and AML datasets and input module eigengenes as latent variables in classification algorithms predictive of clinical parameters.

Rationale

While aberrations at the level of DNA, i.e. mutations and cytogenetics alterations, are the source of heritability and consequent clonal evolution, the genotype-phenotype map that determines the functional state of the cell is governed by a greater gene regulatory network. The complex connectivity between gene products are ignored by conventional differential expression analyses, which are designed to solely detect differences in mean expression and are the methodology of choice for standard clinical marker identification. **We hypothesize that leukemogenesis is often not the result of up- or down-regulated expression of a select few gene products, but rather the more insidious breakdown of highly interconnected regulatory networks involving thousands of genes.** To this end, we will construct gene co-expression networks (GCNs) as a systems-based framework to further elucidate functionally relevant transcriptional modules driving MDS pathogenesis and associate these modules to phenotypic metadata, such as mutations, cytogenetics, peripheral blood counts, and overall survival.

Approach

Aim 1.1: Identify variable and highly connected genes and construct GCNs representative of inter-patient variability.

To assess topology, we will take a systems biology approach and infer features of the underlying gene regulatory network by means of weighted gene coexpression analysis (WGCNA)¹⁵. Topological network analysis is an intuitive and biologically relevant means to understand the interconnectivity of gene products and how perturbation can result in disease. Due to the abundance of healthy control samples to serve as a comparison, we will use the beatAML dataset for initial identification of transcriptional modules and reserve the TCGA LAML dataset for cross-validation of key findings^{16,17}. WGCNA of the bulk transcriptome permits assessment of transcriptional modularity at the level of a population of patients, revealing cryptic inter-patient biology

undetectable by conventional differential expression analysis. A more exhaustive analysis of intra-patient biology at the single-cell level is presented in **Aim 2**.

Hierarchical clustering of each sample will be performed, and samples branching above a threshold cut height will be considered outliers and removed from downstream analysis to prevent skewing. Further, we will perform an initial feature-based filtration step to eliminate uninformative genes. Specifically, we will filter gene products based on greater variance in patients than that observed in healthy controls to focus on gene modules associated with disease by taking the difference in standard deviation ($\Delta\sigma$) (**Figure 3**). Genes with greater standard deviation in the patient cohort than the healthy control (red region) will be kept for downstream analysis; additionally, we will perform a preliminary assessment of connectivity using WGCNA's softConnectivity function, keeping only genes

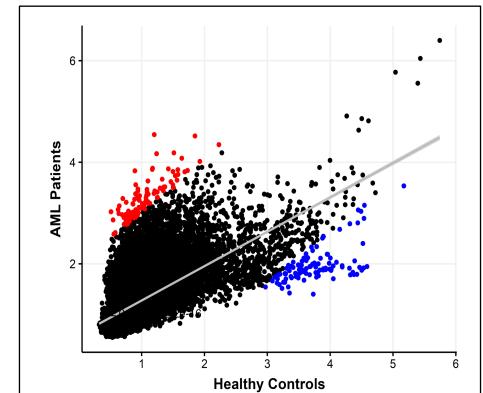


Figure 3: Assessment of gene-level standard deviations across patients. Top 100 genes with the greatest positive difference in standard deviation are shown in red and greatest negative difference shown in blue. The threshold $\Delta\sigma$ will be optimized with iterative clustering.

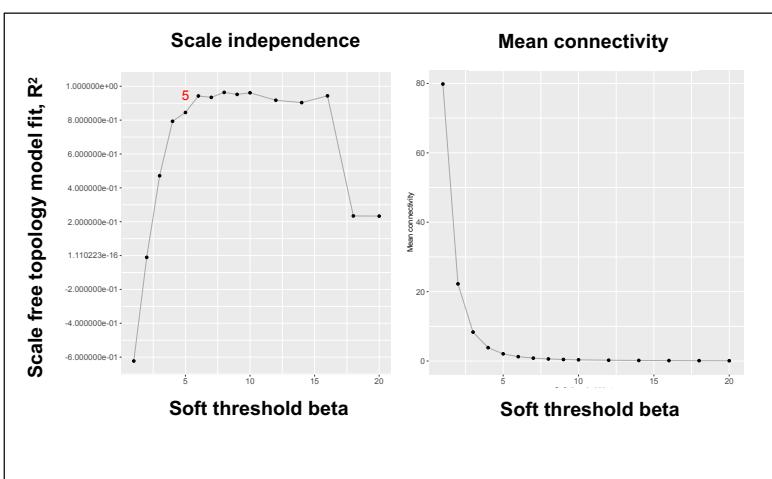


Figure 4: Beta selection for preliminary WGCNA of the beatAML dataset and associated mean connectivity. The lowest power at which the scale-free topology index curve plateaus with a reasonable fit >0.8 is chosen. We will perform a preliminary assessment of connectivity using WGCNA's softConnectivity function, keeping only genes

above threshold connectivity¹⁵. A threshold based on weighted $\Delta\sigma$ and connectivity will be optimized via iterative module clustering and assessment of intramodule collinearity. Beta selection will depend upon the lowest power at which the scale-free topology model curve flattens (**Figure 4**).

Due to strong biological rationale supporting non-linearity in expression networks, we will perform DCOL, a rank-based method compatible with non-linearity, in parallel to WGCNA which utilizes the linear biweight midcorrelation. The rank statistic calculated by DCOL has been shown to exhibit desirable statistical properties, such as asymptotic normality and rigorously defined p-values and FDRs, making it both computationally efficient and amenable to statistic inference necessary for our purposes¹⁸.

Aim 1.2: Evaluate network characteristics and correlate key gene modules with phenotypic parameters using module eigengenes and expression of intramodular hub genes.

To assess whether our results are compatible with properties of robust small-world or scale-free networks, we will perform linear regression on $p(k)$, the frequency distribution of the connectivity, k , and $\log_{10}(k)$. An $R^2 > 0.9$ and negative slope will be accepted as validation that network connectivity does indeed follow a power law, indicating that the network is comprised of a few densely connected hub genes amongst otherwise sparsely connected genes. In addition to whole network connectivity (degree), we will assess intramodular connectivity, topological overlap, and density between networks across patient subsets. Node strength and betweenness centrality will be used for hub gene identification. Node strength, s_i , is defined as the sum of the weights of adjacent edges:

$$s_i = \sum_{j=1}^N a_{ij} w_{ij} \quad (\text{Equation 1})$$

With a_{ij} and w_{ij} representing adjacency and weight matrices between nodes i and j , respectively, while betweenness centrality is defined as:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (\text{Equation 2})$$

In which σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v ^{19,20}.

As an initial assessment of feasibility of this approach in myeloid malignancies, preliminary results using the beatAML dataset illustrated a module largely anchored by Homeobox (hox) hub genes (**Figure 5**). Hox genes are widely known to be both co-localized chromosomally and co-regulated in development, hence their modular cohabitation, and their dysregulation is appreciated to be a hallmarks of aggressive leukemia. While this is only

the first iteration of our analysis, this finding validates our approach, given the notion that hub genes are of particular biological significance^{21,22}.

Upon network construction, we will evaluate associations between identified gene modules and key phenotypic characteristics. Specifically, module eigengenes, defined as the first principal component, and covariances of intramodular hub genes will be used to predict overall survival, blast morphology, cytogenetics, mutations, hemoglobin concentration, platelet count, and white blood cell count (WBC), both as standalone independent variables and as latent variables in more comprehensive logistic regression and support vector machine classifiers. Further, owing to the functional significance of evolutionary conservation, we will also inform our model with evolutionary ages of hub genes in

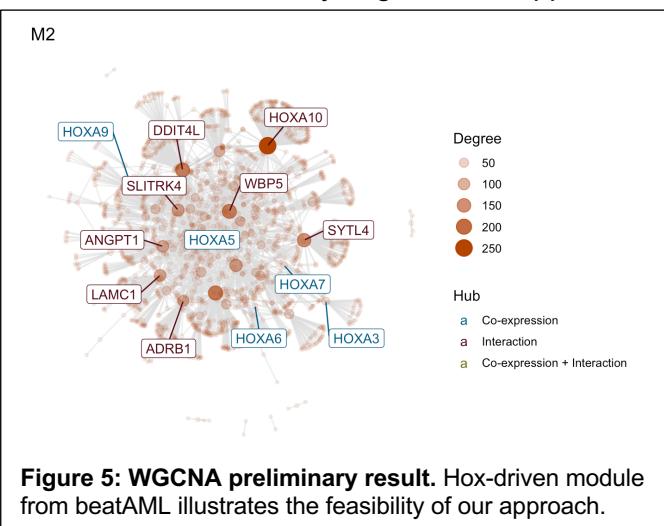


Figure 5: WGCNA preliminary result. Hox-driven module from beatAML illustrates the feasibility of our approach.

order to reduce spurious gene correlations impacting our analysis²³. This would be the first evolutionary biology-driven metric to be used in prognostication of patients with hematological malignancies.

Aim 2: Assess module-specific transcriptional coordination in malignant hematopoietic stem and progenitor cells from early to later disease progression at the single-cell level.

Rationale

Since bulk RNA-seq populations contain heterogeneous mixtures of cells, biological modules present in subpopulations of cells or lower intensity signals may be obscured by clones present at greater proportions or prevalent mature blood cells. While the bulk approach described above can elucidate transcriptional patterns present in dominant clones of AML patients with advanced disease, it is likely to neglect key events driving disease initiation and earlier progression. scRNA-seq permits distinction of heterogeneous clusters of cells present in the bone marrow that may harbor divergent mutations, copy number alterations, and, in turn, transcriptional profiles. Further, the generation of thousands of observations per patient allows for gene coexpression network assessment across different clusters and pseudotime, granting a closer look at how modules are evolving with malignant progression and various stages of differentiation. Our lab has previously used FACS and scRNA-seq to characterize megakaryocyte and erythroid trajectories in HSPCs, as well as targeted sequencing of stem cells to evaluate clonal evolution²⁴. While previous studies have utilized WGCNA to identify modules associated with drug resistance and prognostication in solid tumors, this study will be the first to evaluate GCNs in myeloid malignancies at the single-cell level, spanning early MDS, remission, and progression to secondary AML^{25,26}. Overall, high resolution scRNA-seq is likely to reveal key trends in gene connectivity otherwise imperceptible in bulk RNA-seq samples.

Table 1: Patient Characteristics

Demographic	Mean (range) or occurrence (n=13)
Sex	
Female	10
Male	3
Age at diagnosis (years)	71 (46 – 85)
Treatment	
ESA	7
Lenalidomide	4
None	2
Karyotype	
Normal	6
del(5q)	5
del(13q)	1
t(1;3)	1
Blast count (%)	
Diagnosis	3.9 (1 – 10)
Relapse	6.7 (1 – 19)

Aim 2.1: Characterize single cells based on marker expression, copy number profile, mutations, and pseudotime.

To achieve a more exhaustive evaluation of MDS/AML evolution, we will focus on patients with available samples from both stable and advanced disease timepoints and map expression profiles to known cell types in a healthy control (**Table 1**). To enrich for malignant and pre-malignant stem and progenitor cells (CD34+Lin-), we will perform an initial fluorescence activated cell sorting (FACS) step, thereby depleting mature cells and focusing on the stem-like cells from which MDS and AML originate. Library preparation will be prepared using the 10X Genomics Chromium V3 3' platform, sequenced to a target of 200,000 reads/cell for small variant detection, aligned, and processed as previously described²⁷. To ensure rigorous downstream analysis, only high-quality cells will be included, and cells with low feature counts, high mitochondrial reads, and count distributions resembling doublets filtered out. Seurat will then be used to perform principal components analysis (PCA), generate a k-nearest neighbor (k-NN) graph, and perform Louvain clustering on log-normalized and scaled data, as previously described²⁸. Uniform manifold approximations and projections (UMAP) will be used for data visualization, and cluster identities predicted based upon marker gene expression. Importantly, copy number analysis will be performed using inferCNV and used to distinguish malignant clusters, as well as clusters harboring known disease-driving CNAs present at diagnosis, such as del(5q)²⁹.

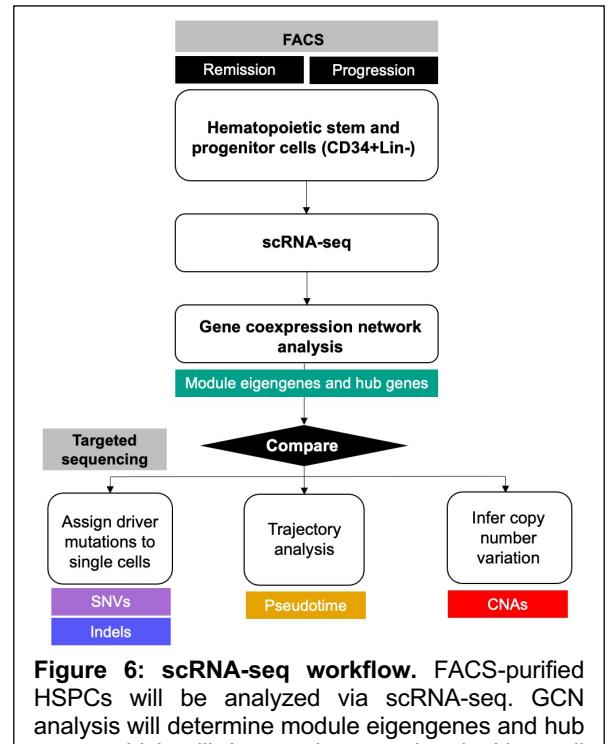


Figure 6: scRNA-seq workflow. FACS-purified HSPCs will be analyzed via scRNA-seq. GCN analysis will determine module eigengenes and hub genes which will, in turn, be associated with small variants, pseudotime, and CNAs.

In parallel to scRNA-seq, we will perform targeted sequencing using genes featured in the Illumina TruSight Myeloid panel. Variant calling will be performed using GATK's best practices workflow, and variants will be filtered based on base quality, mapping quality, variant allele frequency, and a panel-of-normals. Funcotator will be used for variant annotation, and COSMIC, ClinVar, gnomAD, and dbSNP databases will be harnessed in order to generate a concise list of high-confidence variants. Variants passing filtration will then be assigned to single cells based upon UMI barcodes using VarTrix²⁷.

Finally, we will analyze differentiation trajectories in malignant and pre-malignant HSPCs using PHATE and PAGA^{30,31}. All of the aforementioned characteristics, including cell type, copy number information, mutations, pseudotime, and real-time due to the longitudinal nature of our data, provide rich annotations with which module eigenvalues and hub genes can be associated in downstream analysis (**Aim 2.2**).

Aim 2.2: Perform WGCNA at the single-cell level

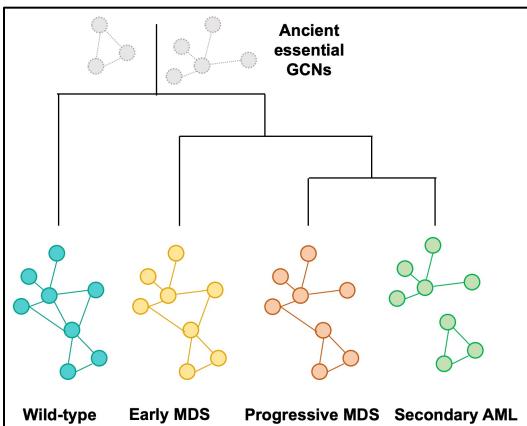


Figure 7: Simplified phylogenetic model of breakdown of gene connectivity with disease progression. The wild-type leaf represents the complete GCN, which undergoes progressive destruction of connectivity until secondary AML. Certain GCNs are maintained even at the AML stage and inferred to be putatively essential ancient modules.

As in **Aim 1.1**, we will identify genes that exhibit greater variability in malignant clusters, as compared to their healthy counterparts. Due to the tendency of sparse single-cell data to exhibit zero-inflated negative binomial distributions, we will perform an additional filtration step based upon bimodality via the R package SIBER, which utilizes a normal mixture model to fit expression data³². Bimodal distributions often represent interesting biology but are frequently missed by conventional measures of central tendency, such as the mean.

WGCNA and subsequent evaluation of network properties will be performed as described in **Aims 1.1** and **1.2**. We will measure transcriptional coordination in different cell types via multivariate dependency of gene subsets between clusters and across both time and pseudotime, and associate these modules to cell type, mutations, and copy number alterations. To validate our approach, we will benchmark novel discovery of module members and hub genes against genes identified in classic mean-based differential expression analysis.

Comparison of longitudinal targeted DNA sequencing and single-cell RNA sequencing of stable vs. relapsed MDS patients will allow us to parse out whether key phenotypic transitions can be explained by mutations alone or are further accounted for by GCN analysis. By virtue of sequencing longitudinal samples, we hope to identify transcriptional modules associated with early disease pathogenesis, as well as transition from remission to relapse and disease progression. Finally, we will cross-validate single-cell-resolved eigenvalues and hub gene expression in bulk RNA-seq patient data and model correlations with overall survival and disease-relevant clinical parameters. We strive to expand this model to include key characteristics of the gene regulatory network critical in fueling high plasticity resistant subclones upon perturbations with treatment, and, ultimately, put forth network-driven detection and management strategies to impede rapid evolution of malignant populations present at relapse.

References

1. H, K. Acute myeloid leukemia--major progress over four decades and glimpses into the future. *Am. J. Hematol.* **91**, 131–145 (2016).
2. Della Porta, M. G. et al. Minimal morphological criteria for defining bone marrow dysplasia: a basis for clinical implementation of WHO classification of myelodysplastic syndromes. *Leukemia* **29**, 66–75 (2015).
3. Kennedy, A. L. & Shimamura, A. Genetic predisposition to MDS: Clinical features and clonal evolution. *Blood* **133**, 1071–1085 (2019).

Google PhD Fellowship Research Proposal, Emily Schwenger

4. Nimer, S. D. Myelodysplastic syndromes. *Blood* **111**, 4841–4851 (2008).
5. Tehranchi, R. et al. Persistent Malignant Stem Cells in del(5q) Myelodysplasia in Remission. *N. Engl. J. Med.* **363**, 1025–1037 (2010).
6. Greenberg, P. L. et al. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood* **120**, 2454–2465 (2012).
7. Mossner, M. et al. Mutational hierarchies in myelodysplastic syndromes dynamically adapt and evolve upon therapy response and failure. *Blood* **128**, 1246–1249 (2016).
8. Salesse, S. & Verfaillie, C. M. BCR/ABL: from molecular mechanisms of leukemia induction to treatment of chronic myelogenous leukemia. *Oncogene* **2002** *21*, 8547–8559 (2002).
9. Zhang, X., Sun, J., Yu, W. & Jin, J. Current views on the genetic landscape and management of variant acute promyelocytic leukemia. *Biomark. Res.* **2021** *9*, 1–15 (2021).
10. Busque, L. et al. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.* **44**, 1179–1181 (2012).
11. Genovese, G. et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
12. Jaiswal, S. et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
13. Xie, M. et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).
14. Schwenger, E. & Steidl, U. An Evolutionary Approach to Clonally Complex Hematologic Disorders. *Blood Cancer Discov.* **2**, 201–215 (2021).
15. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
16. Tyner, J. W. et al. Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**, 526–531 (2018).
17. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
18. Liu, H. et al. Nonlinear Network Reconstruction from Gene Expression Data Using Marginal Dependencies Measured by DCOL. *PLoS One* **11**, e0158247 (2016).
19. Barrat, A., Barthélémy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci.* **101**, 3747–3752 (2004).
20. Kirkley, A., Barbosa, H., Barthelemy, M. & Ghoshal, G. From the betweenness centrality in street networks to structural invariants in random planar graphs. *Nat. Commun.* **2018** *9*, 1–12 (2018).
21. Alharbi, R. A., Pettengell, R., Pandha, H. S. & Morgan, R. The role of HOX genes in normal hematopoiesis and acute leukemia. *Leuk.* **2013** *275* **27**, 1000–1008 (2012).
22. Zhang, H. et al. Functional interrogation of HOXA9 regulome in MLLr leukemia via reporter-based CRISPR/Cas9 screen. *Elife* **9**, 1–30 (2020).
23. Tong, Y.-B. et al. GenOrigin: A comprehensive protein-coding gene origination database on the evolutionary timescale of life. *J. Genet. Genomics* (2021). doi:10.1016/J.JGG.2021.03.018
24. Chen, J. et al. Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. *Nature Medicine* **25**, 103–110 (2019).
25. X, C., L, H., Y, W., W, S. & C, Y. Single Cell Gene Co-Expression Network Reveals FECH/CROT Signature as a Prognostic Marker. *Cells* **8**, 698 (2019).
26. Wu, H. et al. Single-cell Transcriptome Analyses Reveal Molecular Signals to Intrinsic and Acquired Paclitaxel Resistance in Esophageal Squamous Cancer Cells. *Cancer Lett.* **420**, 156–167 (2018).
27. Petti, A. A. et al. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. *Nat. Commun.* **2019** *10*, 1–16 (2019).
28. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
29. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (80-).* **344**, 1396–1401 (2014).
30. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **2019** *3712* **37**, 1482–1492 (2019).
31. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **2019** *20*, 1–9 (2019).
32. Tong, P., Chen, Y., Su, X. & Coombes, K. R. SIBER: systematic identification of bimodally expressed genes using RNAseq data. *Bioinformatics* **29**, 605–613 (2013).

Impact essay

Emerging clonal complexity has brought into question the way in which we perceive and, in turn, treat malignancies of the hematopoietic system. Substantial pharmaceutical investment in targeted therapies are yet to produce appreciable improvements in outcomes in MDS/AML patients, which begs for innovative approaches that defy existing dogmas. This project borrows from perspectives rooted in evolutionary biology to reframe pathogenesis of hematologic disorders as dynamic processes involving complex interplays of genetic and non-genetic subclones and the tissue microenvironment in which they reside.

As a former scientist at AstraZeneca who worked in structural biology, pharmacokinetic/dynamic (PKPD) modeling, and bioinformatics, I observed that the mantra of experimental design is to reduce every system down to its most basic component parts: a pure drug and its receptor hybridized to a chip, an IC₅₀ determined in an isogenic cell line, a “one-size-fits-all” PKPD model in which a kinase inhibitor binds with 100% specificity to its target kinase. While these approaches are important, they overlook the high dimensional nature of cellular systems and encourage a somewhat myopic view of vastly complex diseases. Importantly, this has resulted in an underestimation of tumor evolvability and an overall dogmatic view of cancer biology in which a tumor can be distilled down into its dominant clone’s mutated receptors, which, when inhibited, will lead to a cure.

While interrogating dynamic systems with high dimensionality and both inter- and intraindividual variation poses a major challenge, longitudinal high-throughput single-cell RNA sequencing allows us to observe transcriptional reprogramming in real patient data. Insights derived from direct observation can then be queried on larger scales and in the face of various genomic and environmental insults by means of evolutionary biology models and computational simulations. These approaches will allow us to gain a deeper comprehension of phenomena that cannot be experimentally observed, notably the early mechanisms that ensure that rapid clonal expansion does *not* happen for most of life despite the constant onslaught of genotoxic insults. Ultimately, we must embrace complexity if we are to outsmart it and see tangible improvements in outcomes for MDS/AML patients with clonally complex diseases.

Leadership essay

During medical school, I adopted the position of Education Coordinator for Space in Prisons for Health Education for Re-entry (SPHERE), for which I designed and facilitated health-related workshops at a correctional facility. The club was struggling to stay on its feet, due in part to increasing pressure felt by students to perform on exams, and I found myself spending quite a few weekends at Edgecombe Correctional Facility in Washington Heights due to a lack of volunteers.

Running the workshops was, admittedly, intimidating at first. There were some weekends when I could barely get any control over the room and felt silly talking about seemingly irrelevant details on the inner workings of the endocrine system to a group of adult men; other weekends, the correctional officers would forget to announce that the workshop was happening so no one would show up.

Eventually, I began to realize what the participants truly needed out of the workshops: someone to listen. I ditched the health pamphlets that I had printed in favor of the stories and familiar faces in the room. There was one material, however, I did use as guidance: a guided meditation. This became their favorite part of every session--to be taken away to some place peaceful, a place where they were simply human beings, not inmates or criminals. When I stopped seeing a regular participant in the sessions, I'd like to think that they were enjoying their well-earned freedom, perhaps even using meditation as a tool when times get tough.

When I look at myself in the mirror, I certainly do not see a prototypical strong leader; rather, I see a flawed first-generation college graduate who frequently makes mistakes. Who knows--maybe if I'd grown up in a different neighborhood or with different colored skin I'd be participating in those workshops rather than leading them. But when I think about what true leadership really means, I think of Obama, MLK, Gandhi, all of whom share qualities of radical empathy, courage in the face of crisis, and rising to the occasion to meet the needs of their people. By that definition, I could lead.

MAIN MENU	Personal	Student	Employee	Applicant	Financial Aid/Awards	

McGill

UNOFFICIAL Transcript

* An asterisk appears next to the credit value of courses not counted in the total credits earned.

Remarks column:

- I - Course is included in credits and included in the GPA.
- E - Course is excluded from credits and excluded from the GPA.
- A - Course is excluded from credits and included in the GPA.

Please click [help](#) for more transcript information.

Student Name:	Schwenger, Emily Sachiko
McGill ID:	260397980
Permanent Code:	SCHE08539208
Advisor(s):	Oliooff, Paul R

PREVIOUS EDUCATION

BC High School
McGill University - Bachelor of Science 2014

Credits Required for BSc Bio,Biomed&Life Sc Honours - 120 credits

Fall 2010

Bachelor of Science
Full-time Year 0
Major Freshman Program

Credits/Exemptions

From: Advanced Placement Exams - 3 credits

MATH	140
------	-----

BIOL 111	001 Principles:Organismal Biology	3	A	3	A-
CHEM 110	001 General Chemistry 1	4	A	4	B+

PHYS 101	001 Intro Physics - Mechanics	4	A	4	B+
PSYC 100	001 Introduction to Psychology	3	A	3	B
TERM GPA:	4.00	Advanced Standing & Transfer Credits:	3.00	TERM TOTALS:	14.00 14.00 14.00 56.00
CUM GPA:	4.00	TOTAL CREDITS:	17.00	CUM TOTALS:	14.00 14.00 14.00 56.00
Standing: Interim Satisfactory					
Winter 2011					
Bachelor of Science Full-time Year 0 Major Freshman Program					
BIOL 112	001 Cell and Molecular Biology	3	A	3	B+
CHEM 120	001 General Chemistry 2	4	A-	4	B+
MATH 141	001 Calculus 2	4	A	4	B-
PHYS 102	001 Intro Physics-Electromagnetism	4	A	4	B+
TERM GPA:	3.92	Advanced Standing & Transfer Credits:	3.00	TERM TOTALS:	15.00 15.00 15.00 58.80
CUM GPA:	3.95	TOTAL CREDITS:	32.00	CUM TOTALS:	29.00 29.00 29.00 114.80
Standing: Satisfactory					
Dean's Honour List					
Fall 2011					
Bachelor of Science Full-time Year 1 Major Pharmacology					
BIOL 200	002 Molecular Biology	3	A	3	B
CHEM 204	001 Physical Chem./Biol.Sci. 1	3	B	3	B-
CHEM 212	001 Intro Organic Chemistry 1	4	A	4	B+
MATH 203	002 Principles of Statistics 1	3	A	3	B+
PHGY 209	002 Mammalian Physiology 1	3	A-	3	B
PHGY 212	001 Introductory Physiology Lab 1	1	B	1	B-
TERM GPA:	3.71	Advanced Standing & Transfer Credits:	3.00	TERM TOTALS:	17.00 17.00 17.00 63.10
CUM GPA:	3.86	TOTAL CREDITS:	49.00	CUM TOTALS:	46.00 46.00 46.00 177.90
Standing: Interim Satisfactory					

Winter 2012

Bachelor of Science
Full-time Year 1
Major Pharmacology

BIOL 201	001 Cell Biology & Metabolism	3	A-	3	B
BIOL 202	001 Basic Genetics	3	A	3	B+
CHEM 222	001 Intro Organic Chemistry 2	4	A	4	B+
PHGY 210	001 Mammalian Physiology 2	3	A	3	B-
PHGY 213	001 Introductory Physiology Lab 2	1	A	1	A-

	Advanced Standing & Transfer Credits:	TERM TOTALS:	Att Cr	Earned Cr	GPA Cr	Points
TERM GPA:	3.93	3.00	14.00	14.00	14.00	55.10
CUM GPA:	3.88	TOTAL CREDITS:	63.00	CUM TOTALS:	60.00	233.00

Standing: Satisfactory

Fall 2012

Bachelor of Science
Full-time Year 2
Major Pharmacology

BIOC 311	001 Metabolic Biochemistry	3	A-	3	B
BIOL 301	001 Cell and Molecular Laboratory	4	A	4	B+
BIOL 314	001 Molecular Biology of Oncogenes	3	B+	3	B
ENGL 200	001 Survey of English Literature 1	3	B+	3	B
PHAR 300	001 Drug Action	3	A	3	B+

	Advanced Standing & Transfer Credits:	TERM TOTALS:	Att Cr	Earned Cr	GPA Cr	Points
TERM GPA:	3.68	3.00	16.00	16.00	16.00	58.90
CUM GPA:	3.84	TOTAL CREDITS:	79.00	CUM TOTALS:	76.00	291.90

Standing: Interim Satisfactory

Winter 2013

Bachelor of Science
Full-time Year 2
Major Pharmacology

CEAP 250	703 Research Essay & Rhetoric	3	A-	3	B+
CHEM 302	001 Intro Organic Chemistry 3	3	A	3	A-
ECON 209	004 Macroeconomic Analysis&Applic	3	A	3	B
PHAR 301	001 Drugs and Disease	3	A	3	B+
PHAR 303	001 Principles of Toxicology	3	A	3	B

	Advanced Standing & Transfer Credits:		TERM TOTALS:	Att Cr	Earned Cr	GPA Cr	Points
TERM GPA:	3.94	3.00	TERM TOTALS:	15.00	15.00	15.00	59.10
CUM GPA:	3.85	TOTAL CREDITS:	94.00	CUM TOTALS:	91.00	91.00	91.00

Standing: Satisfactory

Fall 2013

Bachelor of Science
Full-time Year 3
Honours Pharmacology

COMP 202	002 Foundations of Programming	3	A	3	A-
ECON 208	001 Microeconomic Analysis&Applic	3	A	3	B
PHAR 503	001 Drug Discovery & Development 1	3	A-	3	A-
PHAR 562	001 General Pharmacology 1	3	A	3	A-
PHAR 599D1 ✦ 001 Pharmacology Research Project	3	A-	3		

	Advanced Standing & Transfer Credits:		TERM TOTALS:	Att Cr	Earned Cr	GPA Cr	Points
TERM GPA:	3.88	3.00	TERM TOTALS:	15.00	15.00	15.00	58.20
CUM GPA:	3.86	TOTAL CREDITS:	109.00	CUM TOTALS:	106.00	106.00	106.00

Standing: Interim Satisfactory

Winter 2014

Bachelor of Science
Full-time Year 3
Honours Pharmacology

CHEM 382	001 Organic Chem:Natural Products	3	A-	3	C+
PHAR 558	001 Pharmacology Selected Topics	3	A	3	A
PHAR 563	001 General Pharmacology 2	3	A-	3	B+
PHAR 599D2 ✦ 001 Pharmacology Research Project	3	A-	3	A-	
SOCI 377	001 Deviance	3	A	3	B+

	Advanced Standing & Transfer Credits:		TERM TOTALS:	Att Cr	Earned Cr	GPA Cr	Points
TERM GPA:	3.82	3.00	TERM TOTALS:	15.00	15.00	15.00	57.30
CUM GPA:	3.85	TOTAL CREDITS:	124.00	CUM TOTALS:	121.00	121.00	121.00

Standing: Satisfactory

Bachelor of Science Granted: May 2014

Honours Pharmacology

*

Distinction

First Class Honours in Pharmacology

RELEASE: 1.17

FORM NAME: SWFTRAN

NOTICE: Are you receiving "Page not working" or "Page can't be displayed" errors?

If you are using Internet Explorer or Microsoft Edge, try switching to Chrome or Firefox. If the issue still persists, please report it to the IT Service Desk at (514) 398-3398. If you can take a screenshot from Minerva, it will also help us.

© 2020 Ellucian Company L.P. and its affiliates.

Albert Einstein College of Medicine

1300 Morris Park Avenue
 Beifer Building, Room 202
 Bronx NY 10461
 (718) 430-8682

Emily S Schwenger

Date of Birth: 08-MAR

Admit Term: Summer 2019

Program: MD/PHD

Anticipated Degree: PHD

Majorl: Biomedical Sciences

SUBJ	NO	COURSE	TITLE	CRED	GRD	SUBJ	NO	COURSE	TITLE	CRED	GRD
Advisor:											
Ulrich Steidl											
CoAdvisor:											
Aviv Bergman											
Concentration:											
Cell Biology											
Systems & Computational Biol											

Previous Degrees:

Yeshiva University MS 2020-MAY-26

Qualifying Paper:

Qualifying Exam Title

Clonal Evolution in Low-Risk MDS: A Comparison of Stem Cell, Progenitor, and Blast Populations across Diagnosis, Treatment, and Relapse

INSTITUTION CREDIT:

Summer 2019

BIOS 5003 Histology & Cell Structure
LABR 9006 Thesis Research4.00 E
6.00 S

Att-Hrs: 6.00 Ehrs: 6.00

Cumu-Ahrs: 6.00 Ehrs: 6.00

Fall 2019

BIOS 5010 Mem Physiology & Transport
BIOS 7001 Biochemistry
BIOS 7006 Molecular Genetics
BIOS 7026 Intro to Systems Biology
LABR 9001 Thesis Research2.00 E
5.00 E
5.00 P
2.00 P
5.00 S

Att-Hrs: 12.00 Ehrs: 12.00

Cumu-Ahrs: 18.00 Ehrs: 18.00

Summer 2020

LABR 9006 Thesis Research

6.00 S

Fall 2020

BIOS 4004 Genomic Innovation

4.00 H

BIOS 7010A Quant Skill for Biomed Res I

1.00 E

BIOS 7010B Quant Skill for Biomed Res II

1.00 H

BIOS 7010C Quant Skill for Biomed Res III

1.00 H

LABR 9001 Thesis Research

5.00 S

Att-Hrs: 11.00 Ehrs: 11.00

Cumu-Ahrs: 47.00 Ehrs: 47.00

Spring 2021

BIOS 7027 Systems Biology Seminar

2.00 P

LABR 9001 Thesis Research

10.00 S

Att-Hrs: 12.00 Ehrs: 12.00

Cumu-Ahrs: 59.00 Ehrs: 59.00

Summer 2021

LABR 9006 Thesis Research

6.00 S

Att-Hrs: 6.00 Ehrs: 6.00

Cumu-Ahrs: 65.00 Ehrs: 65.00

Spring 2020

BIOS 5007 MSTP Mechanisms of Disease
BIOS 5011 MSTP Cardiac Physiology
BIOS 5012 Renal, R & A Physiology
BIOS 7007 Gene Expression: BTDH
BIOS 7018 Comp Biology of Proteins
BIOS 7020 Responsible Conduct of Res
LABR 9001 Thesis Research1.00 E
2.00 E
2.00 E
5.00 P
2.50 P
1.00 P
3.50 S

IN PROGRESS:

Fall 2021

LABR 9001 Thesis Research

12.00 ...

In progress : 12.00

Att-Hrs: 12.00 Ehrs: 12.00

Cumu-Ahrs: 30.00 Ehrs: 30.00

This officially sealed and signed transcript is printed on blue SCRIP-SAFE® security paper. When photocopied the name of the institution and the word VOID will appear. Translucent globe icons must be visible from both sides of transcript when held toward light source. A BLACK ON WHITE OR A COLOR COPY SHOULD NOT BE ACCEPTED.

In accordance with USC 438(6)(4)(8) (The Family Educational Rights and Privacy Act of 1974) you are hereby notified that this information is provided upon the condition that you, your agents or employees, will not permit any other party access to this record without consent of the student. Alteration of this transcript may be a criminal offense.

Kamala Lusk
Registrar

Albert Einstein College of Medicine

1300 Morris Park Avenue
 Belfer Building, Room 202
 Bronx NY 10461
 (718) 430-8682

Emily S Schwenger

Date of Birth: 08-MAR

Admit Term: Summer 2019

Program: MD/PHD

Anticipated Degree: PHD
 Majorl: Biomedical Sciences

SUBJ NO	COURSE	TITLE	CRED	GRD
***** TRANSCRIPT TOTALS *****				
TRANSFER:		Earned Hrs		
		0.00		
TOTAL INSTITUTION:				
		65.00		
OVERALL:				
		65.00		
***** END OF TRANSCRIPT *****				



EINSTEIN

Albert Einstein College of Medicine
 MONTEFIORE MEDICINE

This officially sealed and signed transcript is printed on blue SCRIP-SAFE® security paper. When photocopied the name of the institution and the word VOID will appear. Translucent globe icons must be visible from both sides of transcript when held toward light source. A BLACK ON WHITE OR A COLOR COPY SHOULD NOT BE ACCEPTED.

In accordance with USC 438(e)(4)(B) (The Family Educational Rights and Privacy Act of 1974) you are hereby notified that this information is provided upon the condition that you, your agents or employees, will not permit any other party access to this record without consent of the student. Alteration of this transcript may be a criminal offense.