Discovering Population-Specific Epigenetic Markers for Pancreatic Cancer Through Examination of Chromatin Accessibility

Krupa Sekhar

Personal Section

In the ninth grade, I had my life all figured out. I was going to be the next Preston Burke: world-renowned physician and master classical musician. Over the summer, when I wasn't practicing violin in preparation for my upcoming international competition, I was watching "Greys Anatomy." After I finished the tenth season, I decided that my life was going to be medical conferences in Switzerland, sleepless nights in the surgical gallery, and my violin to come home to. So, when I got admitted into the 9th grade Science Research Seminar, I decided I had to do a project involving a scalpel. You can guess how that went.

Flash-forward to a Tuesday in the March of ninth grade. I'm at Lennox Hill Hospital about to get surgery for the nerve I cut earlier that day with, you guessed it, my scalpel. The magic of the OR, in the moment, almost made up for the fact that my arm would be in a cast for the next few months. I was woozy from all the blood I had lost, but I know for a fact my surgeon was relieved when the anesthesia kicked in so he didn't have to keep answering my questions about the pulse-ox machine, and what his residency was like, and what all the numbers on the monitor were.

I remember waking up from the surgery, and hoping and praying that I could bounce back as quickly as the damage had been done. I went to school a couple of days after the incident and took the entrance examination for my school's tenth grade Science Research Seminar. I didn't get in. I remember holding my violin one night after a failed attempt to play, my tears leaving streak marks on the shiny wood, wondering if I would ever be able to play again. Needless to say, I missed the international competition that I had been practicing for. I spiraled into nightmares in which I was always lecturing my grandkids about scalpel safety. "I would've been able to play violin like a star, you know," I would say, or "Maybe I could've become a leading researcher if I had just not made that one mistake."

One sticky June morning, I had just woken up from one such night and came to the determined decision that my nightmares would remain just that. I hopped onto my laptop and typed "Independent research project ideas," and it was there that my journey began. Over the summer, I learned how to code and how to mine data, and finished my own research project on the inverse relationship between Cancer and Alzheimer's disease while my classmates did their summer assignments for tenth grade research seminar. I became one of the only tenth graders that year to compete and win in the NYC science and engineering fair, and my love for medical research started

there and blossomed into the future. I practiced violin every day, and slowly the pain shooting up my arm and the fact that I couldn't feel my middle finger failed to faze me. I participated in and thrived at the annual international competition during the tenth grade.

Two of the biggest players in my violin success were my great-aunt and uncle, avid music enthusiasts and my mentors. That same year, I lost both of them to stage-4 cancer, and was by their side as they suffered through years of painful chemotherapy. I applied to Memorial Sloan Kettering Cancer Center's 2019 Human Oncology and Pathogenesis Summer Student Program (HOPP) that fall with the mission of researching early diagnosis, so I could help minimize the pain my family experienced for others.

As a part of HOPP, I performed research in Dr. Christina Leslie's lab under the guidance of my mentor, Dr. Chirag Krishna. This research became the basis of the project I submitted to the Regeneron Science Talent Search. Dr. Krishna taught me how to use R to perform differential accessibility analyses, create heat maps, analyze chromatin accessibility on the Integrative Genomics Viewer, and more. HOPP was an incredibly formative experience-- as my mentor helped me build essential computational biology skills, my love for data analysis and biomedical research continued to grow.

As I gathered data to analyze, I ran into the striking problem of racial bias in all available samples and the disparity in early diagnosis and survival rates between the African American population and European individuals. As a woman of color and youth activist myself, I wanted to start a project that studied cancer epigenetics in the context of population-specific health disparities to begin revolutionizing early and equitable diagnosis methods and treatment options. As I pursued my research into the following summer, I began to see parallels between epigenetic variates and sex-specific higher incidence as well. My research focus shifted to epigenetic markers correlated with larger population-specific incidence (both racial and sexual), and how understanding these epigenetic markers could eventually aid in population-specific early diagnosis procedures.

I believe the most interesting and impactful next question to further the intersection of epigenetic oncology and population epigenetics is why and how do population-correlated cancer-causing epigenetic variations arise, and what can we do to prevent them?

While current research has made preliminary correlations between population epigenetics and cancer health disparities, I envision future research combining computational analysis with population-stratified genetic screening in order to both isolate targetable epigenetic markers of

cancer, and understand how they arise to inform preventative measures. Epigenetic screening data (ATAC-seq, BiSulfite-seq, etc.) performed on samples of 100,000+ patients from variable populations would undergo extensive comparative computational analysis in order to develop both predictive and correlative machine learning models (for early diagnosis and prevention, respectively). The population-specific genetic screening would include racial populations, sexual populations, and environmentally stratified populations (eg. low income vs high income neighborhoods), in order to understand the role of physiological ethnic/sexual epigenetic variation versus transgenerational environmental causes of cancerous epigenetic inheritance (eg. the effects of systemic housing oppression, which has forced African American communities into low-income and polluted neighborhoods for centuries, on environmental modulation of heritable epigenetics that contribute to higher population-wide susceptibility).

In 50 years, my hope is that researchers will have used this epigenetic data to develop comprehensive and accessible early diagnosis models for at-risk populations, isolated population-correlated oncogenic epigenetic markers, and started to develop epigenetic therapies for these markers (which, because epigenetic modifications are reversible unlike mutations, have already been shown to be highly effective in the clinical setting).

I could not have done my research without the guidance of so many. Firstly, thank you to the Leslie Lab, especially my mentor Dr. Krishna, for guiding me these past few years, challenging my thought processes, and letting me learn so much from them. I also need to thank Ms. Gilana Reiss, my high school research teacher, for her continued support, advice, and revisions of my paper. Thank you also to all of you reading-- if you are a high schooler interested in research, my biggest piece of advice is to take the initiative to do it! If outside opportunities don’t seem to welcome you at first, you can definitely use online resources, coding tutorials, and more to get started on your own independent project like I did. My second biggest piece of advice is to find a mentor-- or more than one! Their experience is invaluable, and there’s nothing better than learning from a seasoned professional to grow your skillset and mindset.

Rationale

Health disparities are deeply integrated into the fabric of medicine, contributing to disproportionate deaths in marginalized and under-researched communities. For example, in 2010, African Americans had a 30% higher heart disease mortality than whites, and African American men were two times as likely than whites to die prematurely from stroke (Gudsnuk et al. 2012). Although some health disparities can be attributed to healthcare system biases and disparities in healthcare access, there has been a general failing of medical researchers to acknowledge that the foundation of medical knowledge is fundamentally skewed towards Western populations. The majority of medical academia is from Western and Asian countries, and is also biased because of a lack of representative research staff, studying samples that do not include all races. Rather than understanding the variability of diseases between populations and the necessity to examine them in the context of this variation, current research has targeted the disease in an "umbrella fashion" without acknowledging that a general approach has failed to provide for many unincluded groups.

This study focused on pancreatic cancer (PC), the 3rd leading cause of all cancer-related deaths in the United States (Rahib et al. 2014) and a cancer that is extremely population stratified. PC mortality closely parallels incidence, making it one of the most lethal cancers (Kamisawa et al. 2016). PC incidence is up to 50-90% higher in African Americans (AAs) (Khawja et al. 2015) and is also significantly higher in males-- incidence is 5.5 every 100,000 patients versus 4.0 per every 100,000 for women (Rawla et al. 2019), with no biological explanation or screening diagnosis mechanism to explain or take early action against these population-specific disparities. Minimal progress has been made in the treatment of PC in the past 20 years- survival rates have remained relatively low and constant. Early detection is especially tough for PC because of the lack of common symptoms experienced by patients in the early stages of PC development, but effective diagnosis is paramount due to the lack of treatment options available.

PC arises out of pancreatic intraepithelial neoplasms (PanINs), resultant by excessive multiplication of epithelial cells in pancreatic ducts (Kamisawa et al. 2016). There is no screening mechanism for high-risk individuals/ populations-- PC diagnosis is currently based on external symptoms (jaundice, abdominal pain) which only present themselves during advanced tumor stages and are often misinterpreted as other pancreatic diseases such as pancreatitis and islet cell tumors (Miura et al. 2006). The only potentially curative treatment for PC is resection (Stathis et al. 2010), but half of all PC diagnoses occur when there is already distant metastasis and surgery

is not possible (Freelove et al. 2006). Overall five-year survival rate is ~4% as of 2006, and only 17% for resectable disease (Freelove et al. 2006).

Higher population-correlated incidence has been attributed to higher rates of smoking in males and AAs, higher risk factor incidence in AAs (eg. diabetes, pancreatitis), and healthcare biases, but these associations alone are not enough to explain the aforementioned incidence disparities (Khawja et al. 2015, Rawla et al. 2019). Unexplained population-specific incidence, poor early diagnosis mechanisms, and ineffective treatment options post-diagnosis compound to make prognosis after being diagnosed with PC one of most unpropitious of any cancer. Due to a lack of common symptoms in early stages, population-specific screening and categorization of such symptoms is needed for more effective early diagnosis, prevention screening, and effective treatment options.

Genome-wide association studies (GWAS) are the main tool currently used for detecting susceptibility loci for this disease. These studies use genomic data from a large cohort of patients to determine associations between specific shared mutations and genes, with occurence of PC. However, it has become more apparent that these studies identify mainly peripheral loci, which do not necessarily directly contribute to disease, and therefore are currently unhelpful to individual patients or for developing improved treatment methods (Frayling et al. 2014, Visscher et al. 2012). This study hypothesized that single nucleotide polymorphisms (SNPs) found using GWAS, as well as SNPs within linkage disequilibrium (non-random association) of the polymorphisms found, play a non-coding role in regulating chromatin accessibility and TF binding. Because GWAS does not take into account the holistic genomic landscape, it is not able to draw conclusions involving chromatin accessibility and epigenetic regulation which may be significant when discussing genetic changes correlated with PDAC development. Therefore, analyzing susceptibility loci in the context of chromatin accessibility may be significant in determining how these SNPs could be modulating PDAC development (and whether this modification occurs epigenetically). In addition, locating the regions of these SNPs and their functions can provide insight into locations of driver mutations, causal of PDAC progression, unlike the peripherally involved genes that current GWAS screening methods are hypothesized to detect.

Furthermore, GWAS is primarily based on genetic information obtained from the European population, when incidence of PC has consistently been higher in the AA population than in any other racial group, and the AA population also has the worst prognosis after being diagnosed

(Scarton et al. 2018). SNPs found in GWAS have not been shown to be replicable in the limited studies that have been performed in the AA population (Frayling et al.), demonstrating that this racial bias actively contributes to lack of information about AA-specific (the primary PC-inflicted population's) incidence of PC.

This study is the first of its kind to use epigenetics as a novel form of PC population-specific risk-factor categorization via Assay for Transposase-Accessible Chromatin Using Sequencing (ATAC-seq) used to assess chromatin accessibility (Figure 1), which is significant in understanding epigenetic regulation. This study began its investigation studying how SNPs associated with increased risk of pancreatic ductal adenocarcinoma affect chromatin accessibility and perturbation of transcription factor (TF) binding, and if ATAC-seq can be used to more accurately detect susceptibility loci for PC.

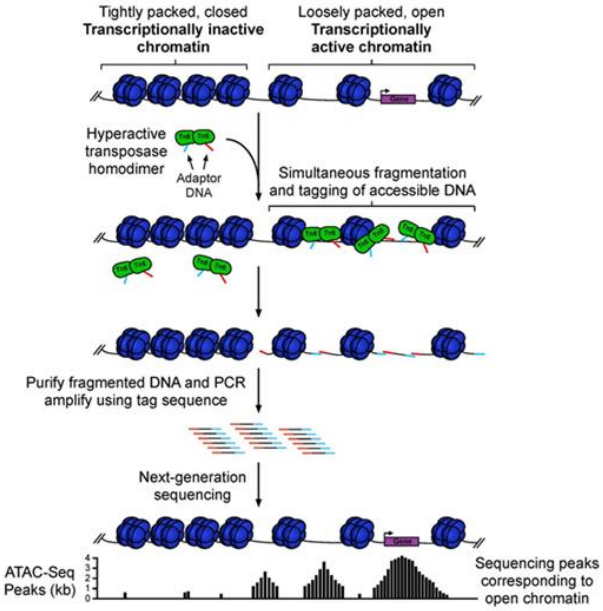

Figure 1: ATAC-Sequencing. (Ahmad et al. 2017)

Epigenetic variations are non-coding variations in DNA, specifically pertaining to structural aspects of DNA such as methylation and chromatin accessibility profiles. The epigenetic landscape differs significantly between racial groups (Mohammed et al. 2017), and is also highly variable between sexual populations (Ratnu et al. 2017). While PC epigenetics and population epigenetics have been explored separately, they have not been jointly explored to explain PC incidence and survival disparities, and no biological or genetic reason has been correlated with population-specific elevated PC incidence. The few studies that have explored epigenetics in the context of other cancer disparities have found significant correlations between variable epigenetic profiles in ethnic/racial populations and higher cancer incidence (Mohammed et al. 2017). Therefore, understanding the epigenetic factors contributing to disease can not only reveal underlying biological drivers of PC as a whole, but also isolate population-specific biological causes of higher PC incidence.

Epigenetics are highly affected by environmental factors including smoking, exposure to toxic substances, stress etc. (Marsit et al. 2015, Herceg et al. 2011, Gudsnuk et al. 2012). However, epigenetic variations caused by these environmental factors can be permanently programmed as epimutations, which allows environmentally impacted epigenetic profiles to be transgenerational (Skinner et al. 2010, Bale et al. 2014). Hence, epigenetic profiles can largely reflect generational socioeconomic lifestyles, and therefore are characterizable based on racial and sexual populations on the basis of their common generational experiences. This is significant in identifying the root cause of population disparities in order to preventatively target them-- higher incidence of certain cancers in racial populations may not be solely associated with race, but with the generational socioeconomic factors that have affected the population's epigenetic profile-- incidence of cancer along with other diseases in socioeconomically poor populations is consistently higher than those of their more privileged counterparts (Ahmad et al. 2017). Epigenetic variations found to be correlated with environmental/ transgenerational lifestyle factors can then be targeted in order to reduce PC risk/ treat PC.

In addition, unlike mutation-based contributors to disease, epigenetic variations are reversible and therefore are much more effective to target for treatment (Ahmad et al. 2017). Epigenetic therapies, using demethylating agents (DNMTs) and Histone deacetylase (HDAC) inhibitors, have had great clinical success for diseases including solid tumor cancers, Acute Myeloid Leukemia, T-cell lymphoma, and myelodysplastic syndromes (Nervi et al. 2015, Kelly et al. 2010).

Epigenetics has the unique potential to explain non mutation-based contributors to disease through examination of chromatin accessibility and methylation profiles, and because it is both associated with heritability and environmental factors, can unearth population specific and effectively targetable biological contributors to higher incidence of disease. This study novelly combined epigenetics with population oncology research in order to analyze population-specific PC disparities, and introduced an interdisciplinary computational population epigenetics methodology that can be utilized in the future to study other population-stratified diseases.