Investigating Atherosclerosis and Chemotherapy-Induced Cardiomyopathy at the Single-Cell Resolution

A DISSERTATION

Presented to the faculty of the School of Medicine in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Experimental Pathology

December, 2022

Wei Feng Ma, BA., MS.

Charlottesville, VA.

Table of Contents

Approval Sheet

This dissertation is in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Experimental Pathology.

Wei Feng Ma
Author

This dissertation has been read and approved by the examining committee:

Clint L. Miller, PhD
Thesis Advisor
Dept. of Public Health Sciences

James C. Zimring, MD, PhD
Committee Chair
Dept. of Experimental Pathology

Melanie R. Rutkowski, PhD
Dept. of Microbiology, Immunology, and Cancer Biology

Heather A. Ferris, MD, PhD
Medicine, Endocrinology and Metabolism

Jeffrey Saucerman, PhD
Dept. of Biomedical Engineering

Nathan C. Sheffield, PhD
Dept. of Public Health Sciences

Eli S. Williams, PhD
Dept. of Pathology

Accepted by the School of Medicine:

Melinda Kibbe, MD
Dean of Medicine
School of Medicine

Dedication

This dissertation is dedicated to my husband, Noah, my parents, and my two dogs: Butter and Noodle.

Dissertation Abstract

Cardiovascular diseases (CVDs) and cancer are the leading causes of death worldwide. Central to many cardiovascular diseases, including myocardial infarction, angina, and stroke, is the process of atherosclerosis, or the hardening of blood vessels. Single-cell sequencing has been used to elucidate many areas of physiology and disease. To understand fundamentals of the cellular processes at play and to derive better treatment and prevention of CVDs, we first examined a publicly available human-derived single-cell dataset of coronary arteries. In doing so, we discovered processes involving smooth muscle cell differentiation (Chapter 2). We also uncovered technical gaps in current practices of single-cell analysis, which gave rise to the development of PlaqView (Chapter 3). PlaqView (www.plaqview.com) is a single-cell web portal aimed to curate, standardize, and provide access for reanalysis of cardiovascular single-cell datasets without the need for technical coding skills.

We further explore the role of cardiovascular diseases in the context of cancer (Chapter 4). For cancer patients, cardiovascular side-effects from chemotherapy are of special concern. To further our understanding of chemotherapy-induced cardiomyopathies, we designed, optimized, and performed single-nucleus multimodal sequencing on left-ventricular tissues from six human patients. Our results suggest that cardiomyopathy arising from chemotherapy side-effects is transcriptionally different from other forms of cardiomyopathy, despite the same standard of treatment currently used clinically. Furthermore, we discovered the specific expression of *XIST* in chemotherapy-induced cardiomyopathy and many potential off-target effects of standard heart-failure therapies. Our data, which will be publicly available through PlaqView, opens up many potential avenues of future study (Chapter 5).

Acknowledgements

First, I thank my husband, Noah Kopp, for all the support and love he has given me throughout the years, my parents, Yan Ling Ma and Hua Xiao Ma, for the unconditional love and freedom they have given me to pursue the things I love, and of course, my two dogs, Butter and Noodle.

I acknowledge and thank my thesis mentor, Dr. Clint Miller, who fostered my growth both as a scientist and as a person, supported me throughout the PhD, and encouraged me to pursue unconventional ideas. I thank Chani Hodonsky, who helped me in my graduate years in so many ways. I would also thank the other members of the lab, including Gaëlle Auguste and Yipei Song for their feedback and help.

I thank my friends I have made in my college and medical school years: (in no particular order) Dana, Cassie, Chris, Courtney, Jasmine, Gustavo, Blair, Ryan, Ashley, Kara, Steven, Eric, Alex B., Alex M., and many others.

I thank my thesis committee, who took the time to listen and challenged me to think of the unexpected (in no particular order): Drs. Jim Zimring (Chair), Heather Ferris, Melanie Rutkowski, Jeff Saucerman, Eli Williams, and Nathan Sheffield. I would also like to thank Dr. Katia Sol-Chuch for her help in the library prep and sequencing for the chemotherapy project.

Lastly, I wish to thank Drs. Frank Fekete, Lynn Hannum, Dasan M. Thamattoor, Allison Galanis, Emese O'Donnell (Prandovszky), and Thomas Leto, who all played an important part in getting me to where I am today.

List of Abbreviations

- C7: complement component C7
- CAD: coronary artery disease
- CH: chondrocytes
- CIPR: Cell Identity PRedictor
- CVD: cardiovascular diseases
- CMP: common myeloid progenitor cells
- DCN: decorin
- DCM: dilated cardiomyopathy
- DGIdb: drug-gene interaction database
- DOX: doxorubicin (an anthracycline chemotherapy agent)
- EC: endothelial cells
- EGFR: epidermal growth factor receptor
- FB: fibroblasts
- FBLN1: fibulin 1
- GMP: granulocyte-monocyte progenitor cells.
- GSEA: gene-set enrichment analysis
- HF: heart failure
- HCM: hypertrophic cardiomyopathy
- LV: left ventricle
- Mø: macrophages
- MYH11: myosin heavy chain 11
- RAAS: renin-angiotensin-aldosterone system
- SC: stem cells
- sc/snATAC-seq: single cell/single nucleus assay for transposase-accessible chromatin sequencing
- sc/snRNA-seq: single cell/single nucleus RNA sequencing
- TI: trajectory inference
- SMC: smooth muscle cells
- UMAP: uniform manifold approximation and projection
- XIST: X-inactive specific transcript

Chapter 1: Background

1.1: Epidemiology of Cardiovascular Diseases (CVD)

Cardiovascular diseases (CVDs) remain the leading cause of death in the U.S., along with cancer and stroke (1–3). According to the 2021 morbidity and mortality report from the Centers for Disease Control and Prevention (CDC), heart disease (excluding stroke) causes about 161 deaths per 100,000 US population. In contrast, cancer and SARS-CoV-2 were responsible for 146 and 85 per 100,000, respectively (3,4). Surprisingly, the mortality rate for heart disease increased from 2019 to 2020 by about 7 per 100,000 (4). As demonstrated by the longitudinal Framingham Heart Study, the lifetime risk of a person developing cardiovascular or heart disease in the U.S. is approximately 50% (5).

Globally, CVDs are responsible for approximately one-third of all deaths, as reported in 2016 (6). In 2012-2013, CVDs were estimated to cause over 17 million deaths globally each year (7). Although the trend of CVD incidence and death is decreasing overall, CVDs remain a major driver of healthcare costs, disability, and quality of life (6).

1.2: Overview of Non-Atherosclerotic Cardiovascular Disease (CVD)

Although the term “CVD” is often associated with atherosclerotic disease (see next section), CVDs are heterogeneous in etiology with an array of genetic, environmental, and lifestyle factors (5). The World Health Organization (WHO) divides CVDs into the following groups of disorders: coronary heart disease, cerebrovascular disease, peripheral artery disease, rheumatic heart disease, congenital heart disease and thrombotic disease (deep vein thrombosis and pulmonary embolism) (8).

Rheumatic heart disease and congenital heart diseases are the two major categories that are distinct from non-atherosclerotic disease, although atherosclerosis may be present in both. In rheumatic heart disease, certain

infections such as *Streptococcus pyogenes* and *Enterococcus faecalis*, can colonize the heart and valves that lead to subsequent destruction and embolism to other parts of the body, such as the brain (9). Other infectious sources, such as the Coxsackie virus and common upper respiratory viral infections can cause viral myocarditis and pericarditis, respectively (10). These types of viral etiologies can cause transient heart diseases that coalesce into long-term restrictive heart disease, where the pericardium thickens and elicits diastolic dysfunction (10,11).

Congenital heart diseases have extremely broad clinical presentations. For example, Tetralogy of Fallot (TOF), the most common cyanotic congenital heart disease, manifests as cyanosis when the child is agitated or crying, and can be managed with supplemental oxygen in some cases (12). In contrast, transposition of the great arteries manifest cyanosis within the first 12h of birth, regardless of supplemental oxygen (13). For a comprehensive review of congenital heart diseases, readers are referred to the review by Puri, et al. (13).

1.3: Overview of Atherosclerotic Diseases

The majority of cardiovascular diseases are caused by atherosclerosis, the hardening of the vasculature via the development of fibrous and fatty lesions typically referred to as an atherosclerotic lesion or atheroma (14). The formation of atheroma marks the start of the atherosclerosis process, and can be observed even in children (15). In a study of teenagers who received heart transplants, about one in six patients were found to have abnormal thickening of the coronary arteries (16). In older adults, atherosclerosis directly contributes to the formation of several vascular pathologies, including coronary artery disease (e.g., angina, myocardial infarctions), peripheral artery disease (e.g., claudications), and cerebrovascular diseases (e.g. stroke).

In young adults and children, atherosclerosis begins with the formation of the atheroma, where the intima of the blood vessels involved thickens, and accumulates “lipid-laden macrophages” (14,17). Over decades, these atheromas continue to develop and grow in complexity. The vast majority of adults have some kind of atheroma or visible atherosclerosis in their vasculature, especially the aorta and coronary arteries. A major focus in the research field is the development and stability of the atheromatous plaques over the adult lifespan, and how these plaques can rupture or cause clinical symptoms such as angina (partial occlusion of the coronary artery leading to poor blood flow and transient ischemia of the heart).

In earlier lesions, a process called positive remodeling dominates as the primary compensatory mechanism. Positive remodeling is when the atherosclerotic plaque and vessel expand together to offset luminal loss, especially in the coronary arteries. In later stages, negative remodeling can progress with the external elastic membrane of the vessel failing to expand and the lumen narrowing from the expanding atheromatous plaque (14,15).

1.4: Known Risk Factors of Coronary Artery Diseases

Clinically, three major risk factors are assessed when determining the risk of an individual for Coronary Artery Disease (CAD) via the Atherosclerotic Cardiovascular Disease Score risk calculator (18): hypertension, smoking, and serum cholesterol. In fact, an additive effect is observed with increased blood pressure and other co-morbidities such as diabetes, age, smoking status, and sex (males have a higher prevalence of CAD) (19,20). Alternatively, key factors have been identified that are associated with good health and promoted by the American Heart Association, which include not smoking, healthy diet, physical activity, and normal blood pressure, glucose, cholesterol, and weight (7).

Although there are clear environmental risk factors that can be somewhat controlled, like smoking, family history, genetic and epigenetic factors that contribute to atherosclerosis and CVDs remain incompletely described (21,22). For example, genome-wide association studies (GWAS) have found over 200 regions within the human genome where one or more polymorphisms (SNPs) are associated with CAD (22). Furthermore, cell-type-specific epigenetic alteration and phenotypic switching have been found to affect both formation and stability of the atherosclerotic plaque (17,22,23).

1.5: Brief Overview of Research Areas in Atherosclerosis Research

Two areas of research currently dominate the atherosclerosis field of research: 1) understanding and slowing down the development of *atheromas*, and 2) understanding factors that contribute to the development of the *fibrous cap* and its stability (22,24–27).

Role of Inflammation

Both *in vitro* and murine models, as well as human observational studies, have implicated the role of the immune system in the formation of atherosclerotic plaques, or atherogenesis (14). For example, expression of VCAM-1 in the endothelial wall attracts leukocytes and is associated with an increase in size of the atheroma in mouse models fed a high-fat diet (14). Furthermore, inflammatory mediators are thought to inhibit collagen synthesis and increase collagenase expression, ultimately decreasing the overall stability of the fibrous cap (14).

Evidence derived from human observational studies suggests that systemic inflammation is associated with worse prognosis, as indicated by levels of

C-reactive protein (CRP), Interleukin (IL)-6, and other inflammatory cytokines. (7,14). Furthermore, elevated CRP is linked with recurrent angina and myocardial infarctions (14). Systemic inflammation is linked to other comorbidities such as diabetes, hypertension, and lifestyle factors such as smoking and alcohol use. Despite the strong evidence linking atherosclerosis and inflammation, clinical trials that have sought to target inflammation in atherosclerosis by reducing IL-1B, such as the Canakinumab Anti-Inflammatory Thrombosis Outcomes Study (CANTOS) trial, have failed to yield efficacious, FDA-approved therapies to-date (28).

Role of Oxidative Stress

As a part of the immune arsenal, oxidative mediators have been extensively studied in many common complex diseases with heritable components (29). Similarly, oxidative stress has been of interest in the cardiovascular disease field. The prevalent hypothesis surrounding the role of oxidative stress in atherosclerosis is that lipid oxidation induces cytokine expression, inflammation, and leukocyte recruitment (30,31).

Despite conventional wisdom suggesting oxidative stress from nitric oxide (NO) and reactive oxygen species (ROS) are deleterious, new evidence suggests that they may play critical roles in maintaining homeostasis and may be associated with better cancer prognosis (29). At this time, clinical trials of antioxidants have largely proven ineffective and further research is needed to clarify the role of ROS and NO in the atherosclerotic and cardiovascular environment.

Role of Genetics

Genome Wide Association Studies (GWAS) have identified over 200 loci associated with CAD (22). While few specific genetic abnormalities have been linked to overt CAD, such as mutations in 5-lipoxygenase activating protein (FLAP) in MI and stroke (32), a large number of genetic analyses in cohort studies imply that the cumulative effects of a large number of variants with small genetic and epigenetic effect sizes may play important roles in cell type-specific contributions to CADs (33–38).

1.6: Overview of Medications for Atherosclerotic and Cardiovascular Diseases

Anti-Inflammatory Agents

Despite the well-known roles of inflammation and ROS in the pathogenesis of CVDs, the role of *direct* anti-inflammatory agents (such as oral steroids or immunosuppressants) are limited except in cases of autoimmunity-mediated diseases (11,39,40). However, *indirect* effects of several drugs, such as statins (discussed in next section), and aspirin (used as anticoagulants), may have some effect in modulating the immune response (40). Direct modulators of the immune system, such as canakinumab from the CANTOS trial, are not currently approved or used in routine treatments and preventions of cardiovascular events.

Statins

The three primary components to manage most cardiovascular diseases remain the same: control of blood pressure, cholesterol profiles, and lifestyle modifications. Clinically, statins, such as atorvastatin and simvastatin, are the most widely used medication to lower serum lipid levels in patients at-risk for

cardiovascular events such as a myocardial infarction or a cerebrovascular event (7,39,40). Statins work by inhibiting the hydroxymethylglutaryl-CoA (HMG-CoA) reductase enzyme in the sterol synthesis pathway (41), thereby lowering the overall serum cholesterol level. Clinically, statins have been shown to prevent cardiovascular events and have mortality benefits (41).

There is also an increasing body of evidence that suggests statins may modulate the stability of the plaque by altering the expression of matrix metalloproteinases (40,42), CRP, and increasing nitric oxide synthesis (43). In clinical trials, statins have been shown to reduce cardiovascular-related mortality, reduce plaque size, and have been the mainstay treatment for atherosclerosis for many years (39). These properties of statins have been actively investigated for patients with heart failure; the plaque-stabilizing and anti-inflammatory properties of statins may reduce the risk for acute coronary syndrome when the heart is failing (44).

ACEs and ARBs

Other than managing lipid profiles, many cardiovascular diseases can be managed with blood pressure medications, when appropriate. Many agents belong in this class and their detailed mechanisms are beyond the scope of this thesis. However, the two mainstays in this class, beta-blockers and angiotensin-converting-enzyme inhibitors (ACE-inhibitors), will be briefly discussed here, as they have been proven to decrease mortality in both coronary artery disease and heart failure (45,46).

ACEs have been widely used to control blood pressure by affecting the renin angiotensin-aldosterone system (20,45,46). ACEs and its related drug ARBs (angiotensin II receptor blockers), are first-line therapy of choice for essential

hypertension (45). The renin-angiotensin-aldosterone system (RAAS) is active in heart failure, where reduced blood flow to the kidneys triggers systemic vasoconstriction to increase blood flow to the kidneys. By inhibiting the RAAS, ARBs and ACE-inhibitors cause reductions in the systemic blood pressure. Furthermore, ACEs have been shown in the CONSENSUS study to reduce the one-year mortality of advanced heart failure from 52% to 36% (47). However, despite the wide use of ACE-inhibitors and ARBs, the overall five-year mortality for heart-failure remains unchanged at 50% (46).

Beta-Blockers

Beta-blockers can have varied effects based on their selectivity to the beta-adrenergic receptor (48). In general, they lower the heart rate and cause some vasodilation. Beta-blockers have been consistently shown to reduce mortality in heart failure patients by 30% or more (48). However, at this time, beta-blockers are usually not used to as monotherapy to treat essential (idiopathic) hypertension, but are used in heart-failure or in combination with other drugs such as ACE-inhibitors in controlling hypertension (20,45,46)

1.7: Interplay Between Cardiovascular Diseases and Cancer

Atherosclerosis is implicated in many other diseases, including heart failure, hypertension, and kidney diseases (14). In particular, the interplay between cancer and heart diseases and the field of “cardioncology” as a whole is gaining significant momentum (49,50).

Together with cardiovascular disease, cancer remains a leading cause of death in the U.S. and developed countries (3). The development of novel cancer

therapies and increase in the molecular understanding of many forms of cancer have significantly blunted the mortality rate from cancer (51). However, many cancer survivors are now dealing with the sequelae from off-target effects of anti-cancer therapies, particularly cardiovascular side-effects.

Overview of General Cancer Treatments

Besides surgical resection, cytotoxic chemotherapies are the mainstays in the cancer treatment arsenal. These therapies can be broadly divided into targeted and non-targeted subtypes. Non-targeted cytotoxic therapies, such as 5-fluorouracil (an anti-metabolite) and doxorubicin (an anthracycline drug), are the most widely-used chemotherapies for a variety of malignancies worldwide, and are particularly effective in solid tumors such as colorectal cancer (51). On the other hand, targeted molecular therapies such as trastuzumab, an anti-ErbB2 (HER2) monoclonal antibody, have given new hope in the treatment of resilient cancer subtypes (52). Non-chemotherapy treatments, such as radiation, either as a monotherapy or in combination with other modalities, can also significantly reduce the disease burden in Hodgkin's lymphoma and other malignancies (53,54).

Despite increasingly sophisticated chemotherapy designs, there is a simultaneous increase in cancer-*related* diseases, often due to the very therapies used to combat cancer (52). Interestingly, almost all of the aforementioned therapies have been implicated in the development of cardiovascular sequelae, even after years of remission. A specialized field of medicine, called ‘cardio-oncology,’ has emerged to deal with patients with pre-existing cardiac conditions, and patients who develop cardiac symptoms as a result of therapy, both of which require specialized therapy plans.

Cardiovascular Complications from Chemotherapies

A wide range of cardiotoxic effects can be observed in each kind of chemotherapy, ranging from acute-onset type, i.e., myocardial infarctions and arrhythmias, to early or late-onset, i.e., fibrosis and cardiomyopathy. Of note, almost all known categories of chemotherapies can cause cardiomyopathy or ventricular dysfunction (52).

Currently, chemotherapy-related cardiotoxicity is mostly thought to affect the myocardium through the generation of reactive oxygen species (ROS), direct DNA damage, mitochondrial dysfunction, and changes in the transcriptome that leads to cardiac remodeling and fibrosis (55–57). Among these mechanisms, changes in cellular transcriptome are of particular interest; recent findings using induced pluripotent stem cell models suggest that expression of certain transporters allow import and accumulation of cytotoxic drugs within the myocardium (58). However, previous studies of chemotherapy-induced cardiomyopathies have been done only in murine and *in vitro* models, and a comprehensive map of the cell-specific processes in primary human tissues is not available.

Chemotherapy Cardiomyopathies Have Varying Clinical Presentation and Onset

Many chemotherapies cause late-onset cardiomyopathies (CMs), especially in patients who underwent treatment as children; up to 60% of pediatric patients who received anthracyclines develop some forms of cardiac dysfunction later in life (55). In pediatric patients, anthracycline causes the insidious onset of asymptomatic CM that progresses to restrictive cardiomyopathy (RCM) in adulthood (or in teenage years), even if treatments were stopped many years ago. In contrast, patients who were treated as adults generally develop dilated cardiomyopathy (DCM) (51,59,60).

For newer types of therapies, such as tyrosine kinase inhibitors used to treat BCR-ABL leukemia in children, the long-term effect is unknown due to the age of the drug. While the molecular underpinnings of these phenotypic differences are unclear, the urgent nature to develop preventative treatments is evident, as there are currently no effective treatments for these late-onset heart failure patients besides a heart transplant, which puts patients on life-long immunosuppressants.

Chemotherapies Adversely Affect Microcirculation

Hallmarks of cardiovascular toxicity from chemotherapies include systemic hypertension, pulmonary hypertensions, myocardial infarction (acute coronary syndrome), claudication (peripheral artery diseases), and stroke (cerebrovascular events). Depending on the therapy used, the side effect profile can vary greatly. However, the recurring theme of damage to the microcirculation persists throughout the chemotherapeutic arsenal (61).

Although the exact mechanisms of actions from chemotherapies on the endothelial tissue is unknown, there are several hypotheses. A number of drugs, like 5-fluorouracil and cisplatin, are thought to cause increased vasoreactivity or spasms via either permanent or transient damage to the tissue (61). Other drugs, such as VEGF inhibitors, may cause damage via inhibiting nitric oxide synthesis through the Akt/Pkb pathway (61). In any case, the common outcome is the development of cardiovascular events like angina and acute coronary syndrome. Furthermore, anthracycline-induced cardiomyopathies seem to be tightly linked with increased reactive oxygen species generation (52,60).

Clinical Practices and Limitations

The current management of chemotherapy-related cardiac dysfunction is essentially the same as other forms of cardiomyopathy. Patients who have preexisting cardiovascular diseases, such as angina (partial blockage of the coronary arteries) have to 'optimize' these preexisting conditions before undergoing treatment with anthracyclines, i.e. getting blood pressure down via the use of beta-blocker, lifestyle modifications, and tobacco cessations (52,53,62). In patients with preexisting low ejection fractions (a surrogate measure of % heart function), ACEs, beta-blockers, and dexrazoxane (a free iron-chelator) may be used (62). Despite being FDA-approved, the clinical effect of dexrazoxane has been controversial and is reserved for a small cohort of breast cancer patients who have received a significant amount of doxorubicin (62). In children, there are very few clinical guidelines in regards to protecting cardiovascular functions (63). To summarize, there are no effective preventative measures or cardioprotectants to date for patients with cancer undergoing chemotherapy.

Although highly desired in the clinical setting, stratification strategies to identify patients who are at the highest risk of cardiovascular sequelae from chemotherapy are not available. Currently, the only clinical measure used to estimate patients' risk is left ventricular ejection fraction (LVEF) (62). Ideally, a future model based on demographic data as well as sequencing data based on either tumor or other biomarkers would be highly informative.

1.8: Current Single-Cell Biology Literature in Cardiovascular Diseases

Single-cell Technologies Hold Promise In Unlocking Novel Preventative Treatments

Beyond understanding the fundamental cellular and genetic underpinning of diseases, single-cell technologies have been expanded in many areas of research

to improve translational and precision medicine. So far, clinical applications of scRNA-seq include: identifying mechanisms of chemotherapy resistance, drug discovery and design, and disease state prediction in combination with machine-learning (64–67). Furthermore, a recently developed machine-learning algorithm applied to single-cell data was able to accurately predict the disease state of multiple sclerosis, a multifactorial disease, using only 13 patient samples (67). In other words, single-cell methods increase the prediction and statistical power when primary samples are scarce.

Single-cell Study on Sporadic Hypertrophic and Dilated Cardiomyopathy

So far, four major studies exist that detail the single-cell environments of hypertrophic and dilated cardiomyopathies arising sporadically or by known genetic mutations (68–71). Koenig et al. described seven groupings of cardiomyocyte states across healthy vs. dilated cardiomyopathy (71). They also noted specific patterns of gene expression in DCM cardiomyocytes, such as lower expression of NPPA, JAK2, STAT4, and increased ADRB2 and ANKRD1 (71), which they termed “disease-associated cardiomyocyte state”. In addition, they note the shift and emergence of disease-associated monocyte and fibroblast populations. This study serves as an important landmark dataset for non-chemotherapy-related dilated cardiomyopathy.

Similarly, Reichart et al. described important findings in dilated cardiomyopathy in the context of certain genetic mutations, such as titin (TTN) and lamin A (LMNA). Previous studies have shown that these mutations, often in anchoring proteins (lamin A/B, plakophilin-2) and structural proteins (troponin T/C, Titin), can cause DCM and heart failure (70,72). In this study, they described various states of cardiomyocytes and fibroblasts as pertains to disease states (70).

Interestingly, they described altered cellular communication profiles when comparing healthy vs. diseased. Taken together, the two studies by Reichart et al. and Koenig et al. provide a useful reference and are particularly useful to discern cellular differences in cardiomyopathies.

In contrast, Chaffin et al. compared healthy hearts against dilated (DCM) and hypertrophic cardiomyopathies (HCM) (68). Although the sequenced cellular compositions are roughly similar, they noted striking differences in transcriptional profiles between healthy vs. DCM and healthy vs. HCM, but not DCM vs. HCM. They observed “activated” fibroblasts using pseudotime and trajectory analysis, but also experimentally induced these states in lentiviral experiments (68). In the context of cardiomyopathies in general, it remains unanswered whether chemotherapy-induced DCM are transcriptionally similar to DCM or HCM.

Lastly, Nicin et al. showed that in pressure-induced cardiac hypertrophy (i.e. aortic stenosis) hypertrophied cardiomyocytes have down-regulated EPH Receptor B1 (EPHB1) expression. In a co-culture experiment, reduction of EPHB1’s ligand, EFNB2, in endothelial cells induced cardiac cell hypertrophy (69). Although EPHB1 is shown to play a role in cardiac tissue architecture (73), the role of EPHB1 in chemotherapy-induced DCM is not investigated.

1.9: Current Gaps in the Field and Justification for this Work

Gap in Methodology and Access

Once published, single-cell datasets are usually deposited in static data banks such as Gene Expression Omnibus (GEO), where reanalysis with newly developed tools is difficult and requires advanced programming skills. As single-cell technologies become more widely available and more data are generated, there has

been a rapid increase in the number of tools designed to help parse out biological and clinical relevance. At the time of writing, there are over 800+ tools listed on scRNA-tools.org (74).

Despite the abundance of computational tools available, the fact that most tools require the users to have substantial programming skills remains a significant barrier to analyzing and exploring single-cell data, particularly across multiple datasets (75,76). Few tools, such as the cBioPortal for Cancer Genomics (www.cbioportal.org), have distilled extremely large datasets and made them accessible to everyone (77). However, no such tool currently exists for the cardiovascular field.

Despite the rapid growth in single-cell technologies and datasets, previously-published single-cell datasets are not readily available to scientists for several reasons: 1) Datasets are often very large, which require specialized computing clusters to access and parse. 2) No standardized guidelines exist for single-cell data, so the format deposited in public repositories such as the Gene Expression Omnibus (GEO) varies greatly. 3) Cell metadata are often not included, which serves as a severe drawback for scientific reproducibility. 4) Advances in single-cell computation methods require the users to have high levels of programming knowledge (e.g. R, Python) and are not always feasible in most scenarios. 5) Datasets are often not publicly deposited and reanalysis of crucial datasets is often impossible. As more datasets emerge, the need for a central repository specifically tailored for single-cell atherosclerosis datasets becomes more pressing.

Gap in Knowledge

Single-cell sequencing methods can uncover the heterogeneous cellular response to diseases and environmental insults, and are particularly useful in CVDs for discovering maladaptive responses and key drivers in various cell types (68–71). In coronary artery disease and heart failure studies, novel cell-specific regulatory and repair mechanisms in response to disease and therapy have been identified using single-cell methods. While some somatic mutations (e.g. Titin) are clearly linked to familial DCM, the risk factors and specific cellular changes underlying chemotherapy-induced DCM are not clearly understood. To date, no single-cell studies or comprehensive atlases exist for chemotherapy-induced DCM in human primary tissues, creating an enormous knowledge gap in its cellular landscape and pathophysiology. Here, I hypothesize that single-cell analysis of primary chemotherapy-induced DCM tissues will reveal clinically relevant cellular mechanisms that may be used for prevention and therapy.

1.11: Overview of Dissertation

In Chapter 2, I detail my original research work on a reanalysis of a previously-published single-cell dataset on human atherosclerosis, and the formation of the initial release of PlaqView, a web-portal for analysis of a few atherosclerosis-related single-cell data.

In Chapter 3, I describe the expansion and rebuilding of PlaqView to PlaqView 2.0, with great enhancement to features, back-end structures, scope of data, and user-interface.

In Chapter 4, I explain my latest work-in-progress, where I generated original human single-cell multimodal data to investigate the pathology underlying chemotherapy-induced cardiotoxicity. Here, I will also discuss shortcomings and future directions.

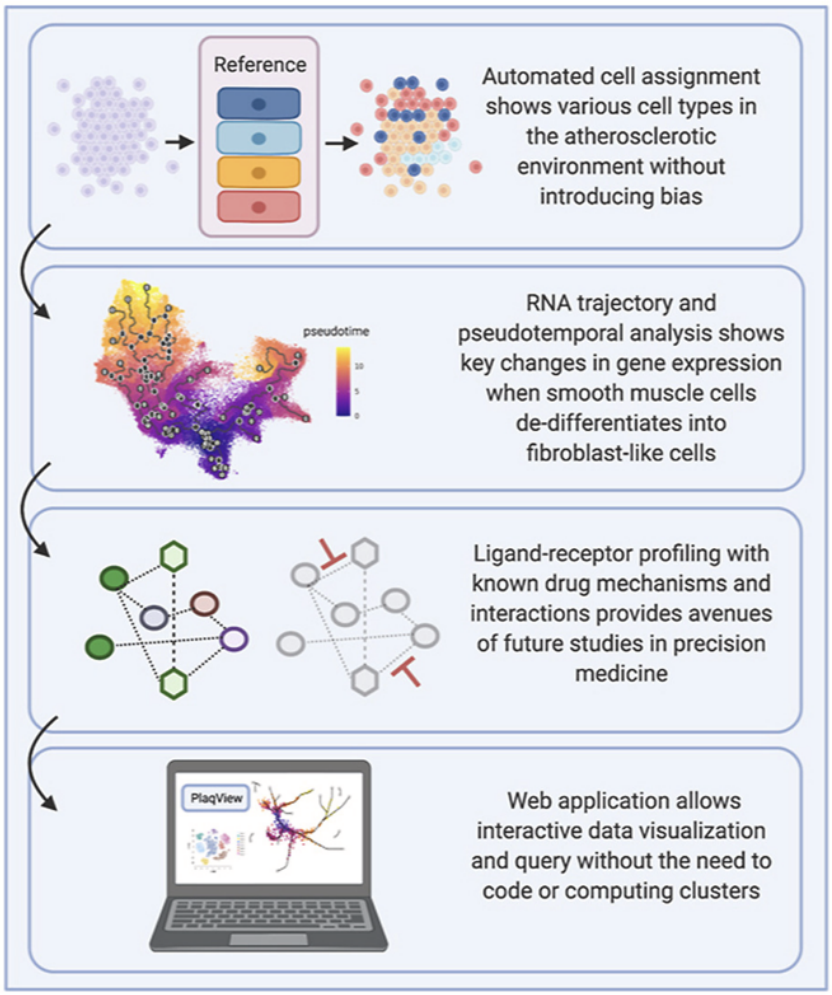
Chapter 2: Investigation of Human Atherosclerosis Microenvironment

This Chapter has been published at *Atherosclerosis* as the following Original Research Article:

Ma, W. F. *et al.* Enhanced single-cell RNA-seq workflow reveals coronary artery disease cellular cross-talk and candidate drug targets. *Atherosclerosis* **340**, 12–22 (2022). https://doi.org/10.1016/j.atherosclerosis.2021.11.025.

Figure and table numbers have been adapted for this thesis.

2.1: Abstract



Background and aims

The atherosclerotic plaque microenvironment is highly complex, and selective agents that modulate plaque stability are not yet available. We sought to develop a scRNA-seq analysis workflow to investigate this environment and uncover potential therapeutic approaches. We designed a user-friendly, reproducible workflow that will be applicable to other disease-specific scRNA-seq datasets.

Methods

Here we incorporated automated cell labeling, pseudotemporal ordering, ligand-receptor evaluation, and drug-gene interaction analysis into a ready-to-deploy workflow. We applied this pipeline to further investigate a previously published human coronary single-cell dataset by Wirka *et al.* Notably, we developed an interactive web application to enable further exploration and analysis of this and other cardiovascular single-cell datasets.

Results

We revealed distinct derivations of fibroblast-like cells from smooth muscle cells (SMCs), and showed the key changes in gene expression along their de-differentiation path. We highlighted several key ligand-receptor interactions within the atherosclerotic environment through functional expression profiling and revealed several avenues for future pharmacological development for precision medicine. Further, our interactive web application, *PlaqView* (www.plaqview.com), allows lay scientists to explore this and other datasets and compare scRNA-seq tools without prior coding knowledge.

Conclusions

This publicly available workflow and application will allow for more systematic and user-friendly analysis of scRNA datasets in other disease and developmental systems. Our analysis pipeline provides many hypothesis-generating tools to unravel the etiology of coronary artery disease. We also highlight potential mechanisms for several drugs in the atherosclerotic cellular environment. Future releases of *PlaqView* will feature more scRNA-seq and scATAC-seq atherosclerosis-related datasets to provide a critical resource for the field, and to promote data harmonization and biological interpretation.

2.2: Background

Atherosclerosis, a complex process involving chronic inflammation and hardening of the vessel wall, represents one of the major causes of coronary artery disease (CAD), peripheral artery disease, and stroke (78). Rupture of an unstable atherosclerotic lesion can lead to the formation of a thrombus, causing complete or partial occlusion of a coronary artery (79). The contribution of smooth muscle cells (SMCs) to both lesion stability and progression has recently been established by numerous groups. However, the exact mechanisms by which SMCs modulate the atherosclerotic microenvironment and whether pharmacological agents can be used to selectively counter SMC-related deleterious effects are still under investigation (17,80,81).

Recent advances in single-cell RNA-sequencing (scRNA-seq) have enabled ultra-fine gene expression profiling of many diseases at the cellular level, including atherosclerotic CAD (17). As sequencing costs continue to decline, there has also been a consistent growth in scRNA datasets, data analysis tools and applications (82). Currently, a major challenge with scRNA-seq analysis is the inherent bias introduced during manual cell labeling, in which cells are grouped by clusters and their identities assigned collectively based on their overall differential gene expression profiles (83). Another drawback inherent to commonly used scRNA-seq protocols is that they destroy the samples, making time-series analyses of the same cells impossible. Instead, these studies must rely on time-points from separate libraries to monitor processes such as clonal expansion and cell differentiation (24,84).

Recently, new approaches have been developed to compensate for both shortcomings, namely automatic cell labeling, pseudotemporal analysis, and

trajectory inference. Tools such as ‘SingleR’, ‘scCATCH’, ‘Seurat’, and ‘Garnett’ have been used to assign unbiased identities to individual cells using reference-based and machine learning algorithms (83,85–87). Moreover, tools such as ‘Monocle3’ and ‘scVelo’ align and project cells onto a pseudotemporal space where each cell becomes a snapshot within the single-cell time continuum (76,88). In essence, the single scRNA-seq dataset is effectively transformed into a time series (76,89,90). Although the pseudotemporal scale does not reflect the actual time scale, it is a reliable approximation to characterize cell fate and differentiation events, e.g., during organogenesis, disease states, or in response to SARS-CoV-2 infections (88,89,91).

In this study, we present the application of an enhanced, scalable, and user-friendly scRNA-seq analysis workflow on four previously published human and mouse atherosclerosis scRNA-seq datasets (17,81,92). Focusing on the only available human coronary artery dataset (17), we performed unbiased automatic cell identification at the single-cell level, pseudotemporal analysis, cell-to-cell communication profiling, and drug repurposing analysis. Our results demonstrate potential new mechanisms by which SMCs contribute to the atherosclerotic phenotype and signaling within the lesion microenvironment. More importantly, we revealed attractive candidate avenues and prospective therapeutic targets for experimental studies of pharmacological candidates. We also developed an interactive web application, *PlaqView* (www.plaqview.com), to allow users to explore this and other human and mouse cardiovascular single-cell datasets. This reproducible analysis pipeline and application can also be easily modified to incorporate different tissue data sources and single-cell modalities such as scATAC-seq or CITE-seq, and will serve as a template to analyze and visualize single-cell datasets in other disease models.

2.3: Results and Discussion

Unbiased automatic cell labeling is comparable to manual annotation, and reveals abundant cells with chondrocyte and fibroblast characteristics.

Recently, automatic cell identification tools have been implemented to overcome the subjective nature of manual cell-cluster labeling (83). Here, we compared two popular reference-based cell type annotation methods, 'SingleR' (83) and 'Seurat'(93), by applying these tools to a previously published human coronary artery scRNA-seq dataset(17), and found that 'Seurat' in combination with the *Tabula sapiens* (TS) reference(94) had the highest concordance with author-supplied manual labels when compared to labels produced by SingleR (Figure 2.1). For example, smooth muscle cells (SMC), pericytes, endothelial cells, and blood cells such as T-cells, B-cells, and mast cells are labeled in similar clusters. From Seurat/TS, we found that the majority of cell types were fibroblasts (FB, 27%), macrophages (Mø, 26%), endothelial cells (14%), and SMCs (11.5%, Figure 2.1B). In contrast, more cells were labeled as endothelial cells by both singleR (16.21%, Supplemental Figure 2.1A) and Wirka et al. (16.3%, Supplemental Figure 2.1B). Furthermore, Seurat/TS identified 11.5% cells as SMCs, whereas singleR and Wirka *et al.* found 13.8% and 6.2%, respectively. Although the true population percentage is likely somewhere in-between, this highlights the need to use multiple tools when analyzing single-cell datasets. We also applied an additional reference-based cell annotation tool 'scCATCH,' and found that it underperforms relative to Seurat and SingleR, and fails to provide consistent cell type assignment when provided with similar tissue priors (Supplemental Table 2.1). These and additional labeling tools have been extensively benchmarked elsewhere (95).

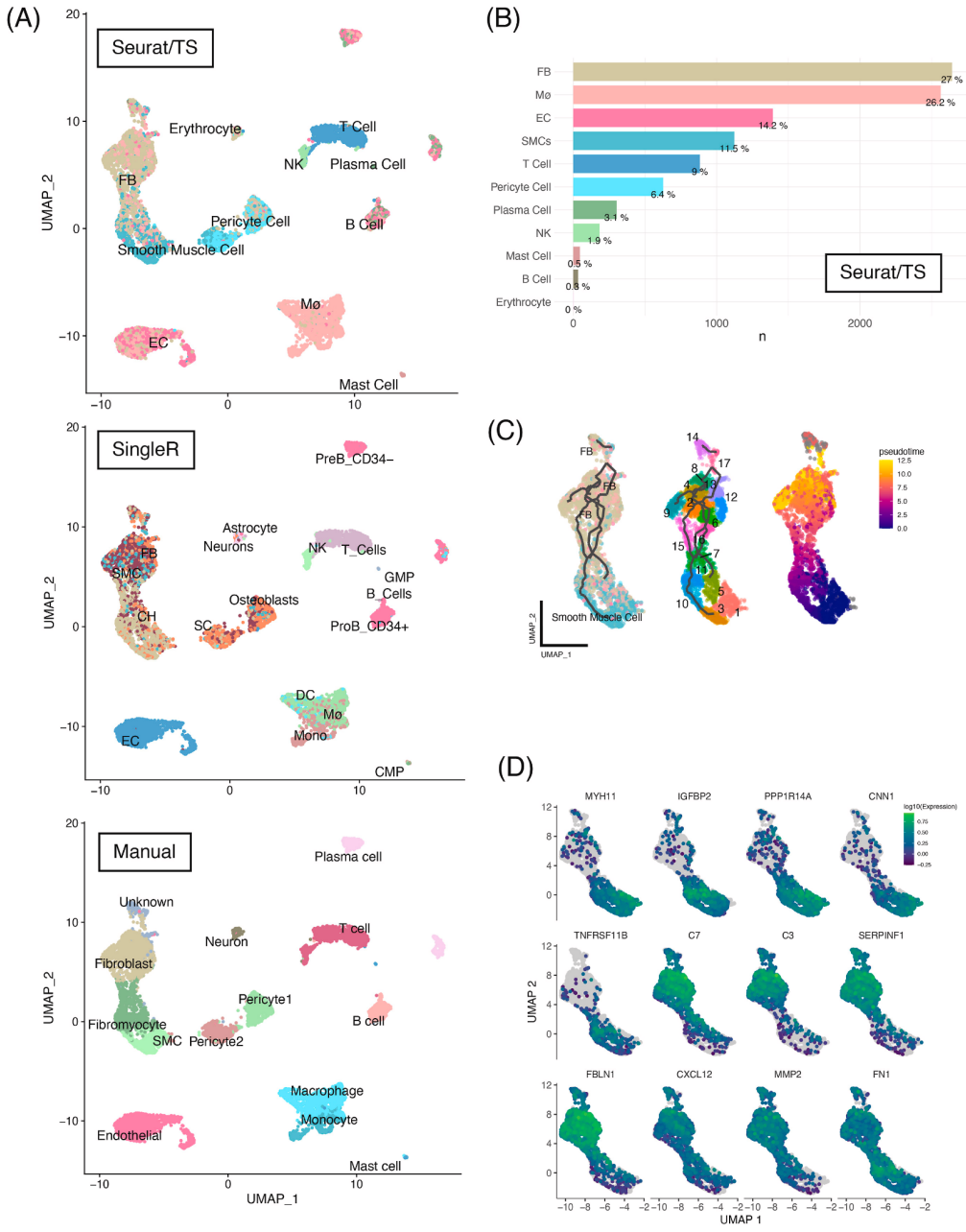


Figure 2.1: Unbiased automatic cell labeling of human coronary scRNA data using 'Seurat v4' had high concordance with manual labels and pseudotemporal and ultra-fine clustering reveals FB-like cells derivation from SMCs. (A) Top panel: UMAP clustering of 9798 cells derived from human coronary artery explants with labels based on Seurat v4 using the Tabula sapiens reference; middle panel: automatic labels provided by singleR using Human Primary Cell Atlas;

bottom panel: cluster-based manual annotation provided by original authors Wirka et al. (B) Population breakdown by percentage based on Seurat/TS labels. (C) Left: RNA trajectory (line) shows direct paths from the SMCs toward FBs; middle: ultra-fine clustering shows the logical transition stages (microclusters) from SMCs to FBs, and each cluster is numbered for clarity; right: pseudotemporal analysis confirms that the cell and clusters existing along a logical single-cell continuum. (D) Selected genes that were shown to vary over pseudotime by Moran's I test were visualized. SMC: smooth muscle cells, EC: endothelial cells, FB: fibroblasts, Mø: macrophages.

Although Seurat/TS did not identify the novel cell state that Wirka *et al*., refer to as ‘fibromyocytes’, a clear transition from SMCs to fibroblasts (FB) within the overall cluster was noted (Figure 2.1, top). However, Seurat/TS underperformed in some clusters. For example, SingleR identified neurons/astrocytes and PreB_CD34- cells in concordance with Wirka et al., whereas Seurat/TS labels neurons/astrocytes as Erythrocytes and did not identify the cluster PreB_CD34-. In their original publication, Wirka et al., did not identify clusters 10, 11, 12, 14, and 18 (17). However, using reference-based annotation, more cells were confidently identified, presumably due to the larger and newer references used and that individual cell expression is not masked by grouping of cells. SingleR identified cells that have osteoblast and chondrocyte (CH)-like cells, whose roles in atherosclerotic plaque stability are under active investigation (92,96). However, the exact identity and behaviors of these cells in this dataset will require further validation by comparing with other tissues with advanced calcification. In the UMAP clusters reflecting single-cell identities provided by SingleR (Figure 2.1, middle), there was an infiltration of osteoblast (OS), stem cell (SC), and chondrocyte (CH)-like cells within and next to the SMC cluster. Similar but more transitional cell-mixing is also

observed using Seurat/TS (Figure 2.1). Although UMAP representations often reflect global gene expression similarity, subtle changes in key phenotypic genes that cause cells to appear more similar to CHs or OSs are often difficult to observe in human single-cell datasets. Although the exact consequences of these phenotypic modulations are still under investigation, data from lineage-tracing mouse models suggest that SMCs transition to a panoply of phenotypes such as stem-like and osteogenic phenotypes, and that the osteogenic transcription factor Klf4 contributes to plaque destabilization (92).

Pseudotemporal ordering identifies distinct fibroblast-like cells originating from SMCs.

To evaluate putative cell fate decisions or differentiation events (e.g., SMC phenotypic transition states), we performed pseudotemporal analysis and ultra-fine clustering using 'Monocle3', a method previously applied to normal and diseased states, e.g., embryo organogenesis and response to coronavirus infection, respectively(88,89,91). We compared over 60 additional trajectory inference (TI) methods including Slingshot (97), PAGA (98), and SCORPIUS using the 'Dynverse' package (99,100). However, we found that most TI algorithms were unsuitable for complex tissue environments such as atherosclerotic plaques due to their inability to distinguish disconnected topologies (Supplemental Table 2). Consistent with Wirka et al., we found that SMCs directly give rise to FB-like cells as revealed by Monocle3 (Figure 2.1C, left). This corroborates earlier findings showing that SMCs may transition or de-differentiate into 'fibromyocytes'—SMCs that have undergone a phenotypic modulation to an extracellular matrix producing cell type within the atherosclerotic lesions (17,24). Although Monocle3 cannot distinguish a clear boundary where SMCs become so-called 'fibromyocytes', it infers a possible course by which SMCs transition to FB or FB-like cells through a series of smaller

‘transition states’ or fine-clusters (Figure 2.1C, middle). Additionally, pseudotemporal analysis using SMCs as a starting point shows clear alignment in the transition pathway, which further strengthens the idea that many of these FB are indeed closely related to SMCs (Figure 2.1C, right). These findings are consistent with mouse data using lineage tracing to establish that multiple cell types including ECM-producing and fibroblast-like cells are derived from SMCs (101).

Further analysis of the pseudotemporal graph using Moran’s I statistic is performed to identify specific changes associated in this transition pathway. Genes associated with healthy SMC phenotypes, such as *MYH11* (a canonical marker of SMC), *IGFBP2* (associated with decreased visceral fat), and *PPP1R14A* (which enhances smooth muscle contraction), are decreased by approximately 50-75% along the SMC trajectory as these cells become more FB-like (Supplemental Table 3, Figure 2.1D, $p < 0.1E\text{-}297$) (80,102). Similar results were found by another group using mouse lineage-traced models where MYH11 expression was decreased in SMC-derived modulated “intermediate cell states” (81).

More importantly, specific inflammatory markers and proteins associated with thrombotic events during CAD, including complement proteins *C7* and *C3, FBLN1*, and *CXCL12,* are increased along the same trajectory (Figure 2.1D) (79,103). Recent evidence suggests that CAD-associated *CXCL12* secreted from endothelial cells may promote atherosclerosis (104). Our results point to a potentially new source of CXCL12 that could be targeted to inhibit SMC-to-FB dedifferentiation. Together, and in corroboration of recent studies, our pseudotemporal analysis demonstrates that SMC and modulated, FB-like cells exist within a continuum, with phenotypes toward the FB phenotype associated with worse clinical observations (17,81). This is further supported by a recent study, in which blocking of SMC-derived intermediate cells coincides with less severe atherosclerotic lesions (81). Precisely how these modulated cells might influence the

overall stability of atherosclerotic lesions and clinical outcomes requires additional longitudinal studies using genetic models and deep phenotyping of human tissues (24,80,81).

Ligand-receptor analysis shows complex intercellular communications in the human coronary micro-environment

To examine the potential cross-talk between different cell types using scRNA-seq data, we performed cell-to-cell communication analysis using 'scTalk', a network-based modeling method that uses confirmed interactions from StringDB(105,106). The resulting networks are highly dependent on the prior labeling methods (Figure 2.2A). While FB cells labeled from Seurat/TS were shown to have weak outgoing signaling, FB from manual and singleR labeling have stronger outgoing signals. The strength of the interaction, or path weight, is calculated based on the summed weight of a four-layer node network as described in (106). Cells that are labeled as osteoblasts from SingleR had significant autocrine signaling as well as outgoing signals, but corresponding cells in the same UMAP regions (e.g., 'pericyte' in the manual labels) do not exhibit the same signaling pattern. This further highlights the need to explore various labeling priors as well as develop label-agnostic inference tools.

Recently, 'CellChat' and its companion database were introduced by Jin et al. CellChat infers the probabilities of cell-cell communications using the law of mass action, and is particularly useful in visualizing pathway-specific interactions (107). We provided CellChat with the labels and produced from Seurat/TS and scRNA-seq count matrices, and discovered several significant pathways: collagen, laminin, fibronectin (FN1), and complement signaling (Figure 2.2B). In particular, SMCs exert a strong signal in FN1 and collagen signaling, and FB have significant

contributions in laminin and complement pathways. Like collagen type IV, laminin isoforms form the basement membrane and have been studied in the context of atherosclerosis, particularly in endothelial cells (108). However, the exact mechanisms of how FB and FB-like cells contribute to plaque formation and/or stability requires future mechanistic studies.

In addition to complement activation, other immune system involvement is of particular interest in the atherosclerosis field (26). We examined the overall MHC-I and MHC-II signaling activation within the atherosclerotic environment (Figure 2.2C-D). We found that almost all cell types exert putative signaling toward NK and T-cells via MHC-I (Figure 2.2C), principally through HLA-A (Supplemental Figure 2.2A), whereas high MHC-II class interactions are focused on only macrophages (Figure 2.2D). While macrophages under basal conditions express low levels of MHC-II molecules(109), our result suggest that a majority of the cells within the lesion may contribute to macrophage activation through interactions such as HLA-DPA1:CD4, HLA-DMB:CD4, and HLA-DRA:CD4 interactions (Supplemental Figure 2.2B). In fact, many of these MHC-II class molecules, such as HLA-DRA and HLA-DMB, have been shown to have possible prognostic value in identifying atherosclerotic plaque rupture (110).

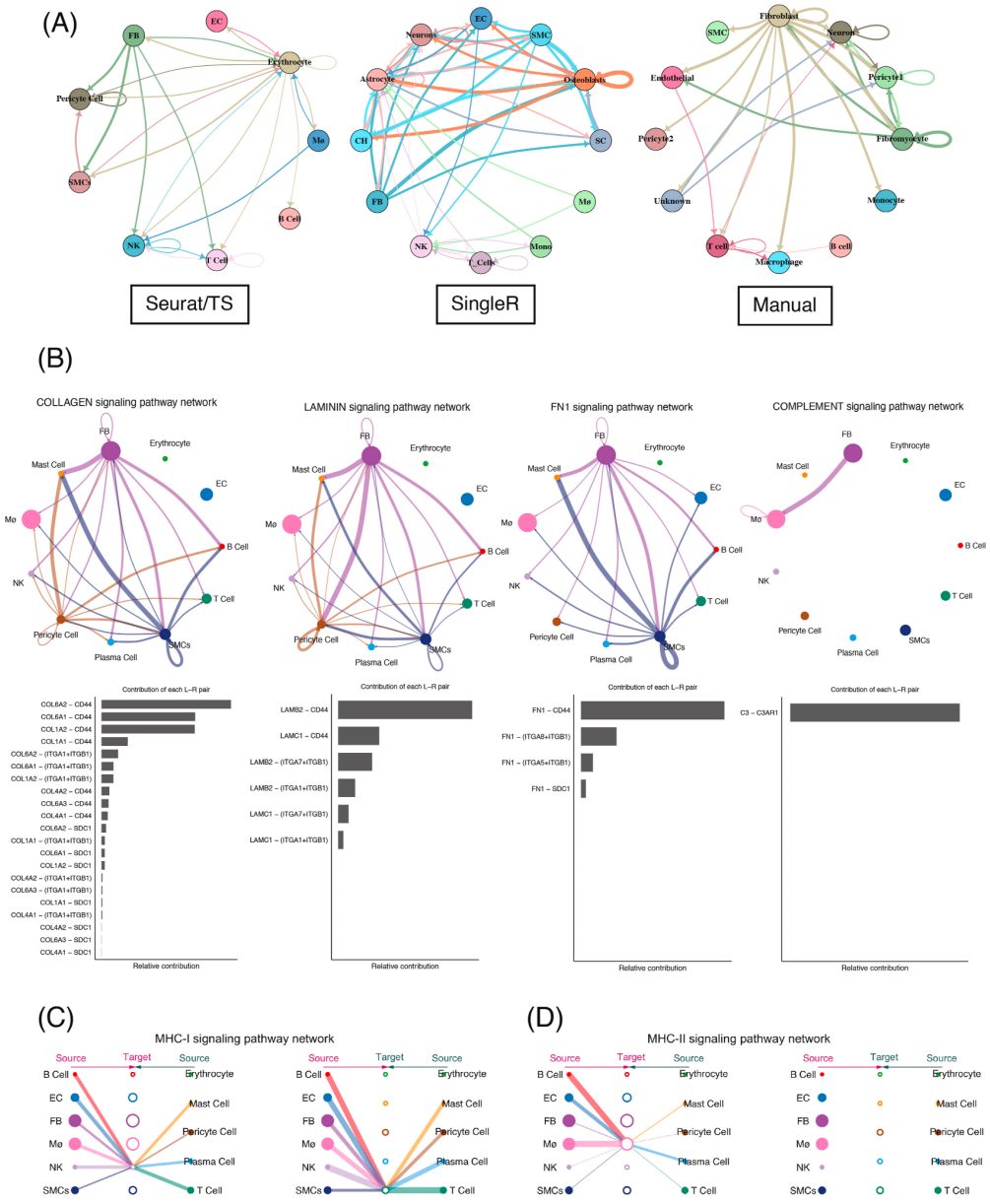


Figure 2.2. Comprehensive ligand-receptor analysis reveals the major inflammatory and immune signaling pathways in FBs.

To identify putative actionable signaling targets, we first evaluated the overall expression and signaling patterns between SMCs, FB, pericytes, and T-cells using singleR, then cross-referenced these interactions against known druggable databases using DGIdb 3.0 (111). As more cell types are included, the interaction network becomes increasingly complex (Supplemental Figure 2.2), thus more

simplified schematics depicting the overall results are presented (Figure 2.3A-B). Unlike scCATCH, singleR does not delineate the interactions by specific signaling pathways but rather by cell type-specific expression, which is particularly useful in providing a global overview of signaling for each cell type. We first highlighted the signaling from fibroblast (FB) or FB-like cells to T-cells, pericytes, and SMCs using labels provided by Seurat/TS. These signaling pairs involved various known inflammatory and repair mechanisms like C3 complement, fibulin-1 *(FBLN1)*, and matrix metalloprotease 2 (*MMP2*, Figure 2.3A, Supplemental Figure 2.2). We then examined how SMCs may signal to the other aforementioned cell types (Figure 2.3B), and we found that COL1A2 and C3 are common ligands used by SMCs and FBs.

Integrative analysis of cell-cell communication reveals cell-specific druggable targets

To investigate potential pharmacological interventions that may disrupt deleterious intercellular communications, we performed an integrative analysis of cell-cell communication with known druggable genome databases. Key mediators C3, MMP2, and integrins (ITGA1, ITGB1, ITGB5, all expressed by FBs) interact with T-cells, pericytes, and SMCs, and can be disrupted by drugs such as compstatin, tanomastat, SAN-300, volociximab, and cilengitide, respectively (Figure 2.3A). Similar components, such as collagen (COL1A2) and C3, also appear to be significantly expressed in SMCs that could signal to FBs, pericytes, and T-cells (Figure 2.3B). Interestingly, multiple studies have linked C3 and the complement system to atherosclerotic lesion maturation in mouse models (112), and a recent case study showed that compstatin-derived C3 inhibitor (AMY-101) may prevent cardiovascular complications in patients with severe COVID-19 pneumonia (113,114). Further, a recent study demonstrated that microRNA-9 repression of

Syndecan-2 (*SDC2)* impedes atherosclerosis formation (115), while MMP2 alteration also contributes to atherosclerosis in mouse models (116). Here, we show that SMCs communicate with FBs within the atherosclerotic environment via SDC2-MMP2, and reveal additional upstream candidate drug therapies that may influence atherosclerosis progression (Figure 2.3A). In general, these results provide a potential mechanistic explanation by which FBs and SMCs can modulate the inflammatory environment and plaque formation.

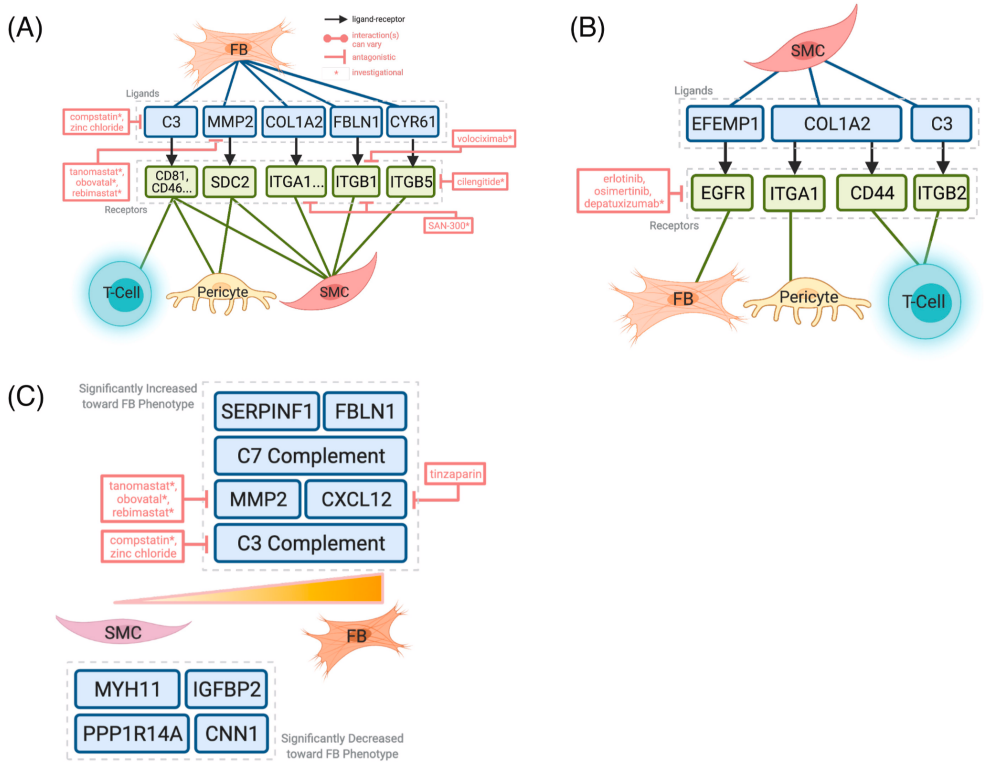


Figure 2.3: Integrative analysis with drug-gene interaction database DGIdb 3.0 reveals potential pharmacological inhibition of SMC and FB signaling and dedifferentiation.

Surprisingly, anti-EGFR (epidermal growth factor receptor)-based cancer treatments such as erlotinib, cetuximab, and gefitinib were identified as potential key mediators of signaling pathways between SMCs and FBs via EFEMP1 (EGF Containing Fibulin Extracellular Matrix Protein 1) and EGFR (Figure 2.3B). EFEMP1

has been suggested as a prognostic marker for atherosclerotic plaque rupture (110). Although the overlap between CAD and cancer etiology has been previously noted, the long-term efficacy and cardiovascular impact of chemotherapy drugs, such as erlotinib, requires further translational studies to investigate their potential use in cancer patients to treat CAD (30,117,118).

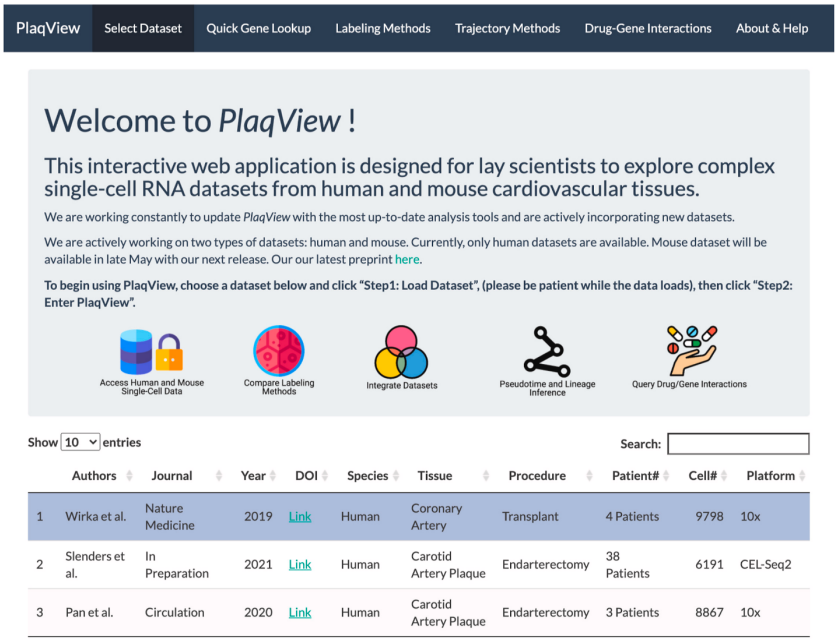
While the exact clinical outcome of SMC de-differentiation is not fully resolved (17,92), such events observed in different single-cell studies present opportunities for pre-clinical pharmacological intervention and testing. Here, we performed integrative analysis of gene expression variation along the SMC-to-FB RNA trajectory with DGIdb. We found that the expression of complement genes such as *C3* and *C7,* and chemokine *CXCL12* are increased as SMCs become more FB-like. Although CXCL12 derived from endothelial cells has been recognized to promote atherosclerosis in mouse models (104), here we provide a potentially new source of CXCL12 using human data and found several pharmacological agents such as tinzaparin, an FDA-approved anticoagulant, to investigate in future interventional studies. Together, this combination of cell-cell communication and trajectory analyses reaffirm the druggable potential of these target genes.

PlaqView **is a user-friendly web application to share and explore atherosclerosis-related single-cell datasets**

To enable other researchers to explore the transcriptomic landscape of the atherosclerotic environment easily and rapidly, we developed a web interface called *PlaqView* (www.plaqview.com, Figure 2.4A). This interactive R- and Shiny-based tool allows for multiple gene queries and comparisons of gene expression, cell-labeling methods such as SingleR, Seurat/TS, and scCATCH, RNA-trajectory tools such as Monocle3 and Slingshot, integrative drug-gene analysis using DGIdb

3.0, and outputs high quality graphs and detailed tables. Calculations are done *ad hoc* according to users' input and the application has been optimized for rapid exploration of these datasets by academic researchers, clinicians, and lay scientists. To our knowledge, there are no publicly available tools to visualize atherosclerosis-related single-cell datasets without prior coding knowledge. Further, *PlaqView* is under active development and will be releasing new datasets coincidently with future atherosclerosis-related publications. At the time of writing, three human and four mouse datasets from four independent studies are available in *PlaqView (Supplemental Table 4)*. *PlaqView* is also open-source and is fully maintained at https://github.com/MillerLab-CPHG/PlaqView. We have written *PlaqView* to be easily repurposed for other single-cell studies; all description files are written in basic R markdown, and the data preprocessing pipeline can be adapted to any tissue from mouse or human. As the database in *PlaqView* expands along with the growing number of single-cell datasets, we anticipate that it will become an essential tool for the atherosclerosis research field. Furthermore, *PlaqView* will serve as a template for other fields of single-cell biology to rapidly disseminate relevant and cutting-edge data with minimal web-development skills required.

(A)



(B)

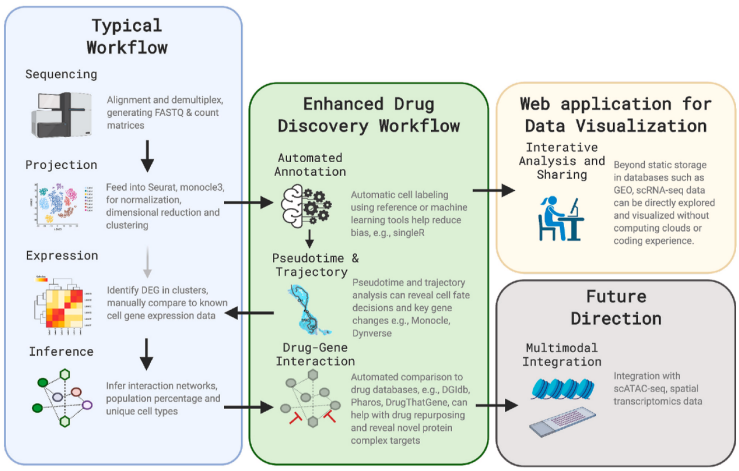


Figure. 2.4: PlaqView and its analysis pipelines are reproducible, user-friendly tools for single-cell data analysis and presentation.

Limitations

Despite the advances presented in this workflow, we are still working to improve several limitations. For instance, the default reference in 'SingleR' and Tabula sapiens (94) do not contain more recently discovered cellular phenotypes such as 'fibromyocytes' (17,83,104,119), which may require a combination of manual and automated labeling methods for accurate identification. Factors such as the intrinsic heterogeneity of the tissue sample, disease stage, and tissue processing artifacts are difficult to isolate computationally and could influence automated labeling methods. Nonetheless, interactive viewers and tools such as *PlaqView*, which incorporate multiple methods and visualizations in one location, could help users separate out the technical and biological variation from various single-cell datasets. Additionally, the modular nature of *PlaqView* will also allow for future improvement of labeling methods as more precise reference datasets are made available.

While we cannot verify the directionality of the RNA trajectories presented, future releases of PlaqView will feature RNA 'velocity' as described in (76) where directionality will be calculated using splice/unspliced RNA ratios from raw data. Lastly, the true efficacy of proposed drugs cannot be verified without extensive pre-clinical testing and clinical trials. Still, these findings may catalyze future investigative efforts to develop more targeted therapies.

2.4: Conclusions

Our findings show that an enhanced, reproducible pipeline for scRNA-seq analysis has the potential to improve upon current standard scRNA-seq bioinformatics protocols. For instance, we provide new insights into intricate vascular cell differentiation and communication pathways while providing actionable and testable targets for future experimental studies (Figure 2.4B). In our combined

analysis, we found that SMCs give rise to a population of FBs that express genetic signatures associated with inflammation and extracellular matrix degradation. For example, SMCs signal to FBs via inflammatory molecules like *C3* complement and *MMP2*, whose expression increases along the SMC-to-FB trajectory. We revealed possible therapeutic avenues that may disrupt these cell-to-cell communications and alter the atherosclerotic pathology. Furthermore, several FDA-approved drugs (e.g., erlotinib, cetuximab, and gefitinib) were shown as potential effectors of SMC signaling to FB, and merit molecular studies to determine whether they may be used to treat CAD in cancer patients to simplify or augment current drug regimens (117). This is consistent with recent reports showing beneficial effects of the acute promyelocytic drug all-trans-retinoic acid (ATRA) in atherosclerosis mouse models (81). Further investment in scRNA-seq may also help resolve the balance of anti-tumor efficacy and atheroprotection for immune checkpoint inhibitors as well as immunomodulators at the interface of cardio-immuno-oncology (49).

Although the utilization of this workflow can compensate for many of the shortcomings of current scRNA-seq analyses, we are still unable to perform cell-lineage tracing that reflects actual timescales without additional gene engineering experiments *in vivo (120)*. However, leveraging mitochondrial DNA variants in snATAC-seq data has enabled lineage tracing analysis in human cells (121,122). Likewise, these analyses can ultimately be extended to integrate spatial omics and other multi-modal data (93). As spatial transcriptomics, scATAC-seq, and/or CITE-seq data become more widely available, this workflow can be easily modified to discover signaling pathways or differentiation events at specific tissue locations and timepoints, allowing for more disease-relevant drug-gene interaction analyses (Figure 2.4B).

Nonetheless, this pipeline can be applied immediately to datasets from other tissues or diseases to generate informative directions for follow-up studies, and is

more user-friendly and reproducible compared to standard scRNA analyses. PlaqView serves as a central repository for interactive analysis and exploration for atherosclerosis-related single-cell datasets. As *PlaqView* incorporates additional relevant single-cell datasets, we anticipate that this application will become an indispensable resource for the community and for the growing field of 'cardioinformatics' (123–125). Most importantly, web applications such as *PlaqView* democratize the access and analysis of single-cell data, which will promote collaboration, reproducibility, and innovation across disciplines (77,126).

2.5: Methods

Data retrieval and pre-processing

Human coronary artery scRNA data read count matrix was retrieved from the Gene Expression Omnibus (GEO) using #GSE131780 and loaded into R 4.1.1, and was preprocessed using standard parameters of the R packages 'Seurat' v.4, and 'Monocle3' as required (75,86,87,89): read count matrix was read into R and converted to a 'Seurat object' using the CreateSeuratObject() function. Then, the object underwent removal of mitochondrial and low-quality reads, followed by normalization and variable feature selection using the NormalizeData() and FindVariableFeature() functions, respectively. Uniform manifold approximation projections (UMAPs) were then calculated using the runUMAP() function using the first 30 dimensions. Custom scripting was created to export UMAP of the clusters from 'Seurat' into 'Monocle3' before pseudotemporal analysis.

Automatic cell Identification

scRNA read matrices were read into ‘SingleR’ as previously described for cell labeling (83). Here, we exported the processed scRNA matrix from Seurat into a ‘SingleCellExperiment’ object using the GetAssay() function, and fed into ‘SingleR’ along with the human primary cell atlas as the reference via the ‘celldex’ package (83). ‘SingleR’ compares each cell’s gene expression profile with known human primary cell atlas data and gives the most likely cell identity independently. ‘SingleR’ first corrects for batch effects, then calculates the expression correlation scores for each test cell to each cell type in the reference, and the cell identity is called based on reference cell type exhibiting the highest correlation. Annotations provided by ‘SingleR’ were then abbreviated for clarity.

To validate the labeling provided by singleR, we tested two additional labeling tools: ‘scCATCH’ (86) and ‘Seurat’ (86,93) using the *Tabula sapiens* data (abbreviated as Seurat/TS). Tabula sapiens is the latest and largest human single-cell reference available to-date (94). For ‘scCATCH’, the findmarkergenes() function was first used to identify marker genes in the query dataset using the minimum percentage of cell expression cut off of 0.25, minimum log-fold change of 0.25, and p-value of 0.05. We then ran the main labeling function scCATCH() and specified the tissue type as ‘Blood Vessel’, ‘Heart’, ‘Myocardium’, and ‘Serum’.

The CZ Biohub recently released the first comprehensive human single-cell blueprint consisting of high-quality sequencing from 25 organs and eight normal human subjects, and a large portion is dedicated to vasculature tissue(94). This atlas is extremely helpful in that many reference-based labeling tools lack the necessary reference for correct tissue types. We obtained this dataset and converted it into a Seurat-ready file using the Convert() function in the ‘SeuratDisk’ package. We then used the FindTransferAnchors() function in ‘Seurat V4’ to find the low-dimensional representations of our dataset, then used the TransferData()

function to label the test dataset. The annotation provided by ‘Seurat’ and the ‘Tabula sapiens’ dataset were abbreviated for clarity.

Pseudotemporal Ordering and Trajectory Inference

Pseudotemporal analyses were performed as previously described in the analysis of embryo organogenesis (89). The processed count matrix was imported into a ‘cds’ object for ‘Monocle3’ using the new_cell_data_set() function, and preprocessed again using the preprocess_cds() function. To have consistent UMAPs across these tools, we used custom scripts to transfer the UMAP calculated in ‘Seurat’ into ‘Monocle3’ before running the main functions ‘learn_graph()’ and ‘order_cells()’ to get trajectory and pseudotime. The starting nodes, or root nodes, were selected in the SMC cluster using an automated function that picks the node most heavily surrounded by ‘early cells. We picked SMCs as the starting cell type based on prior evidence suggesting that SMCs can ‘dedifferentiate’(17,92). The SMCs and related clusters were then subsetted for detailed sub-clustering and analysis. For each cluster, Moran’s I statistics were calculated, which identify genes that are differentially expressed along their trajectories. The Moran’s I statistic helped inform the selection of genes to display in Figure 2.1D. We reviewed the most significant candidates of the Moran’s I calculation and selected 12 to present for clarity. Additionally, we explored 60+ other trajectory inference methods using the ‘Dynverse’ package (99). We first used guidelines_shiny() function to filter out methods that are not suitable for disconnected topography, as in the case of human tissues where cells are developed from multiple progenitors. We then exported the ‘Seurat’ object into a ‘dyno’ object using the wrap_expression() function. infer_trajectory() function was used to calculate the trajectory based on the chosen algorithm, including ‘SlingShot’ (97), ‘SCROPIUS’ (100), and ‘PAGA’ (98). Detailed

codes to reproduce the figures and custom scripts aforementioned can be found at the Miller Lab Github (see availability of data and materials).

Ligand-receptor cell communication analysis

We analyzed candidate ligand-receptor interactions to infer cell communication using the R package ‘scTalk’, as previously described in the analysis of glial cells (106) and CellChat (107). For scTalk, we exported statistically significant differentially expressed genes from ‘Seurat’ using the ‘FindMarkers()’ function and imported the preprocessed data. Then, overall edges of the cellular communication network were calculated using the ‘GenerateNetworkPaths()’ function, which reflects the overall ligand-receptor interaction strength between each cell type. Then, the cell types of interest were specified and treeplots were generated using the ‘NetworkTreePlot()’ function. The final figures were re-rendered in BioRender manually for clarity. For CellChat, the processed Seurat object was fed into the ‘createcellchat()’ function and processed using its standard pipeline. Briefly, the Seurat count matrices were extracted along with the Seurat/TS annotation generated previously. CellChatDB.human was loaded and differentially expressed genes and interactions were identified in the CellChat object via identifyOverExpressedGenes() and identifyOverExpressedInteractions(), respectively. The CellChat algorithm was then run to calculate the probable interactions and pathways via computeCommunProb() and computeCommunProbPathway(). We also ran filterCommunication() to filter out interactions with less than 10 cells in each cell type.

Gene-drug interaction analysis

The above identified ligand and receptor interaction pairs were fed into the Drug-Gene Interaction database (DGIdb 3.0) to reveal candidate drug-gene interactions (111). Ligands and receptors that were deemed significant from 'scTalk' and genes identified in Monocle3 were evaluated using the 'queryDGIdb()' function of the 'rDGIdb' R package (111,127). Additionally, we queried CTD2, OMIM, ClinVar, Pharos, GnomAD, and the ExAC databases using the docker-based tool 'DrugThatGene' (128). We included all top FDA-approved drugs and experimental or investigational drugs with verified inhibitory or antagonistic activities, as well as drugs that may influence or be influenced by changes in the receptor. Figures 3A-D were modified using BioRender for clarity.

Development of *PlaqView*

PlaqView is written in R and Shiny, and is hosted on the web at www.plaqview.com using dedicated DC/OS servers at the University of Virginia, but can be run locally through RStudio. The raw data was first processed as previously described, but packaged into .rds objects. The data were placed into a housing directory within the PlaqView application. Custom scripts were written to recall selected datasets on the user-interface and loaded on-demand. Users-inputs were made interactive using Shiny's reactive scripting practices, and calculations were done *ad hoc.* Packages such as Seurat, CellChat, and singleR are wrapped and rebuilt during each update in the web server. This application is open-sourced and its code and data is available at github.com/MillerLab-CPHG/PlaqView. Additional datasets are actively being recruited and will be made available in future updates.

2.6: Declarations

BBK is the founder of Dock Therapeutics, Inc. The remaining authors have no conflicts of interest to disclose.

2.7: Ethics Approval and Consent to Participate

Not applicable

2.8: Consent for Publication

Not applicable

2.9: Availability of Data and Materials

Data presented in figures 1-4 are publicly available at GEO via accession number GSE131778 as previously described(17). Other data available on PlaqView are available upon request through GitHub. All codes and analysis pipelines can be viewed at github.com/MillerLab-CPHG/Ma_2020 and repurposed for other scRNA datasets. Feature requests for PlaqView can be made at https://github.com/MillerLab-CPHG/PlaqView.

2.10: Competing Interests

BBK is a founder of Dock Therapeutics, Inc. The other authors have no competing interests to disclose.

2.11: Funding

Funding support was provided by grants from the National Institutes of Health (NIH): R00HL125912 (CLM), R01HL148239 (CLM), K12HL143959 (BBK) and T32HL007284 (CJH and DW); American Heart Association (AHA): POST35120545 (AWT) and Leducq Foundation Transatlantic Network of Excellence ('PlaqOmics') (CLM, AWT, GKO, GFA, SWvL, GP, MM).

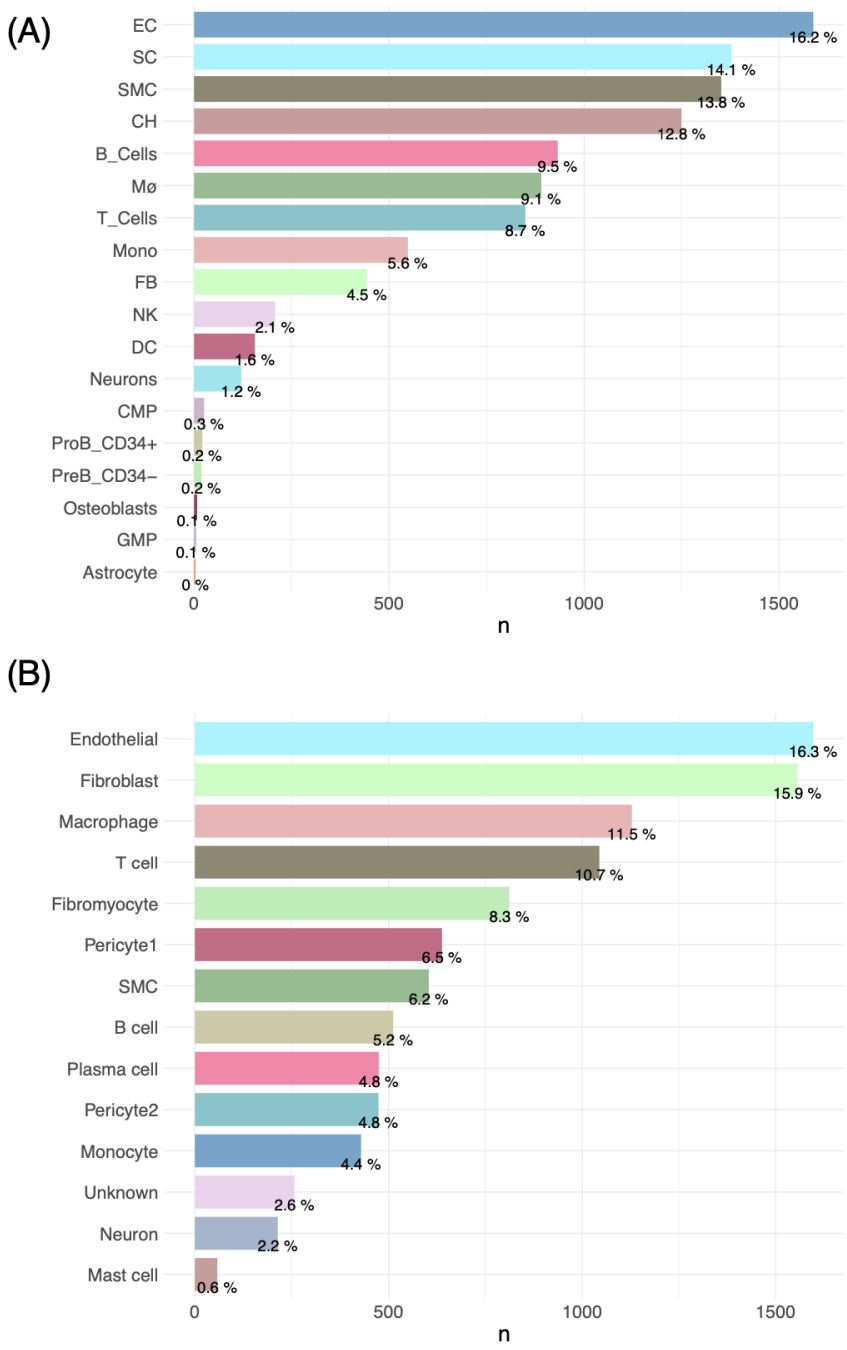
2.12: Authors' Contributions

WFM designed and performed the statistical analysis. CJH, AWT, AVL, BBK, DW and YS refined the methodology and edited the manuscript. JVM, DM, and GFA helped with scripting. AVL, BBK, LS, HZP, NBB, and KO helped with data acquisition and refinement of PlaqView, and edited the manuscript. GKO, MPR, MYL, BBK, GP, MM, and SWvL provided critical feedback on the manuscript. MYL provided statistical review and edited the manuscript. CG provided deployment and application scripting support, and edited the manuscript. CLM and SWvL conceived the project. CLM and BBK refined the project and edited the manuscript.

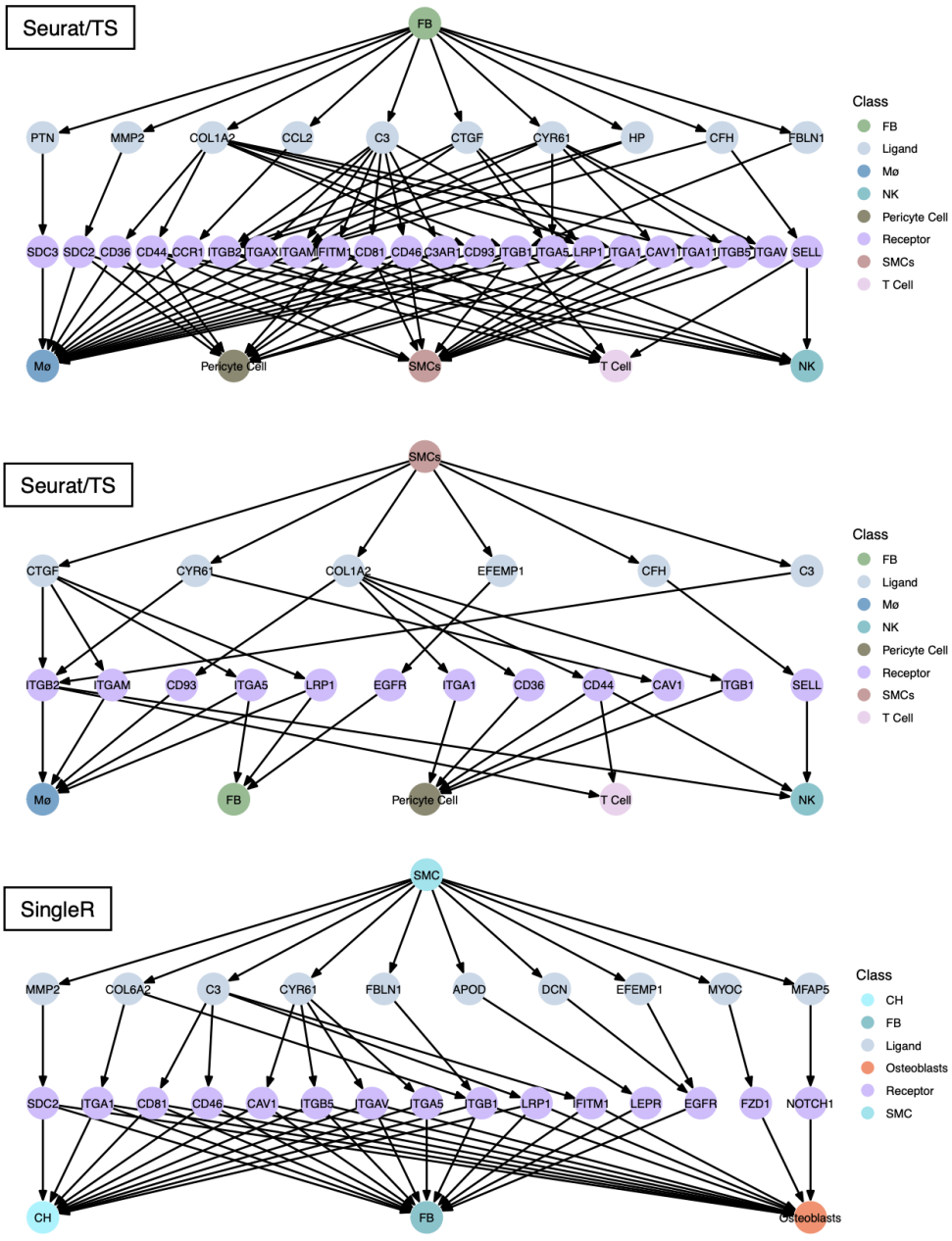
2.13: Acknowledgements

The authors acknowledge Research Computing at The University of Virginia for providing computational resources and technical support that have contributed to the results reported within this publication. URL: https://rc.virginia.edu. The authors also acknowledge Robert Wirka and Thomas Quertermous at Stanford University for providing more details on the coronary artery dataset.

2.14: Supplemental Figures and Tables



Supplemental Figure 2.1. (A) Population percentage when cells are labeled using singleR. (B) Population when using author-provided labels**.**



Supplemental Figure 2.2. Comprehensive ligand-receptor pairs as revealed by ‘scTalk’ when provided with Seurat/TS or SingleR labels. Top: signaling from fibroblasts to macrophages, pericytes, smooth muscle cells, T-cells, or natural killer cells. Middle: signaling from smooth muscle cells. Bottom: signaling from smooth muscle cells but with labels from SingleR. Simplified figures of A and B were presented in Figure 3.3 in the main text.

Supplemental Figure 2.3, and **Supplemental Tables 2.1-4** are available online at the following link:

https://www.atherosclerosis-journal.com/cms/10.1016/j.atherosclerosis.2021.11.025/attachment/ce49374c-f58d-4c15-af48-4c52a15145a9/mmc1.docx

Chapter 3: Construction and Expansion of a Comprehensive Single-Cell Data Portal for Cardiovascular Diseases

This Chapter has been published at *Frontiers in Cardiovascular Medicine* as the following Original Research Article:

Ma, W. F. *et al.* PlaqView 2.0: A comprehensive web portal for cardiovascular single-cell genomics. *Frontiers Cardiovasc Medicine* **9**, 969421 (2022). *https://doi.org/10.3389/fcvm.2022.969421.*

Figure and table numbers have been adapted for this thesis.

3.1 Abstract

Single-cell RNA-seq (scRNA-seq) is a powerful genomics technology to interrogate the cellular composition and behaviors of complex systems. While the number of scRNA-seq datasets and available computational analysis tools have grown exponentially, there are limited systematic data sharing strategies to allow rapid exploration and re-analysis of single-cell datasets, particularly in the cardiovascular field. We previously introduced PlaqView, an open-source web portal for the exploration and analysis of published atherosclerosis single-cell datasets. Now, we introduce PlaqView 2.0 (www.plaqview.com), which provides expanded features and functionalities as well as additional cardiovascular single-cell datasets. We showcase improved PlaqView functionality, backend data processing, user-interface, and capacity. PlaqView brings new or improved tools to explore scRNA-seq data, including New and improved features include: gene query, metadata browser, cell identity prediction, *ad hoc* RNA-trajectory analysis, and drug-gene interaction prediction. We also include 25 new datasets; these new datasets include single-cell atlases, human datasets from other cardiovascular diseases such as aortic aneurysm, and those from mouse experimental models such as atherogenic diets and genetic knock-outs. PlaqView serves as one of the largest central repositories for cardiovascular single-cell datasets, which now includes data from aneurysm, genetic mouse knock-outs, and healthy references. PlaqView 2.0 brings advanced tools and high-performance computing directly to users without the need for any programming knowledge. Lastly, we outline steps to generalize and repurpose PlaqView's framework for single-cell datasets from other fields.

3.2 Introduction

Named “Method of the Year” in 2013, single-cell RNA sequencing (scRNA-seq) technology has now been used in virtually every field of biology and medicine, including cancer biology and cardiovascular medicine (129,130). While scRNA-seq technologies continue to evolve and multi-modal measurements become increasingly more common, there has been a rapid increase in both the number of analysis tools and the complexity of single-cell data. At the time of writing, scrna-tools.org reports over 1,000 tools dedicated to single-cell data analysis (131). Despite the abundance of single-cell analysis tools available, there are two major challenges that affect the single-cell community as a whole: 1) there are no standardized methods of single-cell data sharing, and 2) increasingly complex and large single-cell data require advanced bioinformatic skills and resources for comprehensive analysis and interpretation.

The lack of standardized data sharing leads to the omission of critical metadata needed for reproducible analysis, such as author-defined cluster and cell-type annotations, and variables such as sex and age group (132). Despite the existence of public archives such as the Gene Expression Omnibus (GEO) and Sequence Read Archive (SRA), there is no uniform requirement for the deposition of metadata. Even within SRA, some datasets require specialized cloud computing tools (e.g. Google Cloud Computing) to access, such as the data from Li et al. (133). A recent study has found that less than 50% of the published figures from public single-cell data can be reproduced as published (126). Furthermore, raw sequence files are very large (often >15GB per file) and thus have very high computational resource requirements.

The fact that most tools require users to have significant programming skills remains a significant barrier to reanalyzing and exploring single-cell data, particularly across multiple datasets (75,134). A few tools, such as the cBioPortal

for Cancer Genomics (www.cbioportal.org), have distilled large bulk RNA-seq datasets, made them accessible to everyone, and provided multiple query functions, but none are designed for single-cell datasets (Supplemental Table 1) (77). Other portals, such as DISCO, provide real-time data integration and queries of multiple datasets across different tissues; however, DISCO currently lacks cardiovascular-specific datasets and comparison of analysis tools (Supplemental Table 1) (135). As more datasets emerge, the need for a centralized and domain-specific repository tailored for single-cell cardiovascular datasets becomes more pressing. A centralized repository packaged with tools specific for cardiovascular diseases will facilitate the advancement of the field as a whole and support the democratization and utility of single-cell data.

Previously, we introduced PlaqView, an open-source web portal focused on the exploration of the data generated by Wirka, et al., (17) and a few other atherosclerosis-related datasets (27). Here, we introduce PlaqView 2.0, a significantly improved release with a broader scope to include, among others, datasets from other areas of the cardiovascular field, including human aortic aneurysm (136), healthy human heart atlas (72), human aortic valves (137), mouse models of atherogenesis (31), and others. We describe improvements to the user interface, new *ad hoc* functions to calculate trajectory on cells of interest, metadata exploration, and the ability to export publication-ready figures in addition to existing functions such as basic gene expression query and drug-gene interaction analysis. We also highlight several backend improvements that allow us to bring high-performance, reproducible computing environments to lay scientists via the web browser. Lastly, we outline basic steps to repurpose the PlaqView programming scaffold for other areas of single-cell investigation.

3.3: Results

Plaqview 2.0 Includes Significant Expansion in Data Availability

Since our initial publication introducing PlaqView, which featured 7 datasets from 4 studies (27), PlaqView 2.0 now features 32 datasets: 23 from human tissues and 9 from mouse tissues (Figure 3.1, Supplemental Table 2). New single-cell datasets made available in this release include those from: mouse carotid ligation experiments (138), mouse adventitial cell layer (139), human adult heart compartments (72), human COVID-19 autopsy hearts (140), and human aortic leaflets (137). At the time of writing, PlaqView now contains over 1.7 million total cells. PlaqView will be actively maintained and we will continue to add new datasets upon their publication and release to the public.

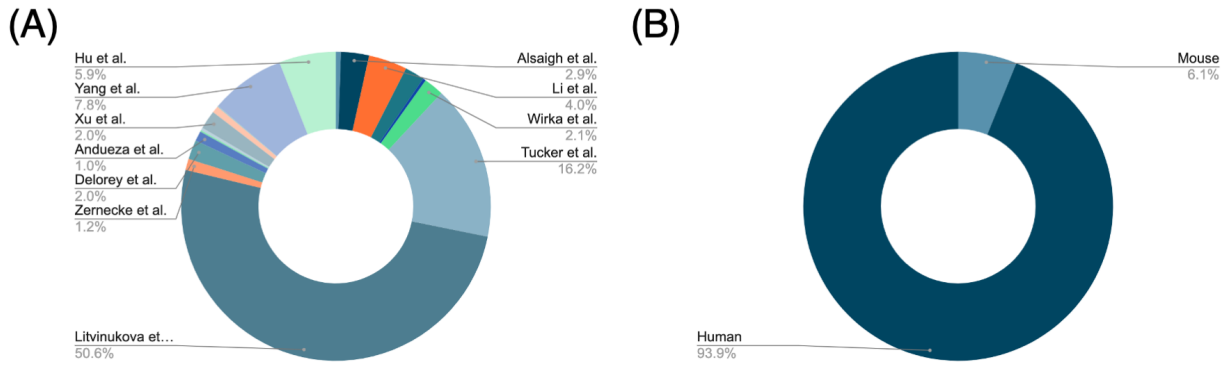


Figure 3.1. PlaqView 2.0 incorporates 32 cardiovascular-related datasets. (A) Major datasets available on PlaqView. PlaqView brings high-performance computing directly to the browser in order to handle large datasets such as ones from Litvinukova et al., which are atlas-type survey data that contain tissues from the entire human heart. (B) Species composition of PlaqView's database.

PlaqView Allows Rapid Query of Gene Expression of Single Cell Datasets

Gene expression-based query is the mainstay for single-cell data exploration. When interacting with single-cell datasets, scientists commonly begin with querying expression levels of one or more genes of interest. However, this is also one of the most time-consuming steps because public data are often shared in various formats that require customized codes to be read and analyzed. In PlaqView 2.0, each dataset has been systematically preprocessed from various stages of upstream analysis to allow efficient querying on-demand. Upon launching the PlaqView homepage (www.plaqview.com), users can open the scRNA-seq portal (Figure 3.2), choose the scRNA-seq dataset of their interest, and load it into the memory of their dedicated session. Manually curated information about each dataset is also available on the portal and on the main homepage. Once the dataset is loaded, users are directed to the "Gene Lookup" tab (Figure 3.3) where they can enter a single gene or multiple genes, select the plot type they prefer (e.g. Dot Plot, Feature Plot, or Ridge Plot), and choose from one of the available cell-labeling methods (see next section) for query. The Gene Query function is powered by Seurat (75). Gene symbol capitalizations are automatically corrected based on species of the dataset selected to conform to Human Gene Nomenclature Committee (HGNC) or Mouse Genome Informatics (MGI) gene nomenclature conventions (e.g. *APOE* for humans and *Apoe* for mice) (141,142). The queried gene(s) plots then will appear alongside the conventional UMAP displaying the cell type, and users can download high-quality, publication-ready figures in .pdf format.

To facilitate functional interpretation within a given single-cell dataset, PlaqView conducts automated Pathway Enrichment Analysis with the queried genes, powered by EnrichR (143–145). Although not commonly a part of a standard single-cell RNA-seq pipeline, pathway enrichments provide additional insights into the biological importance and functional consequences. Users can choose their

preferred databases from a list of well-annotated sources such as ENCODE (Encyclopedia of DNA Elements) and GO (Gene Ontology). For example, querying the two NADPH oxidase genes implicated in redox metabolism and atherosclerosis “CYBB” and “NOX4” (29,146), in the Li et al. (2020) dataset using the “GO_Biological_Process_2018” database shows the top function category as “superoxide anion generation,” and that (FB), macrophages (Mø), and smooth muscle cells (SMCs) highly express these genes (Figure 3.3). This information allows users to quickly assess and generate hypotheses and aid in the design of future experiments.

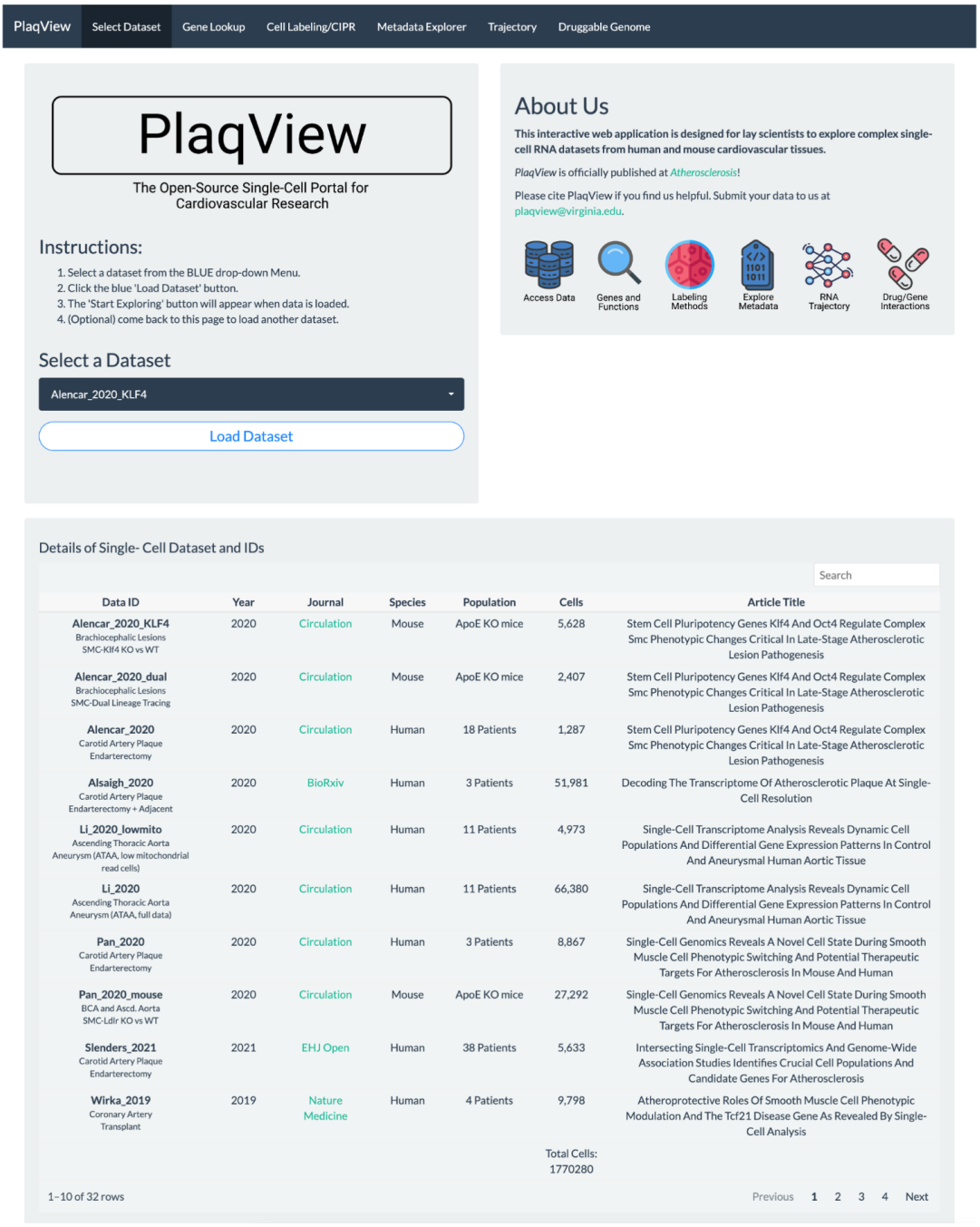


Figure 3.2. PlaqView 2.0 homepage facilitates selection and loading of relevant datasets. The homepage allows users to browse dataset details and load selected datasets into their dedicated session for further exploration.

Figure 3.3. Gene Query page allows rapid visualization of gene expression, UMAP embeddings, and facilitates automated GSEA. PlaqView 2.0 supports

visualization of a single gene or multiple genes via feature, dot, and ridge plots, as powered by Seurat. Users can also choose their preferred annotation methods and download high-quality pdfs for publication. This instance is a demonstration using Li et al. dataset.

Cell Labeling and Differential Gene Expression

Cell identity prediction remains one of the most time-intensive and critical steps in the single-cell analysis pipeline (147). Previously, we found that the upstream cell-state prediction step greatly affects downstream analysis such as cell-cell communication analysis (27), and multiple labeling methods should be compared for consistency. Automated labeling tools are great starting points and help eliminate the inherent bias introduced in cluster-based manual annotation (147). However, current references may not accurately predict novel cell types or transitional states, such as the ‘myofibroblasts’ as shown in Wirka et al. (17). Furthermore, discovery of novel niche cell-types are difficult and require careful examination of the differential gene expression patterns.

PlaqView enables users to run and compare several methods of cell annotation and compare them against other databases, as well as exporting the entire differential gene expression matrix for manual exploration of rare cell types. In many areas of the application and specifically under the “Cell Labeling” tab, users have the ability to explore the cell identities as provided by the original authors (when available), by SingleR (83), and by Seurat V4 “label transfer” using the Tabula Sapiens atlas (75,148). To demonstrate this, we explored the Livinukova et al., annotation (Figure 3.4A) against the annotation provided by Seurat’s label transfer using the Tabula Sapiens reference (Figure 3.4B) (75,148). These identity predictions are pre-run during the data processing stage and stored within each data object. By

running the cell identity prediction during the preprocessing stage (see Methods) and not on-demand, PlaqView can rapidly display results with little downtime. To allow additional flexibility, transparency and further exploration of the differentially expressed genes in each cell type, we provide precomputed tables of the differentially expressed genes based on labeling methods in downloadable .csv format. By precomputing these tables during the preprocessing stage, we cut down hours of computing time for the end-user. These tables provide users the opportunities to review genes or cell groups of interests, and serve as starting points for downstream analysis such as drug-gene interaction analysis.

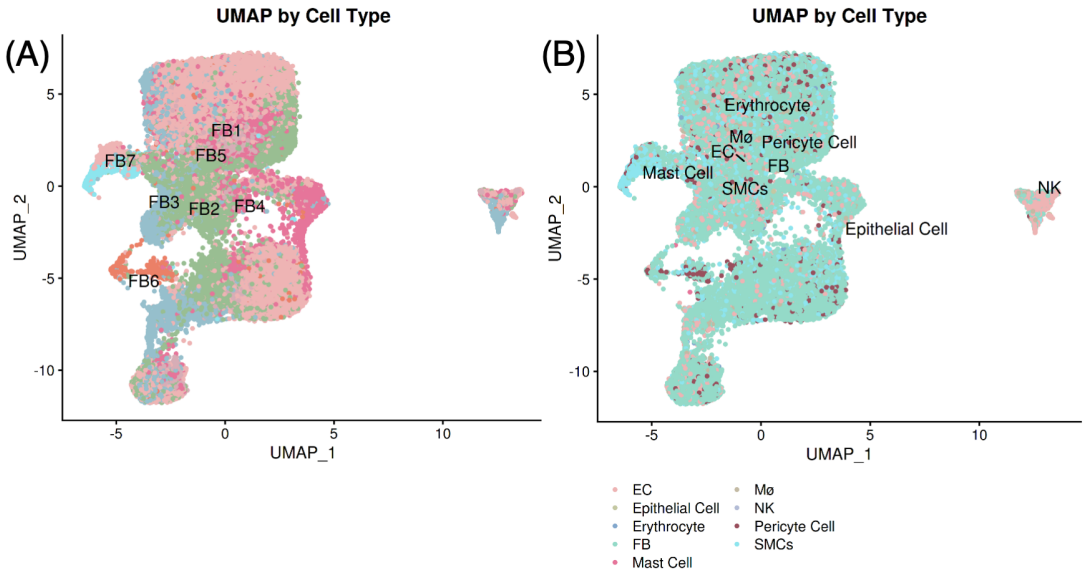


Figure 3.4. Annotation Explorer page facilitates cell-level annotation comparison. Users can compare pre-computed cell identity annotations as well as author-provided labels (when available). Using the Litvinukova et al. dataset, we demonstrate the difference between the annotation from the (A) original authors and (B) Seurat label transfer using the Tabula Sapiens reference.

In PlaqView 2.0, we have incorporated a new interactive feature named Cell Identity PRedictor, or CIPR (Figure 3.5) (149). CIPR provides an additional

opportunity for users to interact with the data and further explore and compare the cell annotations. CIPR calculates, in real-time, cluster-based gene expression similarity index scores against known references, such as the Database of Immune Cell Expression (DICE), Immunological GenomeProject (ImmGen), and the Human Primary Cell Atlas (149). CIPR provides an additional opportunity for the users to interact with the data and further explore and compare the cell annotations. To run CIPR on the loaded dataset, users need to select the starting labeling method (default is the unlabeled Seurat clusters) and the reference to benchmark against (default is ImmGen Mouse). CIPR was designed to be able to run against human and mouse references interchangeably. PlaqView will output an interactive CIPR plot where users can select the cluster(s) of interest and explore the top similar cell types, their descriptions, identity scores (calculated as a function of fold-change dot-product) and percent of genes that are similarly co-expressed. For example, when running the COVID-19 heart autopsy data from Delorey et al. (140) using the pre-sorted human RNAseq reference provided by CIPR, we noted a high concordance between the author-labeled “CD8+ T-cells” with the reference cluster “Effector memory CD8 T cells”, with a 75.5% percent positive correlation in gene signature. Lastly, Users can download a full table of the CIPR results in .csv format as well as the CIPR plot in .pdf format for publication.

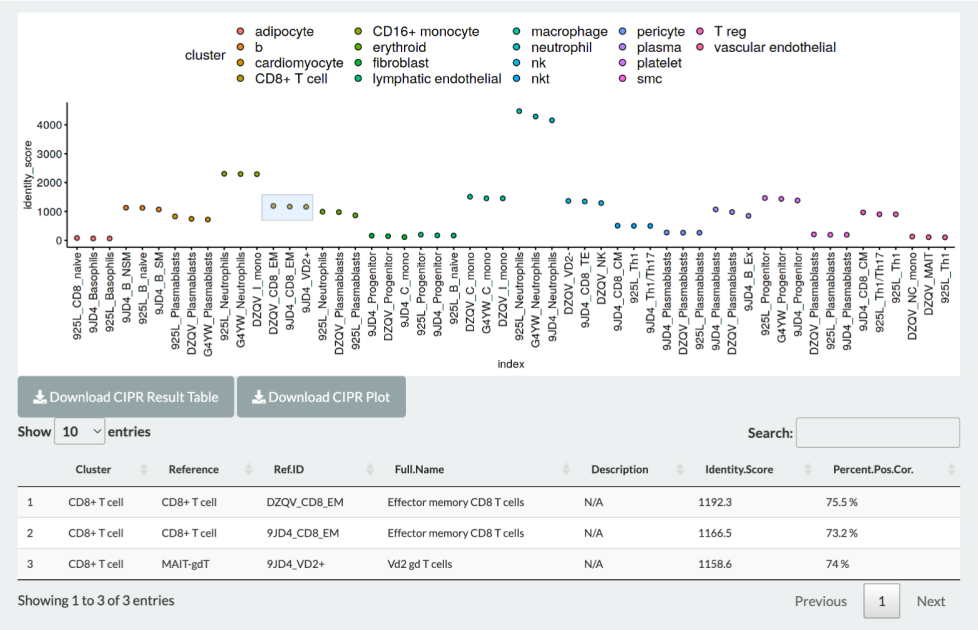


Figure 3.5. CIPR integration allows further exploration and interaction with cell-level annotation. Users can benchmark pre-computed annotations with existing single-cell references, and interactively compare and visualize identity scores and percent correlations with the top candidate reference identities. Light blue box indicates selected groups for detailed tables.

PlaqView Enables Users to Explore Unstructured Metadata

Currently, no systematic convention exists in sharing single-cell metadata that are essential for reproducible analysis and future meta-analysis. These important metadata, such as sex, age, sample location, and author-provided cell-type annotations, are often omitted when submitting data to public repositories. Currently, it is estimated that fewer than 25% of current single-cell studies have provided cell-level metadata (132). PlaqView 2.0 aims to provide a platform for easier standardization and sharing of cell-level metadata in three ways: 1) we curate and reformat existing metadata and append them into the Seurat object, 2) we

require new submissions to have pre-embedded metadata, and 3) we developed features to explore all existing available metadata in their unabridged format.

When available, PlaqView separates the metadata into “Factor-Type” (Figure 3.6A) and “Continuous-Type” variables (Figure 3.6B). Examples of factor-type metadata include sex, age-group, biological individuals, and disease type, whereas continuous-type metadata include percent mitochondrial reads, age, and p-values of singleR annotations. PlaqView will output appropriate feature maps when these data are available. Furthermore, we introduce the ability to query gene expression based on factor-type variables. We demonstrate its use by querying the genes *APOE*, *COL1A1*, *FBLN1*, and *FBLN2* in the Tucker et al. (150) dataset using the “chamber” variable (Figure 3.6C). Interestingly, fibulins (*FBLN1* and *FBLN2*) are more highly expressed in the right atrium (RA) and left atrium (LA), compared to ventricular samples, most likely due to the differential behavior of atrial and ventricular fibroblasts (151). However, further interrogation using the “experiment” and “biological.individual” variables show significant variation of fibulin expression among the atrial samples as well as among individuals sampled (Supplemental Figure 3.1A and 3.1B). This particular example demonstrates the critical need for better metadata sharing as well as the utility of the metadata explorer feature.

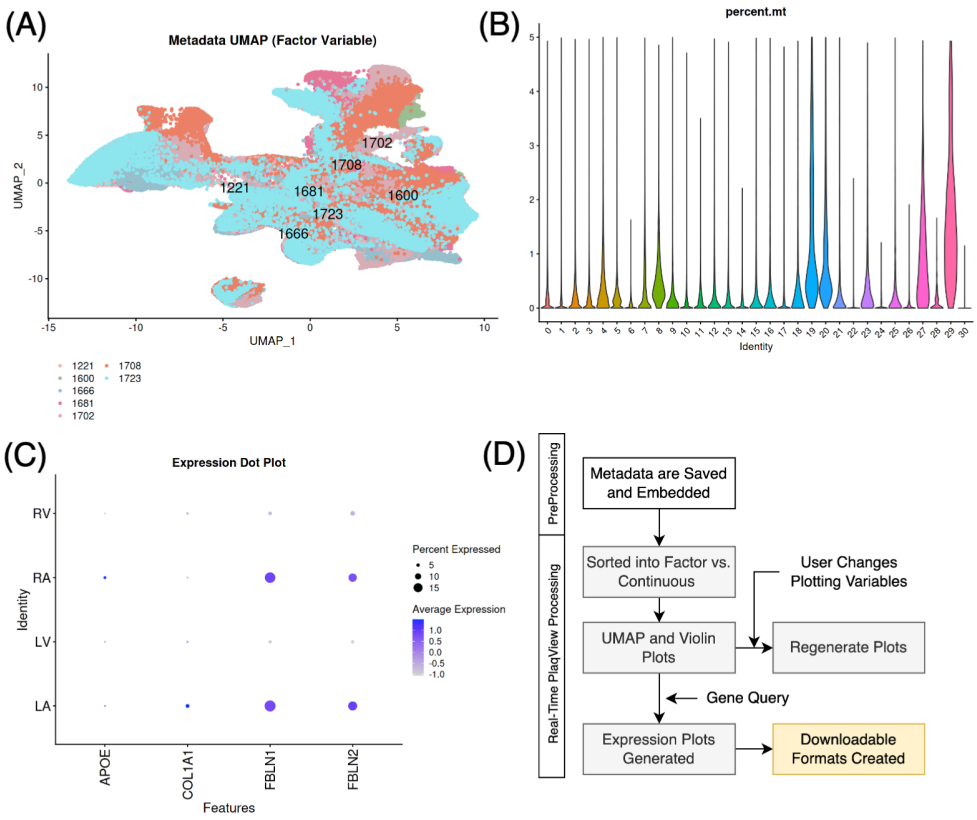


Figure 3.6. Metadata Explorer enables visualization and query of unabridged cell-level metadata. When available, cell-level metadata are divided into (A) factor-type such as cells separated by “biological_individuals,” and (B) continuous-type such as percent mitochondria by seurat cluster for visualization. (C) Users can also query gene expression based on factor-type metadata such as the heart “chamber” in the case of Tucker et al. (150). (D) We embed the metadata during the preprocessing stage and PlaqView sorts the metadata and runs calculations in response to the user selection to generate corresponding UMAPs and Violin plots.

Cell Trajectories and Re-Clustering

RNA trajectory analysis, in conjunction with pseudotemporal ordering, has been widely used as a method to reconstruct cell fate, differentiation, and transition events (22,27,89,152). In terms of cardiovascular data, RNA trajectories have been useful in studying the fate and transitions of fibroblasts, immune cells, smooth muscle cells and other intermediate cell types (22,27,153,154). In PlaqView 2.0, we provide additional functionality in RNA trajectory estimation with Monocle3 (89) by integrating high-performance calculation steps directly within the browser. Upon opening the “Trajectory” page, the full pre-calculated RNA trajectory using the entire dataset is displayed. Now, users can select cells of interest and subset these cells for re-clustering and re-calculation of their RNA-trajectory (Figure 3.7A-C). This interactive feature allows supervised input and selection of relevant cells that can help reduce trajectory “noise.” Furthermore, this feature is helpful because RNA trajectories may not always be applicable to every cell type in the study, such as those that are post-mitotic or slow-dividing. Depending on the number of selected cells, *ad hoc* calculation of RNA-trajectory can take up to 10-15 mins, which is longer than a typical shiny application timeout rule (Figure 3.7D). We made custom services rules in the backend container to ensure that RNA-trajectories are calculated in the most time-efficient manner possible and modified any timeout rules typically applied to reliably deliver RNA-trajectory results. As future datasets get larger, we will implement a notification and ticketing system to allow users to return to the same session when calculations are complete. This new interactive feature allows users to focus on their cell types of interest to better develop future hypotheses and experiments.

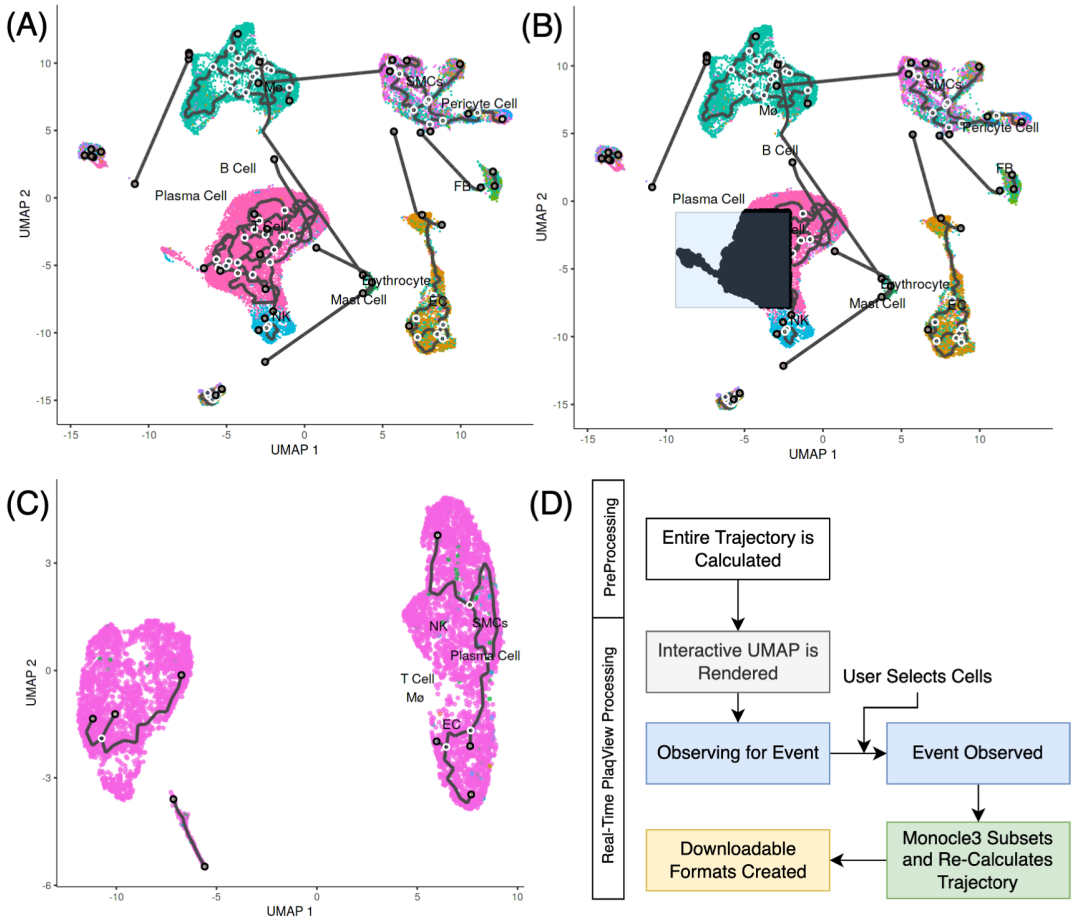


Figure 3.7. PlaqView 2.0 enables interactive RNA trajectory inference. (A) Full cell trajectory inferences are presented to the user, as demonstrated using the [Alsaigh](#) et al. dataset. (B) Selected cells are highlighted in black. These cells can be subsetted and their (C) trajectories can be re-calculated in PlaqView as powered by Monocle 3. (D) We pre-compute the overall trajectory during the pre-processing stage, and recalculations of subset trajectories are done *ad hoc* as needed by PlaqView as an on-demand function.

Druggable Genome and Cell Targeting

Lastly, to enable researchers to rapidly explore current drug databases in the context of relevant single-cell datasets, we integrated the Drug Gene Interaction Database (DGIdb v4) with PlaqView (111,127). In the “Druggable Genome” tab, users can input a gene of interest to simultaneously query the gene expression within the loaded single-cell dataset as well as display potential drug interactions (Supplemental Figure 3.2). Now, users can select multiple databases including Catalog of Somatic Mutation in Cancer (COSMIC), Food and Drug Administration (FDA), and DrugBank as described in DGIdb v4 (155). Additionally, users can download the corresponding UMAP as .pdf as well as the full drug-interaction table as .csv format. This feature is invaluable in rapidly formulating hypotheses and future drug-repurposing experiments.

Goals and Future Updates

We are committed to bringing the most updated and relevant cardiovascular datasets to PlaqView. Currently, we plan to update the PlaqView database at least once monthly as more datasets and studies are published. Furthermore, we are currently developing other multi-modal portals suited for data such as single-cell Assay for Transposase-Accessible Chromatin (scATAC-seq), spatial RNA-seq, and a separate portal to compare healthy and diseased tissues in a systematic manner.

3.4 Material and Methods

Data Formatting

One of the major challenges of reproducible single-cell analysis is the lack of standardization in data-sharing format. We have found that the most commonly used methods are: 1) sharing processed data as matrices (e.g. one file for counts, one file for metadata, etc.), 2) sharing raw FASTQ files, 3) sharing processed Seurat objects as .rds files or equivalent, such as h5ad files. We have found that sharing single-cell data as .rds or equivalent is the most convenient and reproducible method as the metadata are matched at the cell-level. However, as noted above fewer than 25% of published single-cell RNA-seq studies provide this cell-level metadata.

For PlaqView, each dataset curated or submitted is standardized and is ready to be read by the application. Although some efforts have been made to allow interconversions between file formats, such as sceasy (https://github.com/cellgeni/sceasy), manual effort is still required to standardize analysis input for PlaqView. Depending on the incoming file type, they are converted or updated into the latest Seurat object class in R, and are saved as .rds files. These processed files, along with their raw formats, are made publicly available on the PlaqView homepage at https://plaqview.com/data. Systematic processing script is available in in the PlaqView DataProcessing Github page (https://github.com/MillerLab-CPHG/PlaqView_DataProcessing) and can be used as a reference for other areas of single-cell application to readily convert datasets into R-readable format.

Data Processing Steps

Once the datasets are converted into the raw Seurat objects, we process them in the same manner to generate several output files to be read by PlaqView (Figure 3.8). First, we filter out low quality cells that have <200 or >2,500 features, and those with >5% mitochondrial reads. Exceptions are made for particular datasets that evaluate mitochondrial read data, such as data from Li et al. (136). Then, standard Seurat preprocessing using the following functions are conducted: FindVariableFeatures(), NormalizeData(), ScaleData(), RunPCA(), RunUMAP(), FindNeighbors(), and FindClusters().

To infer cell identity using automatic methods, the scaled RNA matrix is extracted using the GetAssayData() function and fed into SingleR as a new singleR object using the SingleR() function. The identities called by singleR are added into the metadata slot within the Seurat object.

Similarly, we use Seurat's label transfer function FindTransferAnchors() and TransferData() to predict cell identity using the Tabula Sapiens and Tabula Muris (156) references, depending on the original species. Identity calls are added to the metadata column using the AddMetadata() function. Some longer cell labels, such as "Smooth_muscle_cells" are shortened to "SMC." The final Seurat object that contains the SingleR calls, Seurat calls, Seurat clusters (numbered clusters) are exported as an .rds file (Figure 3.8).

To enable trajectory analysis, we used custom scripts to extract Seurat data and place them into a Monocle3 CDS object using the new_cell_data_set() function. The new CDS object is preprocessed using the following functions: preprocess_cds(), reduce_dimension(), cluster_cells() (89,152). A custom script is used to overlay the Seurat UMAP embedding into the Monocle3 object for consistent visualization. Then, starting nodes were selected automatically based on

the closest vertex followed by learn_graph() and order_cells(). The resulting monocle3 object was exported as an _cds.rds file (Figure 3.8).

Finally, we compute the entire differential expression for all cell types labeled by different methods (i.e. Seurat clusters, author-provided, singleR, and Seurat/Tabula Sapiens annotations) using the FindAllMarkers() function. This is the most time-consuming and memory intensive part of the preprocessing pipeline and we use the Future package to parallelize this step. The full R script for the preprocessing pipeline is located in our Data Processing Github page (https://github.com/MillerLab-CPHG/PlaqView_DataProcessing).

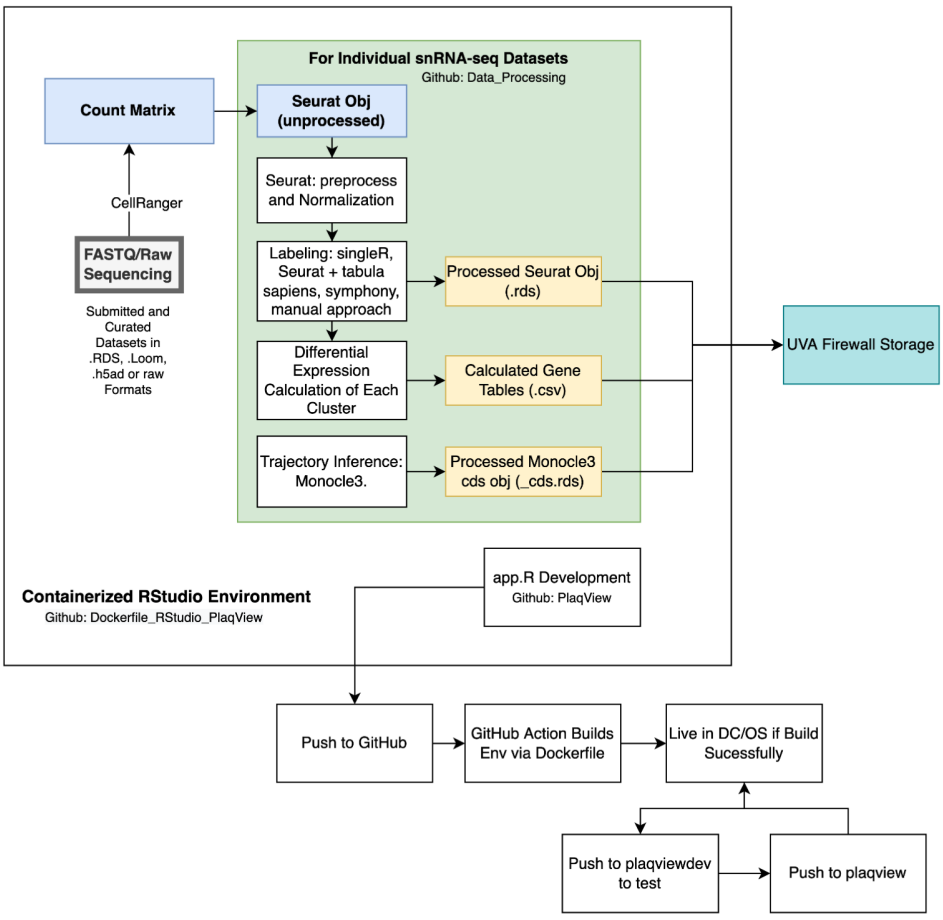


Figure 3.8. Overview of data processing and programmatic strategy for PlaqView. Data submitted to PlaqView are processed systematically and stored as Seurat .rds

objects. These objects, along with calculated differential gene expression tables and trajectories, are stored in secured storage provided by the University of Virginia. Additionally, app development and data processes are all conducted in a cataloged, stable Docker RStudio environment that is registered both on GitHub and DockerHub.

Data Storage, Submission and Requests

Raw human sequencing data often require specialized secured storage both due to their size and institutional review board (IRB) compliance regulations. Currently, PlaqView 2.0 only requires the downstream count matrices and does not use any raw sequencing. When data are submitted to PlaqView as raw sequences (such as FASTQ or BAM files), they are processed offline in a dedicated high-performance computing platform and only the count matrices are transferred to PlaqView storage, which is protected under an institutional firewall (Figure 3.8).

Nonetheless, most datasets currently on PlaqView are already open-access and are deposited in different public storage spaces in various raw formats. All datasets available on PlaqView have undergone systematic preprocessing and are saved in .rds formats that can be requested directly on PlaqView.com under “Data.” Furthermore, researchers can directly and securely submit their dataset to PlaqView on the PlaqView homepage.

Reproducible Computing Environment

Various approaches in computer science and within the R community have been used to create reproducible and stable computing environments to facilitate faster new user setup, scalability for larger datasets, and increased stability of web applications. To our knowledge, popular built-in tools such as *renv* does not

completely enable a reproducible environment and only records the versions of tools in an R computing environment (https://rstudio.github.io/renv/articles/renv.html). Recently, the combination of Docker and R gave rise to the Rocker Project (https://www.rocker-project.org/), which utilizes the image building capability of Docker in conjunction with R. Essentially, a predetermined set of R tools are installed on top of a basic operating system, such as Ubuntu. This enables programmers to capture the entire computing environment- including the base operating system, R, and all dependent packages- and can be downloaded to any computing platform reliably and in isolation. This has an advantage over other tools such as *renv* in that Docker images contain the actual packages and operating system, therefore in the event of version changes or deleted repositories, it will not break with the computing environment. Based on Rocker, we have built a custom Docker image that enables users to run the exact analysis pipelines with all dependencies already installed, and is available via DockerHub at millerlab/plaqviewmaster. This Docker image has two major utilizations: as a stable environment for feature development and as a base for the PlaqView application deployment, as it contains all the necessary packages and serves as a backbone for the web application (Figure 3.8).

Service Structure Overview

PlaqView is designed to enable researchers to interact with high-performance computing via the web browser. The superstructure of PlaqView is to translate user input and selection from the browser via Shiny and run calculations in R, which runs in a clustered container orchestration environment alongside the University of Virginia High Performance Computing system (Supplemental Figure 3.3A). To enable user interactions with the data without coding knowledge, each interactive element in PlaqView is coded as Shiny Reactive

elements, which change the underlying R code snippets in preparation for the analysis. For example, when a user selects a dataset, the corresponding values of the working directory is changed to the selected value, and when the user clicks “Load Dataset”, Shiny monitors and triggers the event to execute the R code to load the .rds files within the working directories.

Currently, we are using DC/OS (Distributed Cloud Operating System) to regulate the amount of memory and processor each user can access. and DC/OS also automatically handles the workload demands to scale up more service instances in response to increased user access. Furthermore, this infrastructure allows for isolated, or “sticky,” instances so each user is given a dedicated R instance and cannot access other users’ instances. This is commonly referred to as “stateless programming.” Currently, PlaqView supports up to 32GB of memory per user (the memory required to access the largest dataset, and can be scaled up as needed). In the near future, we plan to further scale up using Kubernetes on large commercial-grade infrastructures such as Google Cloud Computing.

Development Workflow

Typically, PlaqView development occurs in several stages: edit source code (app.r file), test locally and update Docker images, deployment, and bug fixes (Supplemental Figure 3.3B). The standard development workflow starts with editing the app.R script, which encompasses the UI (user interface) and the Server codes (codes that calculate and compute results). New features and codes are tested locally in the Dockerized RStudio container for bugs. Final edits are pushed to GitHub as “commits” and new changes initiate GitHub Actions to recompile the application from the base Docker image. Normally, this step involves reinstallation of the base operating Linux system and its dependencies, R, Shiny, and all

dependent R packages from scratch, and typically takes about 60mins. However, at this stage, we have simplified the building process by pulling the aforementioned pre-built Docker base image. In our experience, the typical rebuild takes about 60 minutes whereas pulling the stable base Docker image from DockerHub takes only 3-4 mins.

Once the app is live, we begin to capture user feedback and fix any additional bugs. We implement bug fixes and feature requests via GitHub as well as through internal runtime log reviews.

Alternative Service Structure and Adaptation to Other Fields of Research

PlaqView was originally developed for atherosclerosis-related cardiovascular datasets, but the underlying structure was designed to be easily adapted to other fields of research. The entirety of the source-code has been made available on GitHub. Furthermore, each iteration of PlaqView comes with a containerized base image hosted on DockerHub (wfma888/plaqviewmaster), which allows for immediate and reproducible deployment in virtually any computing structure. Essentially, there are two major steps to adapt PlaqView to other fields: minimal processing of the user interface script and deploying to a suitable service structure. are the two major steps to adapting PlaqView to other fields.

Due to the size of the single-cell datasets, PlaqView and adapted versions are best hosted on dedicated, large high-performance clusters. Commercial solutions that would work well are Google Cloud Services as Google Run instances or Amazon Web Services. Shinyapps.io also provides a native and easy way to deploy Shiny apps for beta testing, however limitations in memory per instance and slow performance limits its usefulness in analyzing large datasets.

3.4: Discussion

To date, very few single-cell portals like PlaqView exist for cardiovascular genomics research. ExpressHeart, a single-cell portal dedicated to non-cardiomyocyte cells, has several single-cell datasets but has limited scope and features for data re-analysis (Supplemental Table 1) (157). Other, larger data portals such as the Broad Single Cell Portal (SCP) have many large studies but failed to include many critical cardiovascular datasets such as Wirka et al. (17) and Xu et al. (137), and lack the focus on cardiovascular diseases in general. PlaqView aims to bridge the gap between large, multi-organ portals like SCP and niche portals such as ExpressHeart and serve as a critical resource for the cardiovascular field.

PlaqView helps overcome many modern challenges in the single-cell field, such as the complex coding and computational knowledge needed to explore single-cell data and standardization for data sharing. Since its initial release, we have registered users from 35 countries, with the top being U.S., China, Germany, and the Netherlands. To our knowledge, PlaqView is the most comprehensive single-cell portal dedicated to cardiovascular research. We are committed to the longevity of PlaqView and are working on furthering PlaqView's capability as multimodal datasets are released, such as spatial and scATAC-seq data. Our immediate goals are 1) including additional relevant single-cell datasets, 2) creating a subportal for live single-cell dataset integration and comparison, and 3) creating a subportal for multimodal single-cell data visualization.

3.5: Contribution to the Cardiovascular Field

Single-cell data has always been challenging to share, analyze, and visualize. Typically, these datasets require specialized computing knowledge and high-performance computing tools not readily available. Previously, we presented PlaqView, a web-portal to allow lay scientists and benchtop researchers to rapidly view single-cell RNA-seq data for atherosclerosis. Here, we introduce PlaqView 2.0, a significant improvement to the PlaqView application. In this second major release, we introduce many new features, such as a metadata explorer, cell identity prediction, and *ad hoc* RNA-trajectory calculations. We further improved the usability, speed, stability, and scale of the application. For example, we significantly expanded the scope of data to include data from other areas of cardiovascular disease, such as human aortic valve stenosis and aortic aneurysms, and high fat-induced atherosclerosis in mice. PlaqView serves both as a repository of single-cell data for cardiovascular diseases as well as a tool to rapidly visualize, renalyze, and share scRNA-seq data without the need to code or have specialized computing tools. PlaqView is an invaluable resource that bridges the gap between computational and experimental research to advance cardiovascular medicine.

3.6: Data Availability Statement

All data described herein are made available on the PlaqView data portal (www.plaqview.com) with links to the respective datasets and their DOIs. All code is also made available on GitHub (https://github.com/MillerLab-CPHG/PlaqView).

3.7: Ethics Statement

All research described herein complies with ethical guidelines for human subjects research under approved Institutional Review Board (IRB) protocols at the University of Virginia.

3.8: Author Contributions

WFM designed, tested, and wrote the PlaqView application. AWT provided feedback, programming help and application testing. CG provided back-end infrastructure and deployment aid. DW, YS, JVM, GA, CJH, AP, provided feedback and application testing. SWvdL and CLM guided and helped in the development of PlaqView. HAE developed the CIPR module. All authors provided feedback and editing for this manuscript.

3.8: Funding

Funding for this research was provided by National Institutes of Health (NIH) grants R00HL125912, R01HL14823, and a Leducq Foundation Transatlantic Network of Excellence (‘PlaqOmics’) Young Investigator Grant, Netherlands CardioVascular Research Initiative of the Netherlands Heart Foundation (CVON 2011/B019 and CVON 2017-20: Generating the best evidence-based pharmaceutical targets for atherosclerosis [GENIUS I&II]), and the ERA-CVD program ‘druggable-MI-targets’ (grant number: 01KL1802). Dr. Sander W. van der Laan is funded through EU H2020 TO_AITION (grant number: 848146).

3.10: Disclosures

Dr. Sander W. van der Laan has received Roche funding for unrelated work.

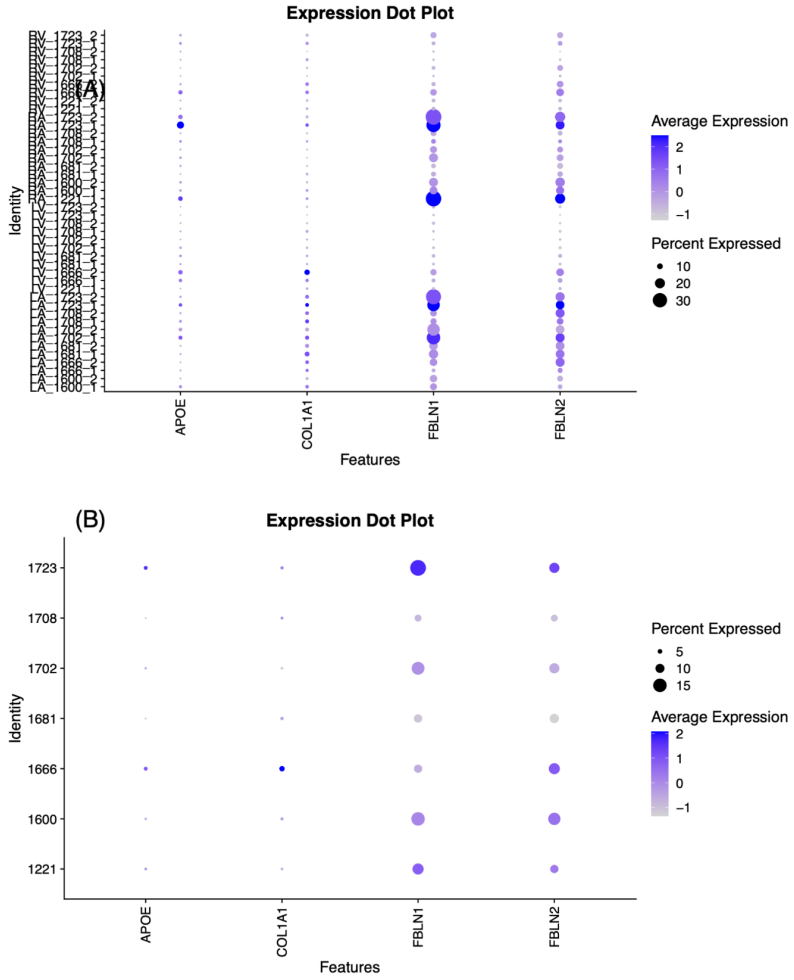
3.11: Conflict of Interest

Authors declare no conflicts of interests.

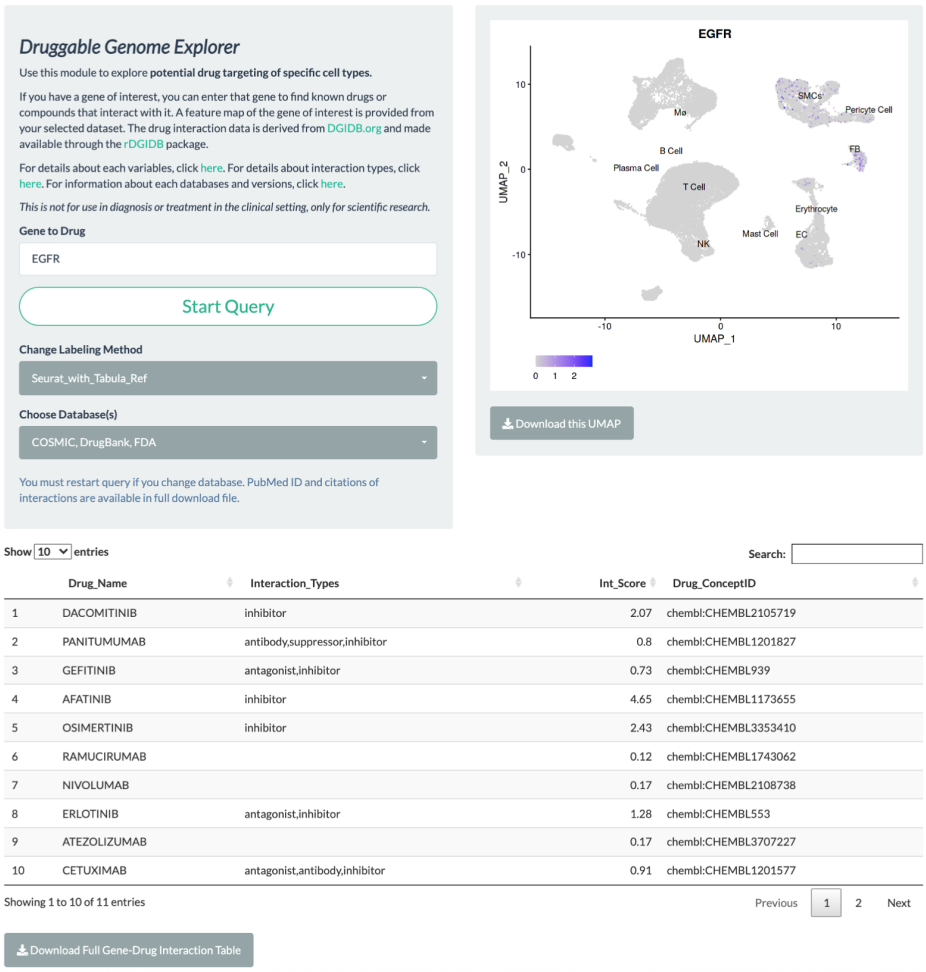
3.12: Acknowledgements

We acknowledge the Research Computing group at the University of Virginia for their helpful suggestions and assistance in deploying PlaqView.

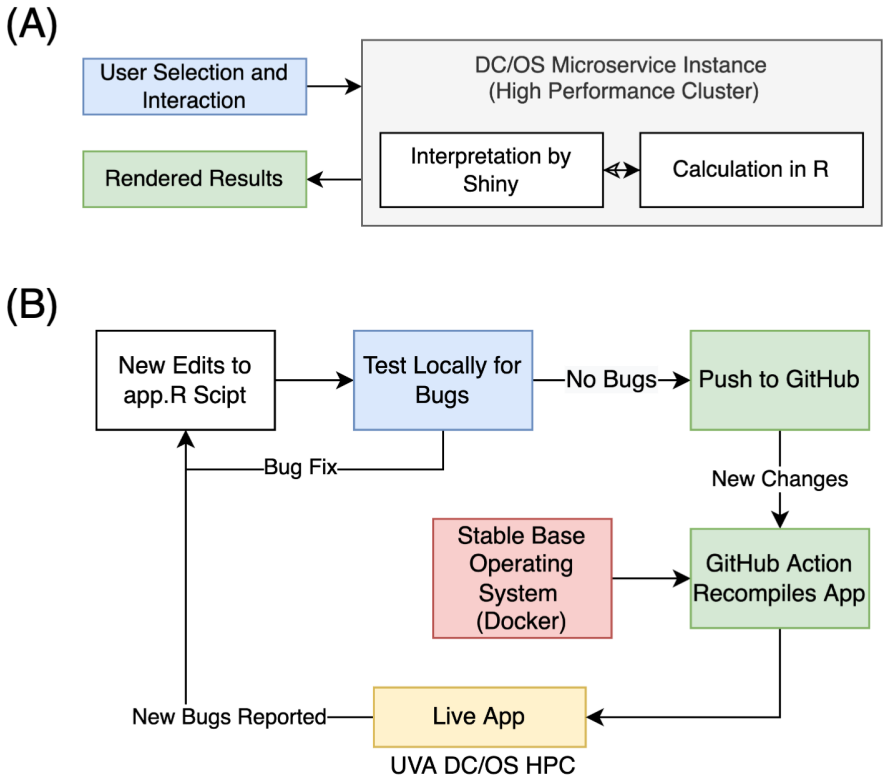
3.13: Supplemental Figures and Tables



Supplemental Figure 3.1. *APOE*, *COL1A1*, *FBLN1*, and *FBLN2* expression in Tucker et al. dataset. Expression data is divided by the metadata column **(A)** "experiment" and **(B)** "biological.individual."



Supplemental Figure 3.2. Druggable genome facilitates drug-repurposing and targeting of specific cell populations. PlaqView 2.0 integrates the latest data from DGIdb to enable side-by-side queries of gene expression and drug-gene interaction. Users can download full drug-interaction tables as well as publication-read expression feature plots.



Supplemental Figure 3.3. Superstructure of PlaqView and application lifecycle. (A) PlaqView enables user interaction with underlying R code using Shiny. When users interact with PlaqView, our high-performance computing cluster at UVA interprets and calculates the output. The resulting outputs are rendered and formatted by Shiny. (B) when new features are being developed or bugs reported, we edit the main app.R script, which we then test locally and push to GitHub. This triggers a rebuild action using the base Docker image and new changes are built on top and new builds are made live automatically.

Supplemental Tables 3.1 and 3.2 can be found Online at the following web address (shown as Supplemental Table 1 and 2 on website):

https://www.frontiersin.org/articles/10.3389/fcvm.2022.969421/full#supplementary-material

Chapter 4: Investigating Cell-Specific Responses in Chemotherapy-Induced Cardiomyopathy

This Chapter is a work-in-progress and contains work that has not been previously published.

4.1: Brief Background and Rationale

Interplay Between Cardiovascular Disease and Cancer

Cardiovascular disease and cancer are the leading causes of death in the U.S. (3,4). Although new antineoplastics and treatment regimens have greatly increased survival for many types of cancer, there is now a special population of patients affected by both cancer and cardiovascular disease, which gave rise to the specialized field of cardio-oncology (49,52,59).

The interplay between cardiovascular diseases and cancer is complex and under active investigation. In the aging population, existing CVDs can complicate cancer treatment regimens, and the very same antineoplastic regimens can also adversely affect the development of CVDs. On the other hand, childhood and adolescent cancer treatment can lead to early-onset of CVDs and heart failure. In general, most systemic antineoplastics and radiation to the chest pose some form of risk for the development of cardiovascular problems (52,61).

The current clinical treatment and management for chemotherapy-induced cardiomyopathy and heart failure are essentially the same as with all other kinds of

heart failure; these treatments focus on reducing systemic volume overload via diuretics and ace-converting enzyme inhibitors. Unfortunately, the prognosis for all heart-failure patients remains poor- at about 50% survival in five years (46,47,55,158).

Current Understanding and Proposed Mechanisms

Currently, the prevailing mechanism of cardiomyopathy induction by chemotherapy (especially anthracyclines) is centered on the formation of reactive oxygen species (ROS). Anthracyclines such as Doxorubicin (DOX) are thought to enter the cell's nucleus and intercalate with GC-rich regions of the chromosomes (159). It appears that DOX can modulate ROS formation via the NADPH oxidases and nitric oxide synthases (NOS) in the mitochondria, which ultimately triggers downstream Bcl-2-associated X protein (BAX) activation and apoptosis (159). Furthermore, the intercalation by DOX into the DNA results in the formation of DOX-DNA-Topoisomerase complexes that ultimately lead to DNA-repair response and subsequent cell death (159,160).

In small animal models such as rats, hypertension has been shown to sensitize the animals to the cardiotoxic effects of chemotherapy (159). This suggests that underlying systemic processes may accelerate the cardiotoxic effects of chemotherapy, rather than localized direct damage alone. However, as with most small animal models, these observations are hard to generalize to the human population as rodents have a vastly different cardiovascular structure, and many physiological and metabolic differences (161–163). Additionally, the administration of DOX in mice (usually just one intraperitoneal injection) does not reflect the complex regimen that human cancer patients typically receive (159).

A step removed from animal models is using human induced-pluripotent stem cells (iPSCs). While several studies have used iPSC to test for acute cellular toxicity, these models are limited in that they are short-lived and do not fully capture the dynamics and stochasticity of the human heart (164). Nonetheless, patient-derived iPSCs seem to recapitulate susceptibility to cardiomyopathy phenotypes of the patients they were derived from, likely because these patient-derived iPSCs inherit the same susceptibility genetic backgrounds (164,165). Because of the ability to recapitulate the genetic susceptibility of individuals, iPSCs could be good models for individualized predictive medicine (164,165).

Justification for This Work

Recently, single-cell sequencing technologies, e.g. scRNA-seq and scATAC-seq, have emerged as the methods of choice for elucidating the cellular changes in various pathophysiologic contexts, including non-chemotherapy-related dilated cardiomyopathies (e.g., idiopathic and genetic mutation-related), pressure-induced cardiomyopathy and hypertrophic cardiomyopathy (68,69,71,72). These studies provide insight and opportunities for therapeutic intervention and a better understanding of the cellular heterogeneity in these diseases. Furthermore, these data allow for integrative comparison to show the different cellular mechanisms between various forms of cardiomyopathy. Unfortunately, no single-cell sequencing dataset exists for chemotherapy-induced cardiomyopathies and the cellular pathophysiology of this particular form of cardiomyopathy remains poorly understood. This is likely due to scarcity and difficulty in obtaining high-quality samples from human hearts, compounded by the difficulty of obtaining high-quality sequencing from multinucleated tissue.

Currently, most cardiomyopathies, including chemotherapy-induced cardiomyopathy, are treated symptomatically, such as by reducing volume overload via diuretics, ACE-inhibitors, and beta-blockers, which may improve survival in the clinical settings (59,166). However, there is very little knowledge informing the exact molecular differences between chemotherapy-induced cardiomyopathy and other forms of cardiomyopathy, and how common heart-failure drugs may be affecting the heart in these different settings.

As cancer survivorship increases, understanding the long-term effects from prior chemotherapies becomes more pressing. Furthermore, understanding how DOX-induced cardiomyopathy relates to other forms of DCM by single-cell sequencing may inform treatment decisions and reveal new therapy targets or potential off-target effects. The study conducted in this chapter will not only amend the aforementioned knowledge gap in the field, but will create the foundational reference for additional hypothesis generation and testing.

4.2: Results

Informed Consent and Patient Demographics

All patients were informed regarding tissue harvesting and usage, and IRB protocol has been approved with the IRB #20007. No identifiable information has been used, and each patient's characteristic is found in Tables 4.1 and 4.2.

Sample ID	Status	Tissue Type	Age	Sex	Diagnostic Code
CH1 (UVA109)	Chemo	Left Ventricle	53	Male	I50.84
CH3 (UVA174)	Chemo	Left Ventricle	55	Female	I42.0

CH4 (UVA120)	Chemo	Left Ventricle	59	Female	I42.8
CH5 (UVA198)	Chemo	Left Ventricle	59	Female	I50.2
N1 (UVA017)	Normal	Left Ventricle	55	Female	NA
N3 (UVA014)	Normal	Left Ventricle	49	Male	NA

Table 4.1: Patient metadata for multiomic sequencing.

Sample ID	Status	Tissue Type	Age	Sex	Diagnostic Code
CH1-V (UVA109)	Chemo	Left Ventricle	53	Male	I50.9
CH4-V (UVA174)	Chemo	Left Ventricle	55	Female	I42.0
N3-V (UVA014)	Normal	Left Ventricle	49	Male	NA
NID1-V (UVA008)	Non-ischemic DCM	Left Ventricle	61	Male	I42.0

Table 4.2: Patient metadata for spatial RNA sequencing. Note that CH1-V and N3-V were the same samples as CH1 and N3, respectively, that were also used in multiomic sequencing.

Optimization of Nuclei Isolation Protocol

One of the central technical barriers in single-cell and single-nucleus studies is isolating cells and nuclei that are intact and free of debris, while protecting RNA and DNA from endonucleases (167). While many previous studies have published proposed protocols, the most effective method is study-specific, depending on tissue type, storage medium, and goal of study (17,69,167–169).

For previously frozen tissues and tissues for which it is difficult to get intact whole-cells such as brain and muscle, single-nucleus RNA-seq is the preferred method (170). We began testing for nuclei quality on pilot normal ventricular tissues using our previously published method designed for snATAC-seq (22). The major steps of the protocol are as follows: 1) freeze dry the tissue in liquid nitrogen and break apart with a mortar and pestle, 2) transfer tissue into a Dounce homogenizer and homogenize, and 3) debris clean up with Opti-Prep. When we tested this protocol on the healthy LV samples that were previously frozen and kept in -80C, we observed $2x10^6$ nuclei/mL, but ample debris remained even after repeated filtering with 70μm cell strainer, and high viability was observed (16.5%, FIGURE 4.1). For single-nuclei isolation, high “viability” suggests that either there is ample auto-fluorescent debris or that many cells are intact and the nuclei are not released. Typically, we aim for 5% or lower, which means 95% are clean, isolated nuclei.

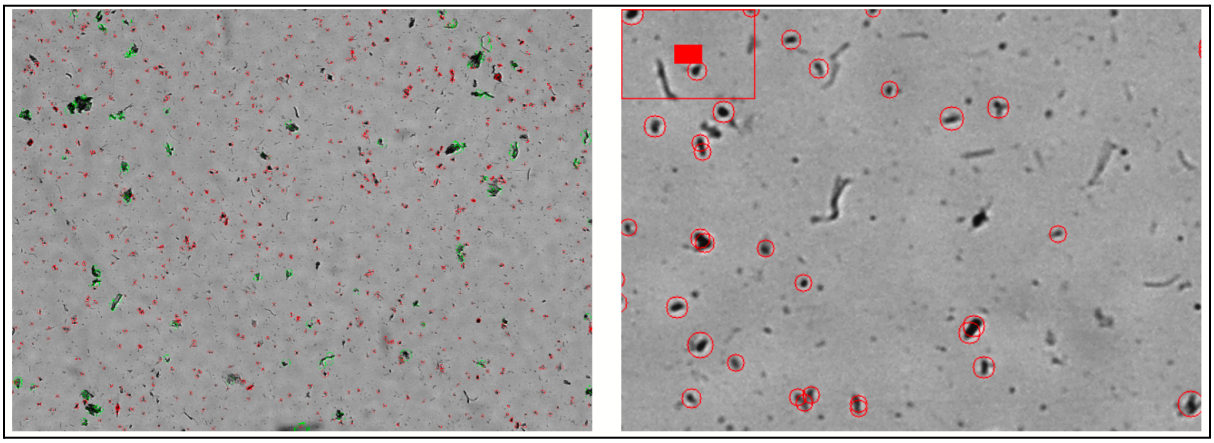


Figure 4.1: Luna images at 1X (left) and 10X (right) of LV nuclei isolated using protocol previously reported in Turner et al., 2022. Nuclei isolates were stained with Trypan Blue and imaged using the Luna counter via the Bright Field Count function.

We sought to reduce the amount of debris using another recently published protocol with some in-house modification to optimize for multiomic analysis (171,172). We embedded the same healthy pilot tissues into OCT (Optimum Cutting Temperature) media, and sliced the tissue at 20μm, 50μm, and 100μm thicknesses for a total of 600μm of tissue at each thickness. This step replaces the liquid nitrogen-freezing and mortar and pestle step to give a more gentle tissue-dissociation step. Then, the tissue slices were dissolved, filtered with a 70μm cell strainer, and their nuclei released using the 10X Multiome Nuclei Isolation Kit (see Section 4.3 Materials and Methods for detailed steps and product numbers).

With this in-house protocol, we were able to isolate over 160,000 nuclei with a 6% viability by slicing the tissue at 50μm for the pilot healthy tissue (Figure 4.2). Microscopic examination of the nuclei revealed significantly cleaner isolate with very few observable debris (Supplemental Figure 4.1). We proceed with this exact protocol for the samples assigned for actual sequencing.

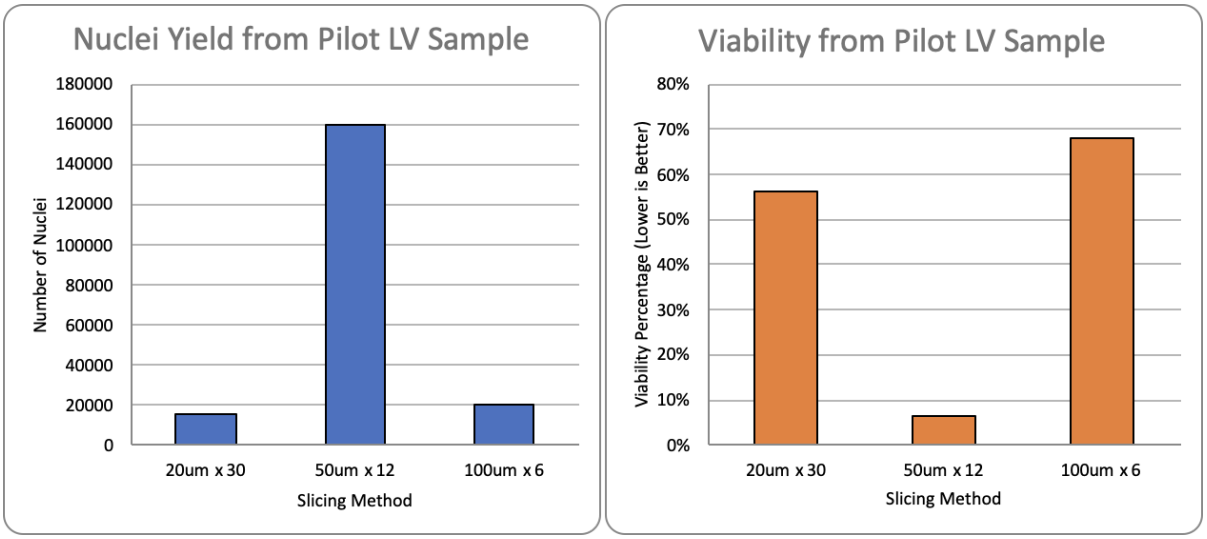


Figure 4.2: Nuclei yield (left) and viability (right) for pilot nuclei isolation protocol.

Sequencing Quality Control for Multiomic Analysis

The sequencing was handled in two batches to minimize handling time of each sample. Additionally, samples CH1 and CH3 both had additional adjacent tissues that were sequenced to supplement the original tissue due to small tissue sizes, and are labeled as CH1_B and CH3_B, respectively (Table 4.3).

Sample	Thickness	Number of Slices	Concentration (nuc/uL)	Nuclei Viability	Batch
N1	50μm	12	1930	20%	First
N3	50μm	12	3070	19%	First
CH1	50μm	20	1300	40%	First
CH3	50μm	20	360	5%	First
CH1_B	50μm	12	2960	0.7%	Second
CH3_B	50μm	12	4560	0.2%	Second
CH4	50μm	12	9000	0.1%	Second

CH5	50μm	12	8250	0.1%	Second

Table 4.3: number of frozen slices and thickness used in multiomic analysis. We typically aim for <5% viability, which means 95% of the nuclei are free from the cell. Nuclei are resuspended for an approximate final volume of 50uL.

The nuclei concentration and viability suggest that both original chemotherapy-related samples, CH1 and CH3, were heavily degraded. Macroscopically, these two original samples were very small (~1mm in diameter each, data not shown). Therefore, in the second batch, we added adjacent LV tissues from these two patients (CH1_B and CH3_B, respectively) to supplement the cell count. The tissues from the second batch appeared more intact upon visual inspection, and the overall viability was lower than 1%, suggesting that most nuclei were free from the cell with limited debris (Supplemental Figure 4.2).

Differential Gene Expression and Clustering Reveals Highly Distinct Cell Type Signatures

Nuclei isolated from the LV tissue were processed using the 10X Genomic Multiome library kit and sequenced using NextSeq 2000 (see Materials and Method), and raw BCL files were aligned using Cellranger and Cellranger-atac. After basic filtering for low-quality nuclei (removing high mitochondrial read percentage and cells with low read mapping), 16,422 high-quality nuclei remained–12,612 cells from the chemotherapy group and 3,810 from the control group (Figure 4.3, Left). The distribution and range of contribution from each donor varied (Figure 4.3, Right). However, all downstream statistical methods are non-parametric and do not assume even sample distribution.

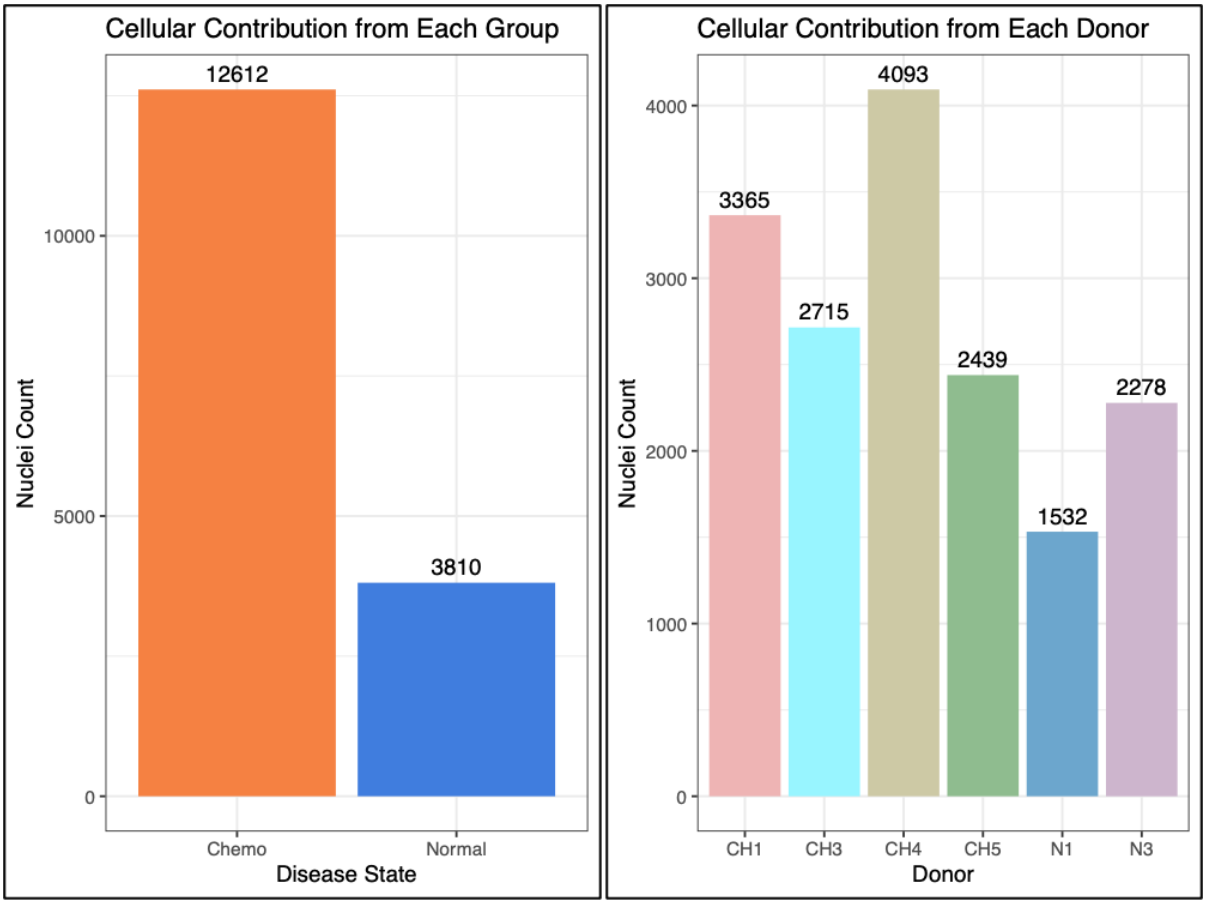


Figure 4.3. Left: cellular composition of the final scRNA-seq dataset by disease state, and by Right: donor. “CH” denotes chemotherapy donor and “N” denotes normal donor.

Once scaled, aligned, and mapped, we integrated the two batches of data using Seurat’s integration function to remove batch effects (75). We then used Seurat’s label transfer function to provide preliminary calls for cell-type labels using the scRNA-seq data from Chaffin, et al., as the reference (68). We further manually refined and confirmed each cluster’s identity by manually examining the top markers (Supplemental Figure 4.3). In the final high-level (labeled as LVL1) annotation, we observed the presence of all expected cell types, including

endothelial cells, macrophages, cardiomyocytes, pericytes, and fibroblasts (Figure 4.4).

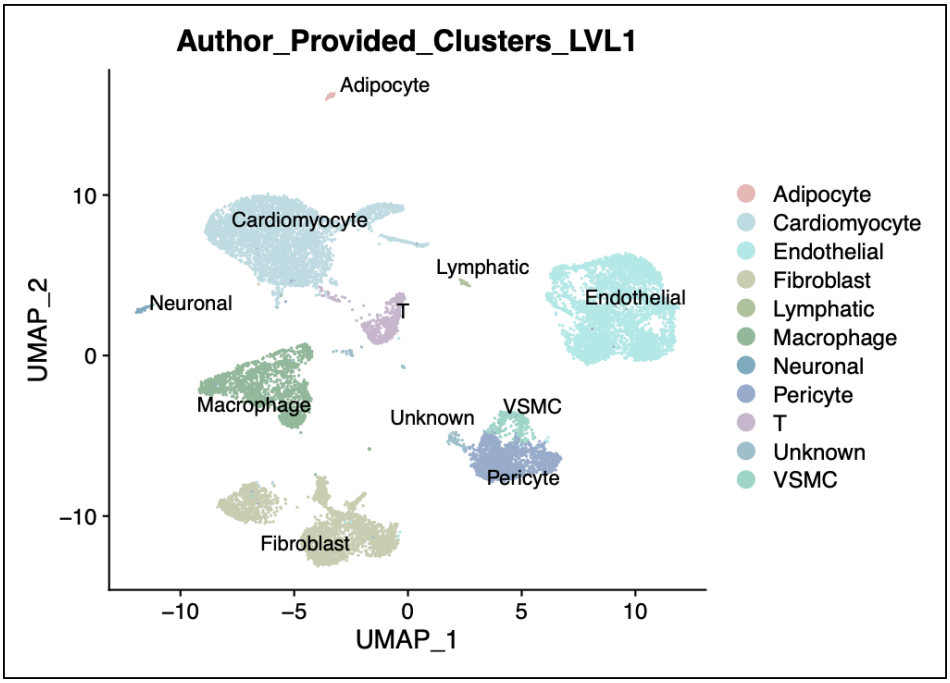


Figure 4.4: Initial clustering and high-level (level one) annotation showed all major cell types present.

To examine for batch effects, we plotted the cells by batch and did not observe the addition of new major clusters or features (Figure 4.5). However, the increase in cell number for each cluster significantly increases the statistical power for downstream analysis. Furthermore, plotting by donor showed no significant skewing for each cell type (Supplemental Figure 4.4)

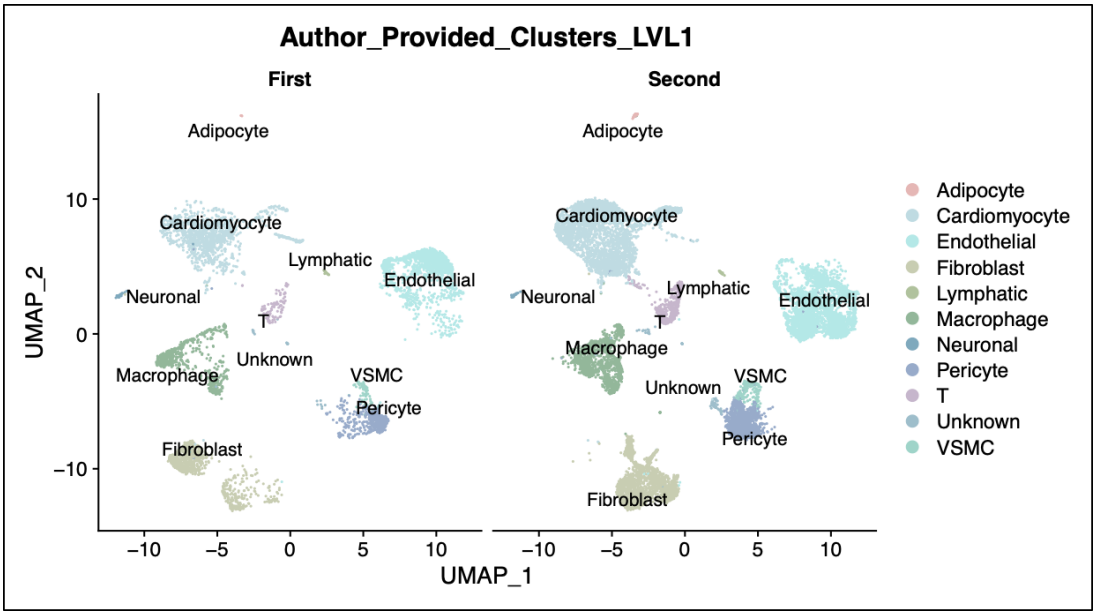


Figure 4.5: UMAP plot separated by batch showed no new cell types or clusters. However, the second batch added a significant number of cells and will increase downstream analysis power.

Fibroblasts Show Distinct Grouping and Activation in Chemotherapy Patients

While most cell types showed expected harmonization and close clustering, fibroblasts showed two distinct subclusters that are closely related to chemotherapy status despite the previous integration step (Figure 4.6). While canonical fibroblast markers such as vimentin (*VIM*) and fibronectin (*FN1*) are expressed in both clusters, we found that chemotherapy-fibroblasts are associated with increased expression of activation markers (173), such as Tenascin-C (*TNC*), Transgelin (*TAGLN*), and Periostin (*POSTN*, Figure 4.6). Interestingly, in another study, the overexpression of activation markers was found in fibroblast derived from post-myocardial infarction (173). Furthermore, fibroblast activation was also observed in idiopathic DCM, supporting previous studies implicating fibroblasts as critical to the heart-failure response (71).

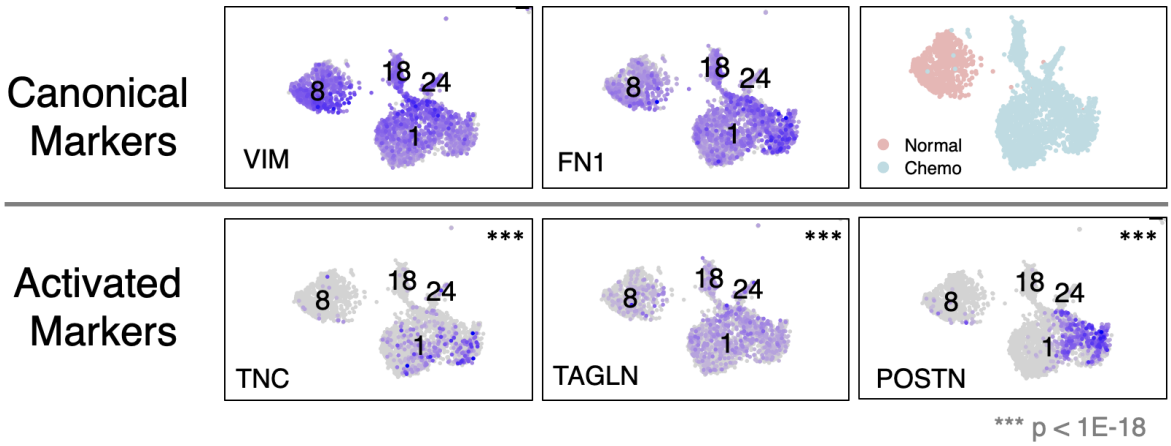


Figure 4.6: Fibroblasts from chemotherapy donors showed increased activation markers, such as tenascin-C (*TNC*, p-adj.=2.45E-18), transgelin (*TAGLN*, p-adj.=1.10E-25), and periostin (POSTN, p-adj.=1.50E-25).

Cellular transition states were previously observed in atherosclerosis, where smooth muscle cells take on a fibroblastic phenotype referred to as “fibromyocytes” (17,27). To look for transition states that may exist between normal and chemotherapy fibroblasts, we performed pseudotime analysis of the subsetted clusters using Monocle3 (Figure 4.7). Interestingly, Monocle3 could not find logical transitions between the two fibroblast clusters, suggesting that they are transcriptionally distinct. This is likely due to the fact that the chemotherapy patients from this study were already in end-stage heart failure requiring a transplant, so we did not capture the acute response and transition of inactivated fibroblasts. Although they both express canonical fibroblast markers like vimentin and fibronectin (Figure 4.6), the unique activation profiles of chemotherapy fibroblasts present possible avenues of therapy-testing.

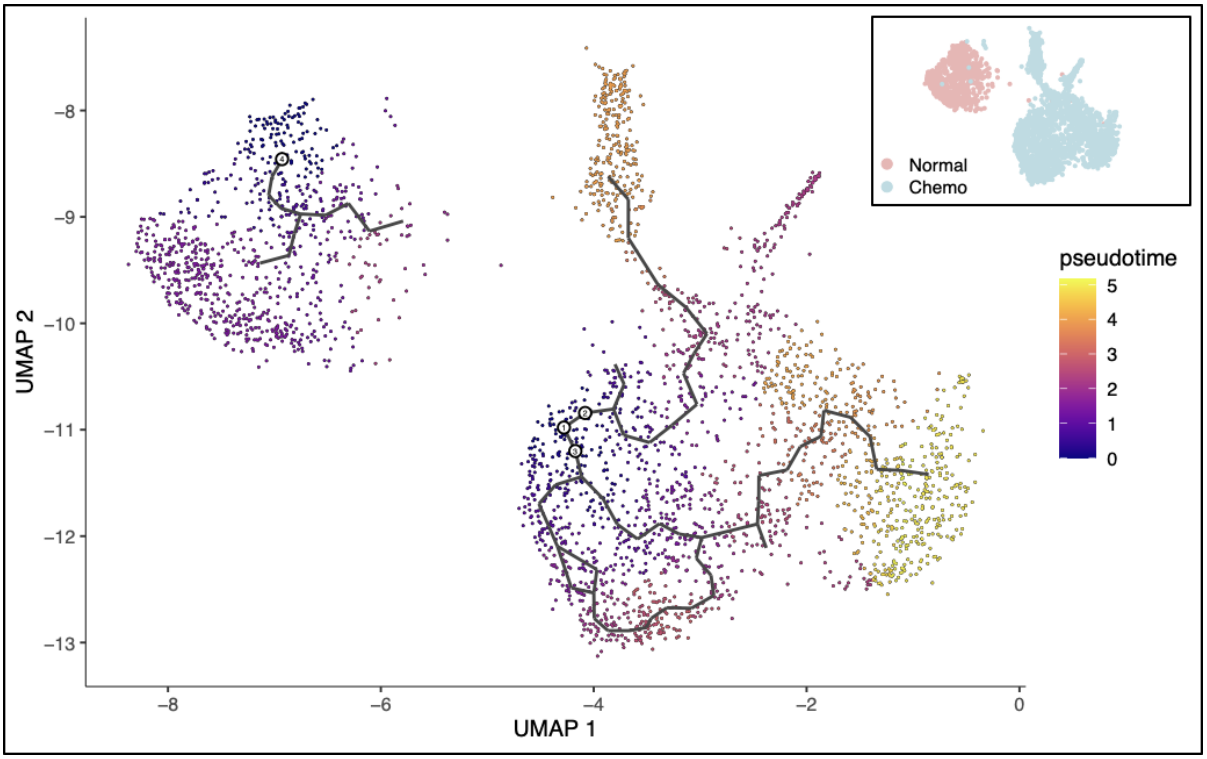


Figure 4.7: Pseudotime analysis with Monocle3 on fibroblast clusters could not identify a logical path connecting the two fibroblast clusters.

Fibroblasts from Chemotherapy Group Showed Transcriptional Signatures that Mimic Idiopathic Cardiomyopathy

To define chemotherapy-specific changes in fibroblasts, we integrated our fibroblast subset with fibroblast data from the non-chemotherapy DCM dataset by Koenig, et al. *(71).* Overall, the transcriptional profiles of fibroblasts from the chemotherapy group mimic those from idiopathic DCM and overlap the same UMAP (uniform manifold approximation projection) dimensions, while control (labeled as “Donor” in the Koenig, et al., data) fibroblasts from DCM closely overlap with our normal fibroblasts (Figure 4.8).

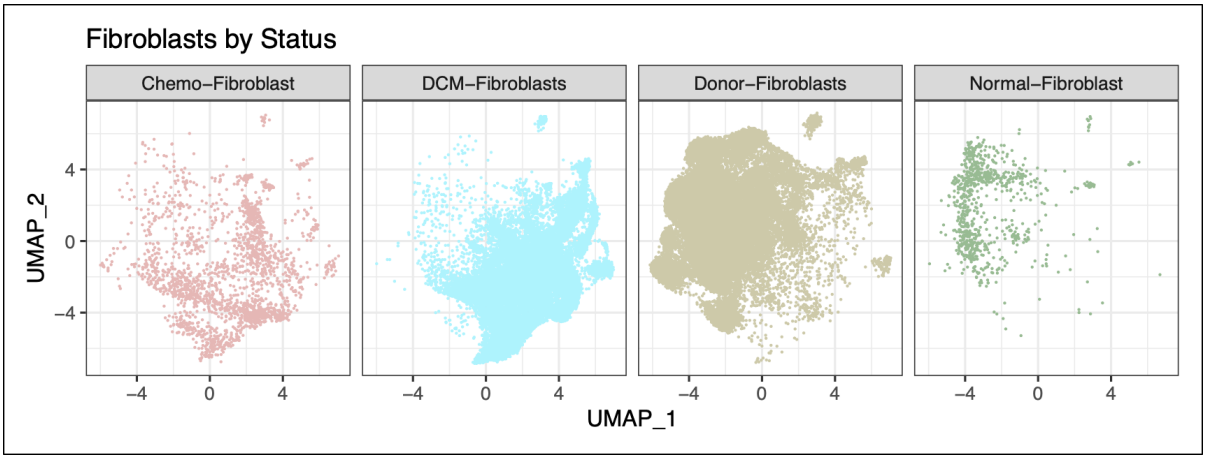


Figure 4.8. Fibroblasts from the chemotherapy group and idiopathic DCM data (Koenig, et al.) clustered in overlapping UMAP coordinates, suggesting overall similar transcriptomes. Note: Chemo and Normal-Fibroblasts are from this study, and DCM and Donor-Fibroblasts are from Koenig, et al. Donor is the non-diseased reference in Koenig, et al.

Despite overall transcriptional similarities, global UMAPs cannot distinguish subtle changes and differences in gene expression or signaling pathways (67,174). Therefore, we performed formal pathway enrichment analysis using ClusterProfilers and bespoke codes (see Material and Methods). We found expected differences between chemotherapy and control fibroblasts, such as altered extracellular matrix and binding proteins (Figure 4.9, Left). Surprisingly, GTP regulation and G-protein coupled receptor (GPCR) signaling were among the major differences between chemotherapy and idiopathic DCM fibroblasts (Figure 4.9, Right). Nucleotide-triphosphate exchange factors, GTPases, and nucleoside triphosphatase regulators are highly involved in regulating the GPCR signaling cascade that modulates extracellular matrix and cytoskeletal dynamics via proteins like Rho and Rac (175). Furthermore, current heart-failure (HF) medications, such as βARs agonists (β-adrenergic receptors, e.g., propranolol) and ACE-inhibitors (e.g., lisinopril) act on GPCR signaling (175,176).

While GPCR signaling is complex and is involved in many pathways, our data suggests that fibroblasts from end-stage HF have distinct signaling profiles when compared to idiopathic DCM and should be investigated further.

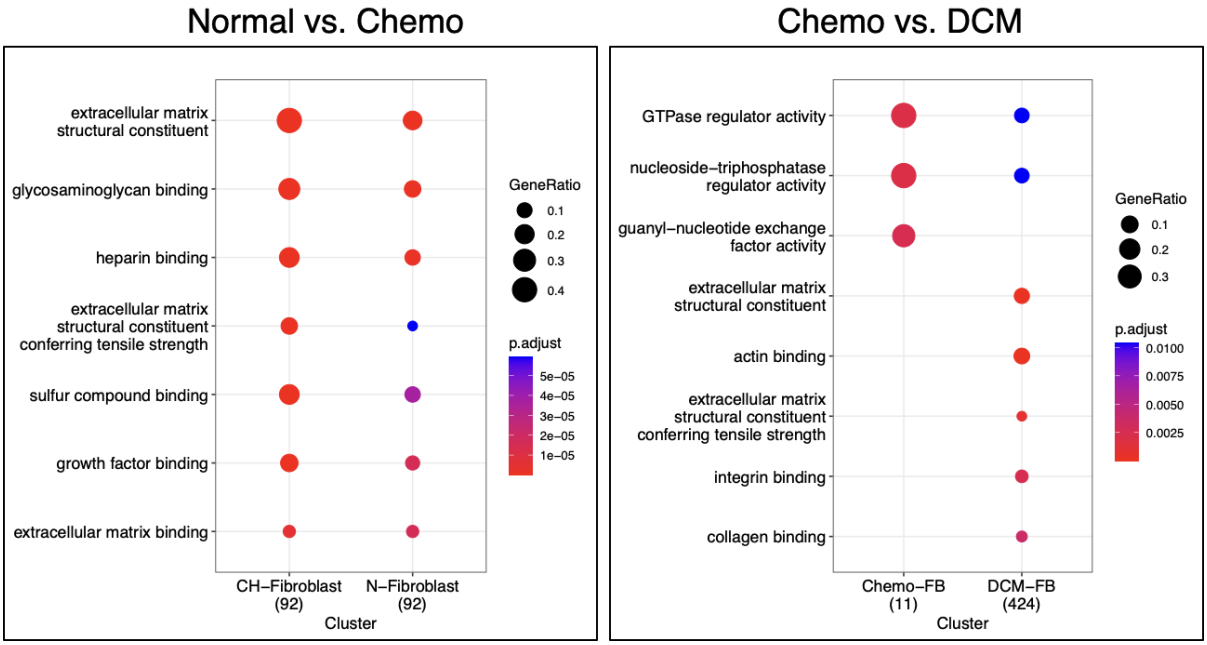


Figure 4.9: Formal pathway enrichment using the GO (Gene Ontology) Resources Database and ClusterProfilers on fibroblasts. Top differentially expressed genes were calculated and analyzed using ClusterProfiler via custom codes.

NADPH Oxidase 4 is Mostly Expressed in Fibroblasts, and Not Cardiomyocytes

Contrary to existing literature that suggests a central role of reactive oxygen species (ROS) in cardiomyocytes in the pathogenesis of chemotherapy-induced DCM, we observed very low oxidative gene expression, such as NADPH oxidases and nitric oxide synthases, in cardiomyocytes. Instead, we found that fibroblasts expressed a relatively high level of NADPH Oxidase 4 (*NOX4*, Figure 4.10, left), which has been shown to generate ROS in the settings of cancer and fibrosis (29). Interestingly, there is no significant difference in *NOX4* expression level when we

compare by chemotherapy status (Figure 4.10, Right). Furthermore, major redox signaling players, such as *NOX1*, *NOX2* (same as *CYBB*), nitric oxide synthases 1, 2 and 3 (*NOS1-3*), and myeloperoxidase (*MPO*) were not significantly expressed in any cells types (Supplemental Figure 4.5). Our data suggests that redox signaling may not be a significantly altered pathway in most cells as a response to chemotherapy, at least in the end-stage of chemotherapy-induced heart failure.

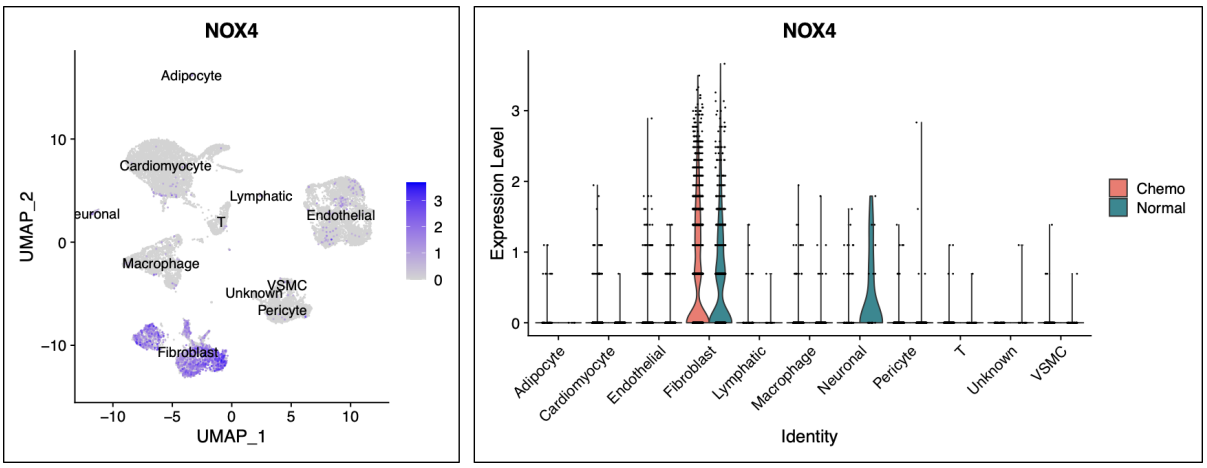


Figure 4.10: Left: expression of *NOX4*, a major constituent of the redox signaling pathway and is involved in many pathogenic processes. Right: *NOX4* expression when grouped by chemotherapy status.

Cardiomyocytes Showed Significant Expression Changes in Response to Chemotherapy

Cardiomyocytes have been the main focus on chemotherapy-induced cardiotoxicity, with many iPSC cardiomyocyte models currently employed in *in vitro* studies (164,165). However, single-cell mapping of the chronic response in cardiomyocytes in human patients has not been performed. We compared the top differentially expressed genes in cardiomyocytes from healthy and chemotherapy groups. Of note, housekeeping genes such as *ACTA1* (alpha-actin 1), *TNNC1*

(troponin C1), and *MYL2* (myosin light-chain) were relatively downregulated in chemotherapy-cardiomyocytes. Mutation in *MYL2* has been linked with cardiac hypertrophy, and all three of the aforementioned genes are important components in cardiac contractility (177). The reductions in MYL2 and ACTA1 could affect the efficiency of crossbridge binding between actin and myosin heads during the production of isometric contraction (178,179).

More importantly, *XIST* (X-inactive specific transcription), a long non-coding RNA, was highly expressed (Figure 4.11). Although *XIST* has been studied in mouse models and has shown to play a role in cardiac hypertrophy, the exact effect of *XIST* overexpression in the context of chemotoxicity remains to be confirmed.

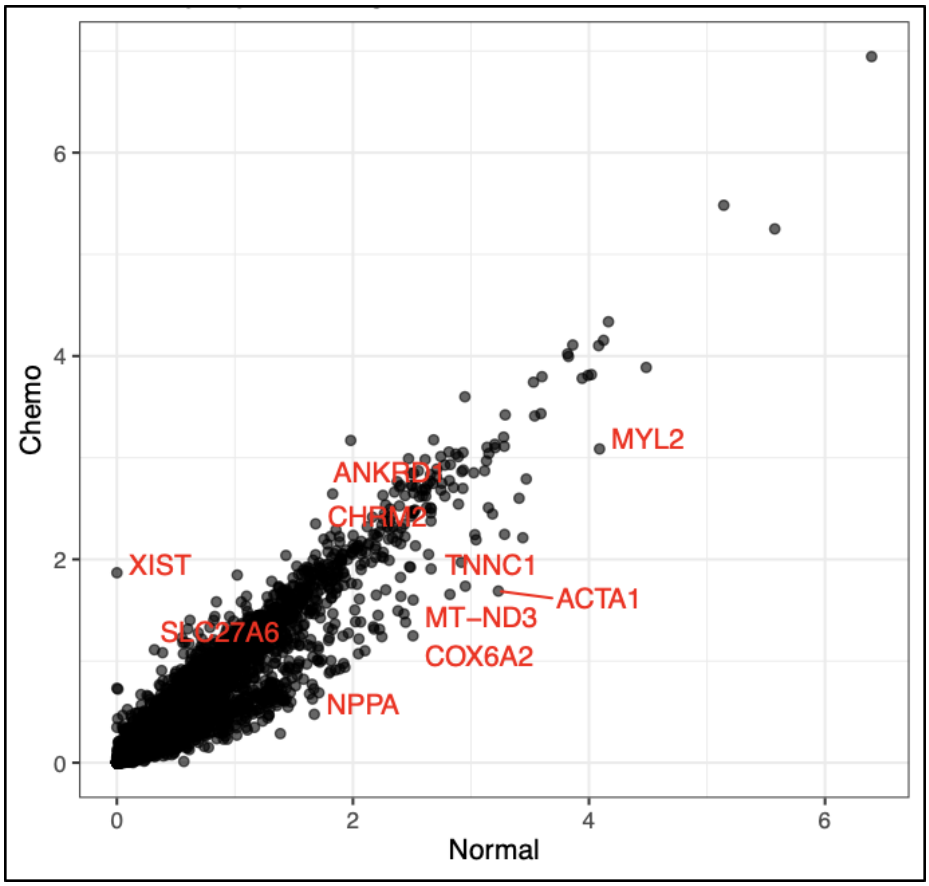


Figure 4.11: Cardiomyocyte relative expression differences between chemotherapy and normal group, with variable genes highlighted in red.

XIST is Highly and Specifically Expressed in Chemotherapy LV

X-inactive specific transcript (*XIST*) was highlighted as one of the genes predominantly expressed by chemotherapy-cardiomyocytes. Surprisingly, we found that *XIST* is specifically expressed by most cell types in the chemotherapy group but not in normal control LV tissue, regardless of cell types (Figure 4.12). As a long non-coding RNA, *XIST* transcripts coat the X-chromosome for inactivation (180,181). Recently, the roles of *XIST* in other disease processes have emerged. Most notably, *XIST* has been shown to regulate hypertrophy of the mouse heart (181). Although *XIST* has been shown to be mainly expressed in females, detailed analysis showed that *XIST* was only highly expressed across multiple cell types in the chemotherapy group even when factoring in sex category (Supplemental Figure 4.6). The differential expression of *XIST* could in part explain the sex differences observed in chemotherapy-induced cardiomyopathy, where female are generally more at-risk for severe HF following chemotherapies (182). Further integration of the RNA-expression data with coupled snATAC-seq data revealed no potential co-accessibility regions (Supplemental Figure 4.6). At this time, the exact effect of overexpression of *XIST* in chemotherapy-induced DCM– whether it is a response or a risk-factor– remains unclear, and more patient sequencing data is required to generalize this finding.

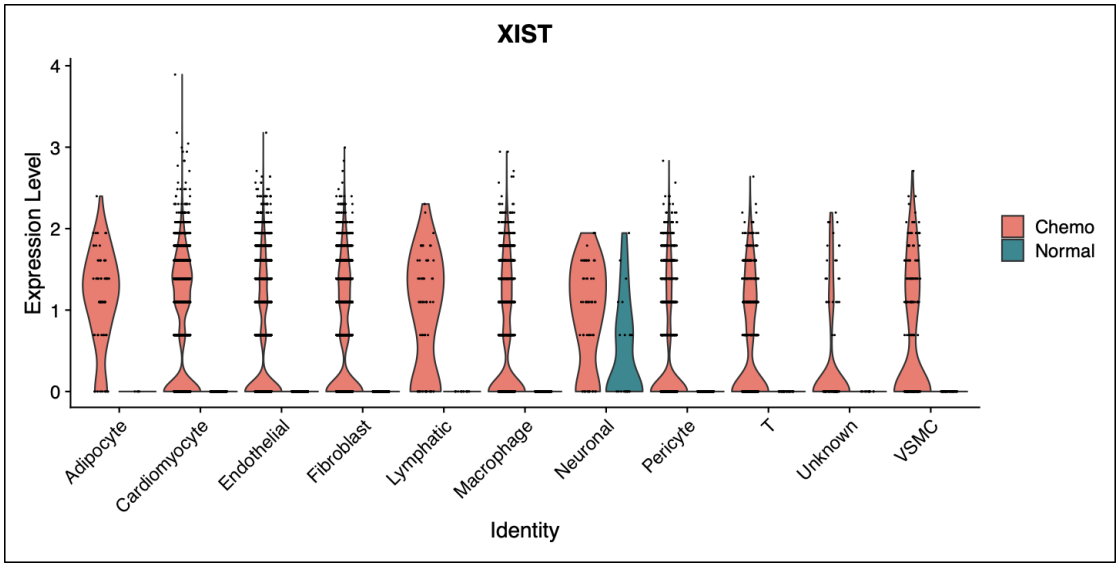


Figure 4.12: *XIST* is overexpressed in almost all cell types in the chemotherapy-induced DCM group.

Fatty Acid Oxidation Presents a Potential Avenue of Therapy for Chemotherapy-DCM

Fatty acid metabolic pathways have been targeted as a potential mechanism to treat angina (183). It is thought that inhibiting cardiac utilization of FA decreases lactic acid production and improves contractile functions during cardiac ischemia (183). Interestingly, although we found several genes related to FA metabolism to be upregulated in cardiomyocytes, such as *SLC27A6*, a fatty acid transporter, we found substantial increase in FA metabolism pathways in macrophages from chemotherapy-DCM compared to macrophages from idiopathic DCM samples (Figure 4.13). Although macrophage uptake of fatty acid has been studied extensively in the context of atherosclerosis (14,17,24,139,184–186), and cardiac injuries such as myocardial infarctions (187), how and why FA oxidation in macrophage occurs in the context of HF is not well understood. Nonetheless,

increase in FA oxidation could be a response to the increased GPCR signaling from fibroblasts (187,188).

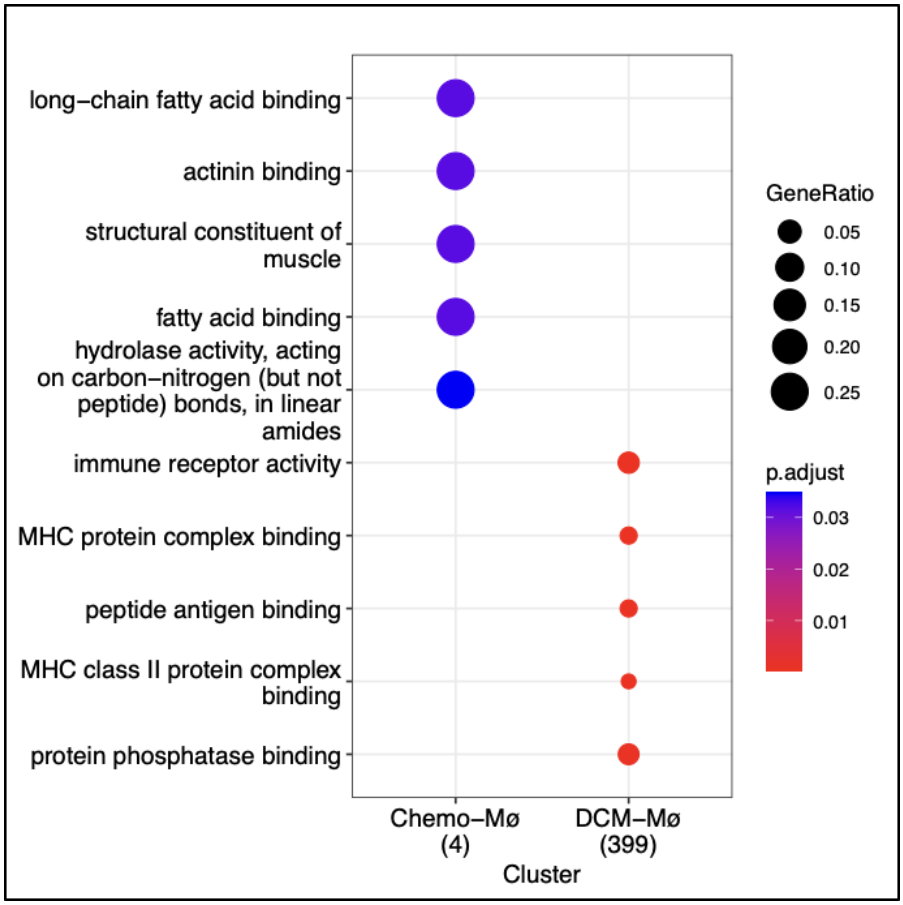


Figure 4.13. Macrophages demonstrate increased expression in genes related to fatty acid binding and hydrolase activities when compared to macrophages from idiopathic DCM.

GPCR Signaling May Be Involved in Endothelial Dysfunction

While endothelial dysfunction has been documented in many cardiovascular diseases, such as atherosclerosis, and many forms of cardiomyopathy, the endothelium may be particularly susceptible to chemotherapy-induced dysfunction

(189,190). In mouse and human ex vivo experiments, the chemotherapy agent docetaxel appears to cause endothelial dysfunction through *NOX4* (190). While we did not see high expression of *NOX4* except in fibroblasts in our end-stage heart failure tissues (Figure 4.10), we noted that chemotherapy-endothelial cells exhibited increased *RHOA* GTPase and *CDC42* GTPase signaling (Supplemental Figure 4.7). GTPases have been implicated in many cardiovascular diseases, including hypertension and diabetes, and understanding the role of GTPase affect cell contractility and eccentric hypertrophy in the chemotherapy response could provide future avenues to prevent cardiac toxicity (191,192).

Current Treatments for Heart Failure May Have Unwanted Effects on the Chemotherapy-Induced Cardiomyopathic Heart

To investigate how current heart failure medications may be affecting the heart in chemotherapy and non-chemotherapy-induced heart failures, we calculated overexpressed genes (log2-fold change >0.5) for each diseased cell type in our dataset and in Koenig, et al. We then cross-referenced overexpressed genes with known interaction targets of common heart-failure treatments, including diuretics, beta-blockers, and dexrazoxane, a drug used to prevent heart-failure in the setting of chemotherapy (full list of drugs in Material and Methods section). The results were surprising: we found that there were more potential interactions with most cell types in the chemotherapy-DCM group relative to the idiopathic-DCM group (Figure 4.14). This result is concerning because most current heart failure treatments act on the periphery to alleviate the symptoms, i.e., removing excess volume by means of diuresis and reducing systemic blood pressure by acting on the RAAS (166). For example, furosemide, a widely-prescribed loop diuretic, targets APOA1 (155) in chemotherapy group cardiomyocytes, yet no target was found in the idiopathic group cardiomyocytes (Figure 4.14). Nonetheless, how these differential interactions

affect the heart as well as patient outcomes require additional studies. Full list of drug tested and drug-gene interactions by cell type are available in our GitHub as supplemental files.

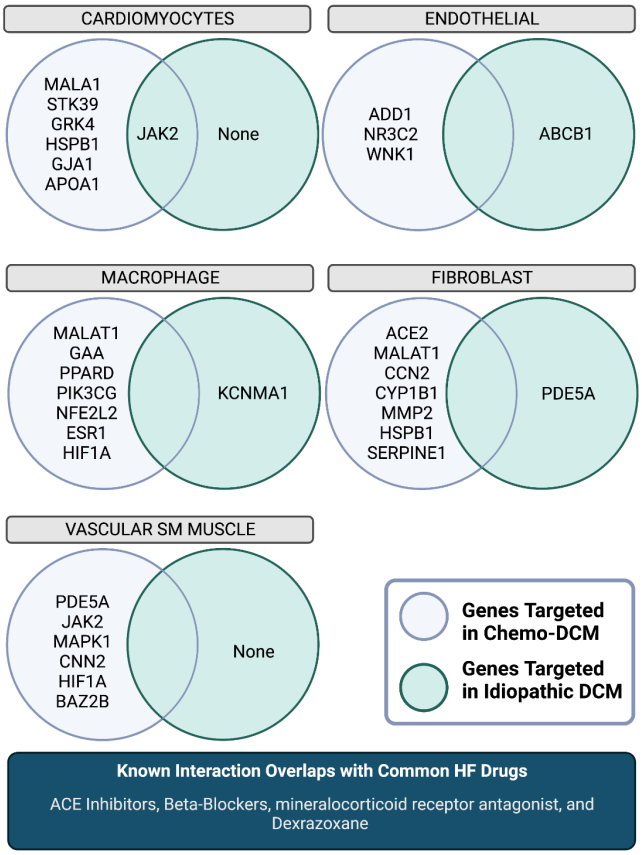


Figure 4.14: Differential targeting analysis in different cell types by current heart-failure drugs reveals many potential targets expressed by cells in chemotherapy-DCM cells and not in idiopathic DCM cells. Full list of drug-gene interaction is available in our GitHub as supplemental files.

snATAC-seq Provides Opportunities for Future Putative Driver Interrogation

In conjunction with the snRNA-seq, snATAC-seq were performed in the same nuclei. Although fewer nuclei passed quality control criteria, we were still able to

visualize the snATAC-seq data overlaid with the snRNA-seq (Figure 4.15). Although we observed strong batch-effects, we were able to correct them with only two iterations via Harmony (Supplemental Figure 4.8). In all, we were able to recapitulate the same major clusters and were able to map the accessibility region of *XIST* in all cell types. Interestingly, although *XIST* is highly observed in most cells in the RNA level, many cell types, including fibroblasts, showed no open chromatin in the corresponding gene region (Supplemental Figure 4.9). This suggests to us that the data we have may be too sparse to conduct formal statistical analysis on its own and must be used in conjunction with the RNA data. Nonetheless, we will make this joint dataset as well as individual snRNA and snATAC-seq data available through our single-cell portal, PlaqView, at www.plaqview.com.

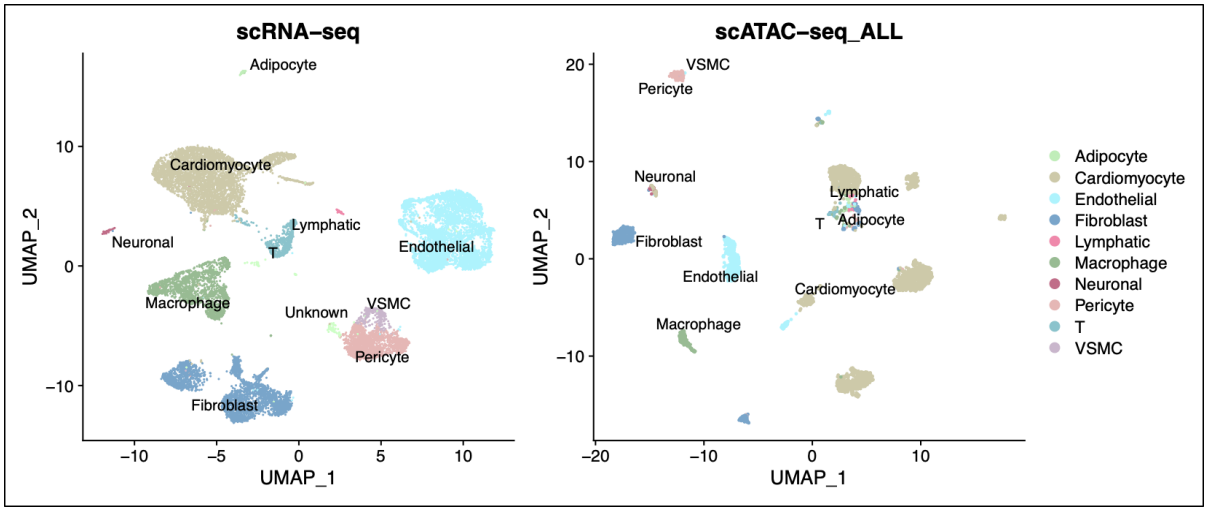


Figure 4.15: Sparse snATAC-seq data still recapitulated major cell types present in snRNA-seq.

Spatial Transcriptomics Provide a Starting Point for Future Confirmatory Studies

In addition to snATAC-seq and snRNA-seq, we performed spatial RNA-seq on two chemotherapy, one normal, and one idiopathic DCM LV samples as a proof of concept and to provide a starting point for additional confirmatory tests. One section was sequenced from each sample and spatial distribution of estimated cell types were calculated using Seurat's label transfer protocol. We found no abrupt or focal changes in cardiomyocyte (Figure 4.16), fibroblast, endothelial or macrophage distributions (Supplemental Figure 4.10). Unfortunately, we found it difficult to discern specific clusters of cells or spatially variable genes, even after attempting resolution enhancement for sample CH1 by Baysian statistics models with BayesSpace (Supplemental Figure 4.11).

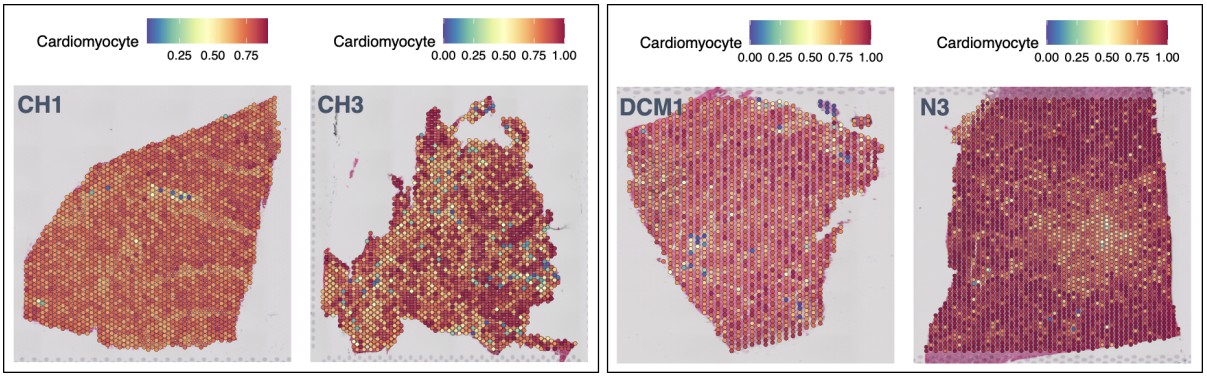


Figure 4.16: spatial RNA-seq was performed on four samples, CH1 and CH3 were from the chemotherapy group, DCM1 is from a donor with idiopathic dilated cardiomyopathy, and N3 is a normal control.

4.3: Discussions

The present study is the first to interrogate the cellular landscape of chemotherapy-induced dilated cardiomyopathy (chemotherapy-DCM). We

performed multimodal sequencing on left ventricular (LV) tissues from chemotherapy-induced cardiomyopathy and normal patients. Furthermore, we integrated our data with existing single-cell idiopathic DCM data to examine the specific long-term effects and cellular response unique to chemotherapy.

Our innovative procedure allowed for isolation of high-quality nuclei from aged, flash-frozen LV tissue, a challenging tissue type due to the multinucleated nature and the abundance of myofibrils. We also performed comprehensive analysis of the cellular heterogeneity between normal, chemotherapy, and idiopathic cardiomyopathy groups.

In our cell-specific analysis of fibroblasts, we showed that chemo-fibroblasts are transcriptionally distinct from normal-fibroblasts and showed markers of activation, even at the end-stage HF. Furthermore, these chemo-fibroblasts, although transcriptionally similar to idiopathic DCM fibroblasts, have different G-protein coupled receptor (GPCR) signaling activity. Comparatively, this suggests that chemo-fibroblasts may have different cytoskeletal dynamics and attachment to the extracellular matrix, and the link between GPCR signaling in fibroblasts and fatty acid oxidation in macrophage should be further explored. More importantly, we found that *NOX4*, a key player in redox signaling, was mainly expressed in fibroblasts, and not cardiomyocytes or any other cell types. In fact, cardiomyocytes exhibited very low expression of key reactive oxygen species-generating (ROS) and nitric oxide synthase genes. This finding is contrary to existing knowledge about ROS being the central mechanism involved in the pathophysiology of chemotherapy-DCM. However, since we are capturing these primary tissues at the end-stage of heart failure, it is possible that genes involved in ROS production are relatively down-regulated because the acute effects of chemotherapy have likely passed. Nonetheless, our data suggests that chemo-fibroblasts, although

transcriptionally similar to DCM-fibroblasts from idiopathic origin, have subtle differences that can be the basis of off-target effects.

Our study also showed a potential new player in chemotherapy-DCM: *XIST*. The X-inactive specific transcript (*XIST*) is a long non-coding RNA expressed mainly in females in the inactivation of the second X chromosome (181). *XIST* was specifically and highly expressed in many cell types in chemotherapy-DCM, and presents an interesting avenue of research. It is also important to point out again that our sample size is limited, and we only have one female control sample. At this time, it is unclear from our preliminary data whether *XIST* is a result of chemotherapy or a risk-factor for the increased susceptibility for chemotherapy-induced cardiomyopathy in female patients. *XIST*, along with other sex-specific genes, should be further investigated.

Overall our data and comprehensive analysis present the first cellular insights into chemotherapy-induced cardiomyopathy in humans. We believe our study presents many new avenues of research and will be a valuable resource for the cardiovascular research community.

4.4 Study Limitations and Future Directions

Several limitations exist in our aforementioned study, despite the strong statistical values from our calculations. First, our chemotherapy-induced cardiomyopathy groups were all in the end-stage of heart failure. Future work should focus on including patients from varying stages of disease although heart biopsies are exceedingly rare. Second, our study included only two control patients who self-identified as male and female, which is not representative of our gender spectrum. Further studies with larger cohorts of both male and female, as well as person who identifies as transgender, should be conducted. Lastly, at the time of

writing, we are unable to obtain the exact amount of chemotherapy received for each patient, making stratification of disease stage difficult.

Please see Chapter 5 for future directions.

4.5: Materials and Methods

Tissue preparation for Multiomic Analysis

Roughly 300mg of left ventricular tissues from either explant (chemotherapy group) or rejected donor hearts (normal references) were excised from the whole heart and either flash frozen in liquid nitrogen or embedded in OCT (Optimum Cutting Temperature media, Tissue-Tek Sakura, Cat #4583) as blocks, and stored at -80°C. The day before nuclei isolation, twelve tissue sections were cut in a cryostat at -18°C using Epredia™ MX35 Premier™ Disposable Low-Profile Microtome Blades (ThermoFisher, Cat# 3051835) at 50μm and stored in LoBind 1.5mL tubes (eppendorf, Cat# 022431021). If the tissue was not embedded in OCT at the time of tissue harvest, it was separately embedded on the day of sectioning without thawing. The cut tissue slices were kept overnight at -80°C.

The sliced tissues were then processed using the 10X Genomics Multiome Nuclei Isolation Kit (10X Genomics, Cat # 100494) with slight modifications. First, the sliced tissues were homogenized and lysed using 500uL of Lysis Buffer. The lysate was transferred to a 70μm sterile cell strainer over a 50mL conical tube (pre-prechilled, on ice) and centrifuged at 200g for 30s at 4°C. The filtrates were loaded into the filter columns and washed as described in the manufacturer's protocol. The final nuclei isolates were resuspended in 50uL of Resuspension Buffer and counted with a Luna Cell Counter using Propidium Iodine/Acridine Orange dye (Vita Scientific, Cat # F23001) at 10:1uL, respectively. All instruments and material were kept on ice whenever possible.

Single Nuclei Library Preparation for Multiomic Sequencing

The generation of single nuclei-indexed gene expression (GEX) and ATAC-seq libraries was performed by the UVA School of Medicine Genome Analysis and Technology Core (RRID:SCR_018883) using the 10X Genomics Chromium Controller platform and the Chromium Next GEM Single Cell Multiome Kit (10X Genomics, Cat # 1000285). Briefly, the nuclei isolate concentrations were adjusted with a Nuclei Dilution Buffer containing RNase Inhibitor to aim for approx. 16,000 cells-per-5uL prior to incubating with the Transposase Mix. The transposed nuclei were then loaded on the Chromium Next GEM Chip J, combined with barcoded Gel Beads and partitioned into single-nuclei Gel Bead-in-Emulsion (GEMs). After GEMs incubation, the resulting barcoded DNAs from both the transposed DNA and full-length cDNA are pre-amplified. Indexed ATAC libraries generated via PCR from the pre-amplified barcoded fragments are visualized on the Agilent 4200 TapeStation Instrument, using the Agilent D5000 kit. For gene expression, pre-amplified barcoded DNA are further subjected to cDNA amplification and the dual-indexed gene expression libraries are pooled by mass prior to sequencing. At this stage, a routine quality-control sequencing run was performed on the Illumina MiSeq using the Nano 300 Cycle Kit (1.4 Million reads/run, Illumina, Cat # MS-102-2002), to estimate the number of targeted nuclei-per-sample via CellRanger v3.0.2. The estimated count enables us to re-balance the pooled sample prior to deep sequencing. We proceeded with the final sequencing with the NextSeq 2000, using the P3-100 Cycle Kit (Illumina, Cat # 20040559). After run completion, the binary base call (.bcl) files are generated and stored for downstream analysis.

Multiome Sequence Alignment

Raw .bcl files were demultiplexed and aligned to the human (Hg38) genome using three iterations: CellRanger v7.0.1 (gene expression), CellRanger-atac v.2.0.0 (ATAC-seq), and CellRanger-ARC v.7.0.0 (Multiome) to generate data for scRNA-only, scATAC-only, and combined scRNA/ATAC output files, respectively. Full alignment scripts can be found at the GitHub page for this article: https://github.com/wfma/Ma_2023.

RNA Cell Labeling and Differential Gene Expression Analysis

We imported the Multiome and RNA-only count matrices in Seurat V4 and performed built-in basic filtering. Using the RNA-only matrices, we performed removal of low-quality nuclei (<500 read/cell and >5% mitochondrial RNA), principal component analysis, and dimension reduction using uniformed manifold projection approximation (UMAP). Gene expression differences were calculated using Seurat's FindAllMarkers() function and p-values reported were all adjusted using the default conservative Bonferroni correction method.

RNA Cluster Pathway and Pseudotime Analysis

To compare the large list of differentially expressed genes across clusters and conditions, we composed bespoke code scripts to separate and annotate each cluster and condition. These annotated cells were then analyzed using ClusterProfiler, a technology-agnostic tool for pathway enrichment analysis (193). For pseudotime and trajectory analysis, we first use bespoke codes to transform our data into a cds object, which is then processed using Monocle3. The full code script is available at https://github.com/wfma/Ma_2023.

RNA Drug-Gene Expression Analysis

To explore how common heart-failure drugs can interact with cells in the heart, we first subsetted and defined all positively overexpressed genes by cell type using the FindAllMarkers() function in Seurat. We then extracted the known targets of common heart failure drugs in DGIdb (111,155). These drugs are: captopril, enalapril, fosinopril, lisinopril, perindopril, quinapril, ramipril, trandolapril, candesartan, losartan, valsartan, sacubitril, ivabradine, bisoprolol, metoprolol, carvedilol, spironolactone, eplerenone, hydralazine, furosemide, bumetanide, torsemide, chlorothiazide, amiloride, indapamide, metolazone, triamterene, and dexrazoxane. The full list of drug-gene interaction and references associated with each interaction can be found on our GitHub page. The known interactions were cross-referenced with the positively overexpressed genes by cell type, and the final figure was created in BioRender.

Tissue Preparation for Spatial RNA-seq

Visium Tissue Optimization (TO) slides (10X Genomics, Cat # 1000193) with 10 micron sections were subjected to methanol fixation and H&E staining prior to tissue optimization following the manufacturer's protocol. All imaging (bright field and fluorescence) was completed using an Olympus BX51 microscope with a motorized stage, and 12 minutes was identified as the optimum permeabilization time.

Visium Gene Expression slides (10X Genomics, Cat # 1000187) with 10 micron sections were fixed, H&E stained, and imaged as described for TO slide processing. Tissue was permeabilized ([insert minutes}), released mRNA was

captured on the spatially barcoded spots and subjected to cDNA synthesis per manufacturer's protocol. The generated cDNA was subjected to real-time PCR quantification to determine optimal cDNA amplification. Amplified cDNA was subjected to library preparation following the 10x Genomics Spatial Expression Protocol.

Final library quality was assessed using the LabChip Touch HT (PerkinElmer) and libraries are quantified by Qubit (ThermoFisher). Pooled libraries were subjected to paired-end sequencing according to the manufacturer's protocol (Illumina NovaSeq 6000). Bcl2fastq2 Conversion Software (Illumina) was used to generate demultiplexed Fastq files and the SpaceRanger Pipeline (10X Genomics) was used to detect tissue, align reads, generate feature-spot matrices, perform clustering and gene expression analysis, and place spots in spatial context on the slide image. Visium spatial transcriptomics assays were carried out in the Advanced Genomics Core at the University of Michigan.

Spatial Feature and Resolution Enhancement

Baysian statistics and other modeling methods from image analysis have been used to enhance the resolution of Visium slides (194–196). Here, we employed the package BayseSpace, a Baysian statistics tool that enhances the resolution of spatial transcriptomics and impute features (gene expression). We first read the 10x output into a Seurat object by subject and performed BayseSpace imputation individually. The enhanced objects were then saved and merged into a combined Seurat obj for downstream analysis. The full codes used can be accessed at our GitHub page at github.com/wfma/Ma_2023.

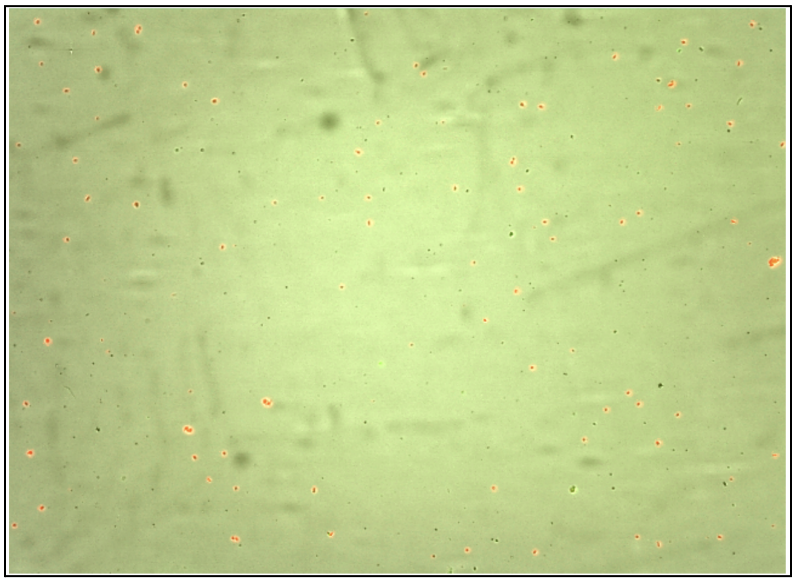
Data and Code Availability

All data and code will be available through our single-cell web portal, PlaqView.com. Code will also be available at github.com/wfma/Ma_2023.

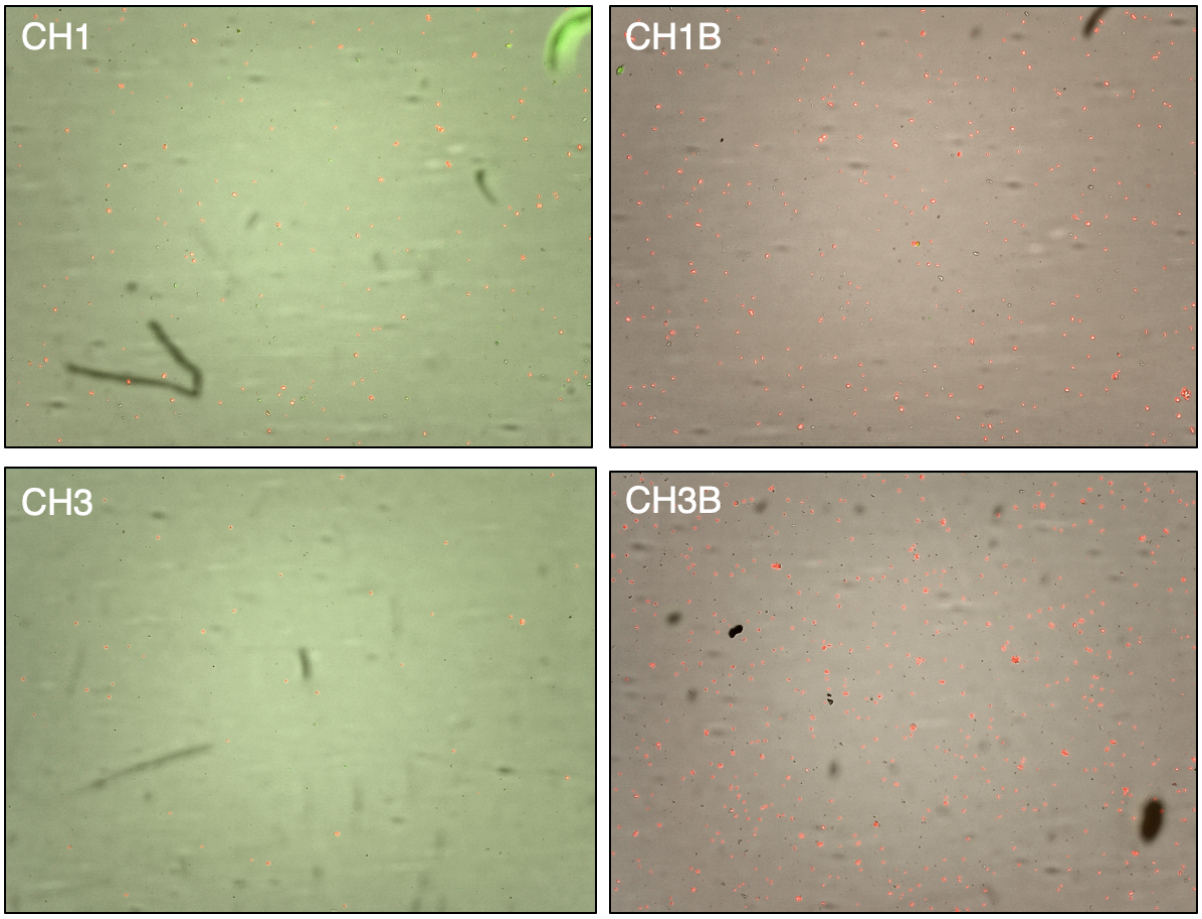
4.6: Acknowledgement and Funding

This research was supported by the NCI Cancer Center Support Grant 5P30CA044579 and by the UVA Genome Analysis and Technology Core, RRID:SCR_018883. Clint Miller's lab is funded by the NIH NHLBI grant R01HL148239, the Leducq Foundation PlaqOmics grant, and the NIH MSTP T32GM007267 grant.

4.7: Supplemental Figures



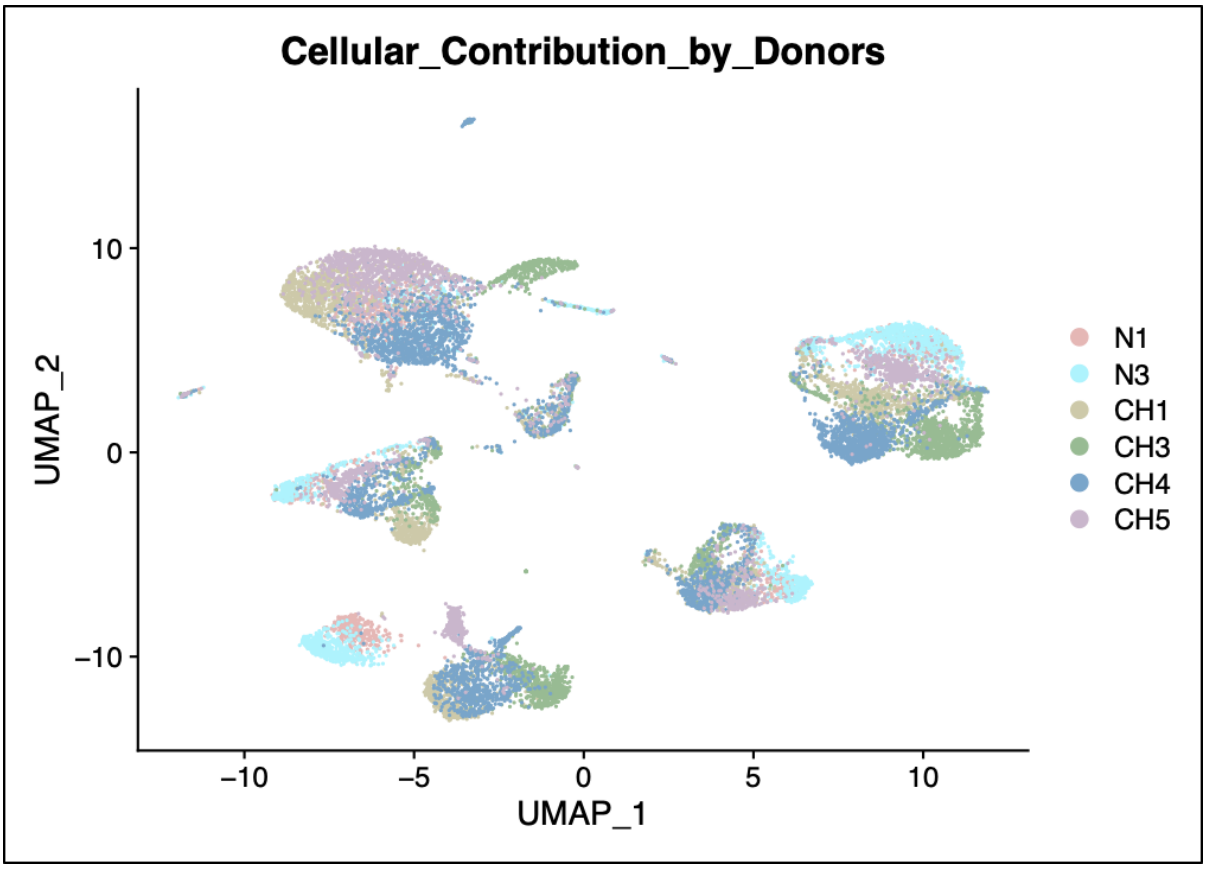
Supplemental Figure 4.1: 10x microscopic image of nuclei isolated from 50µm sections from healthy pilot tissue. Red highlights were automatically generated by the Luna counter to indicate free nuclei.



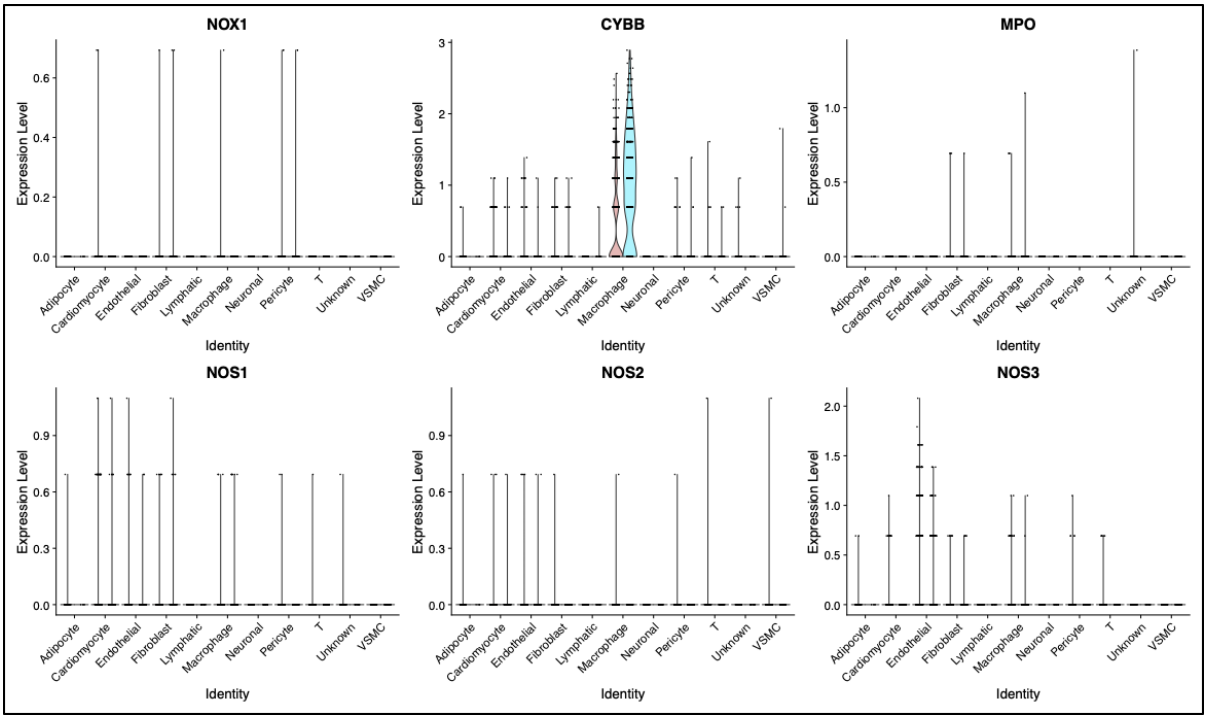
Supplemental Figure 4.2: Representative images of nuclei concentration from two batches of tissue. CH1 and CH3 were run on the first batch, and CH1B and CH3B (adjacent tissues from the same patients) were run on the second batch. Red dye indicates free stained nuclei, and green stains intact cells.


0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Identity
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
Expression
2
1
0
−1
−2

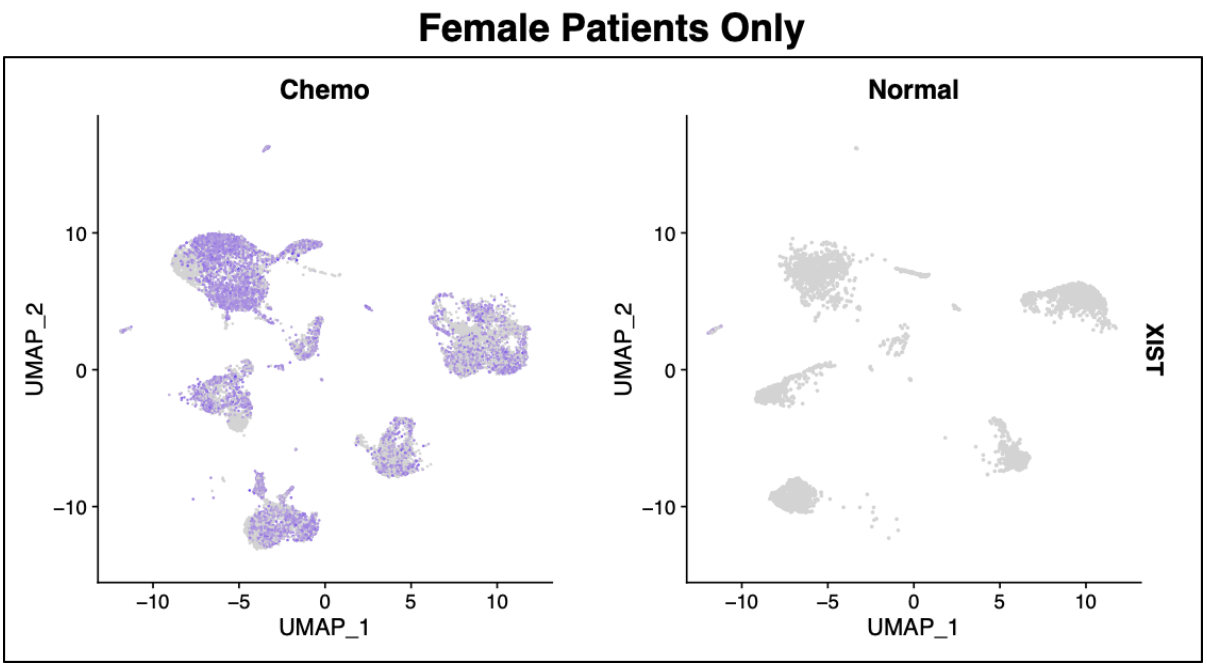
Supplemental Figure 4.3: Top differentially expressed markers generated by the gene expression (snRNA-seq) dataset by cluster.



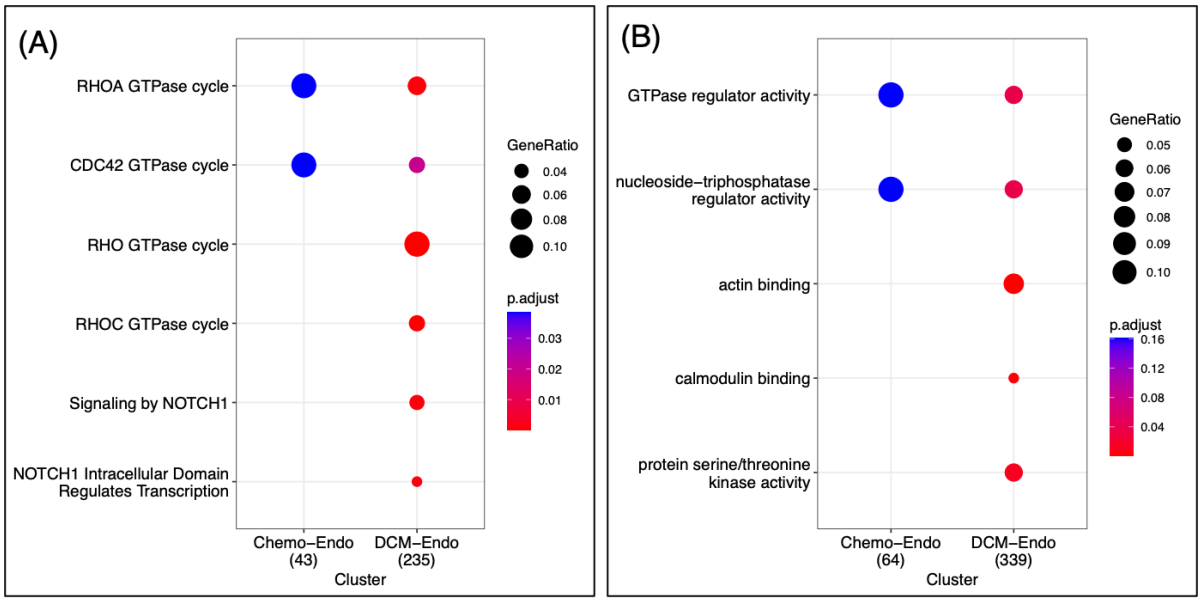
Supplemental Figure 4.4: UMAP projection revealed no significant variations in clusters by patient. In most clusters, all patients were represented. The only notable exception is the fibroblast clusters, where chemotherapy and normal fibroblasts clustered into distinct sub-clusters.



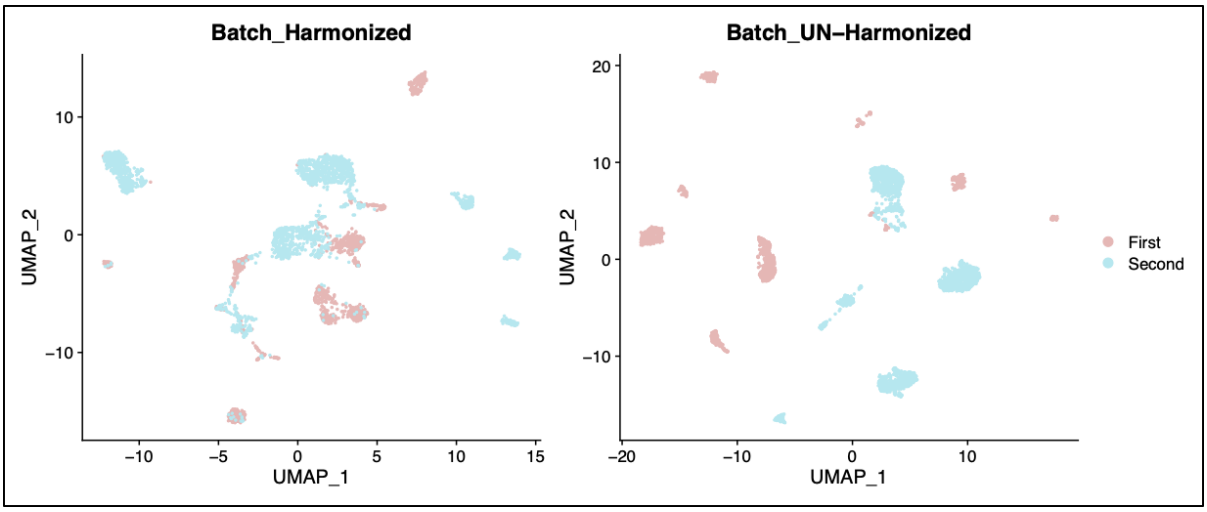
Supplemental Figure 4.5: Major constituents of the redox signaling pathway were not significantly expressed in any cells, except for CYBB (NOX2), a known marker for macrophages. NOX1: NADPH oxidase 1, CYBB is known as NOX2, MPO: myeloperoxidase, NOS1-3: nitric oxide synthase 1-3.



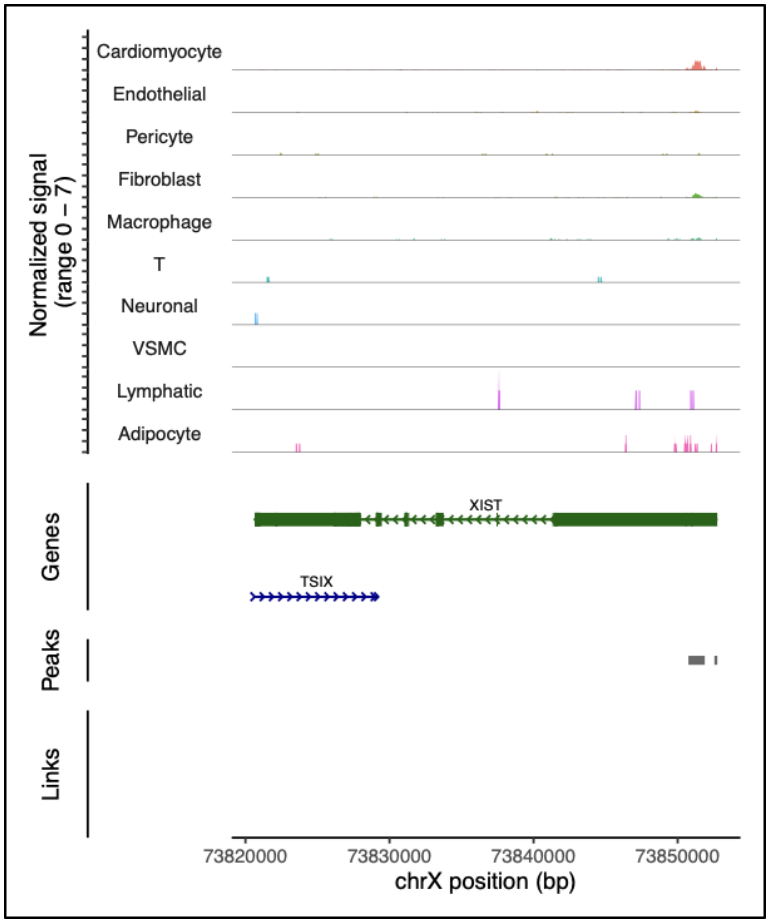
Supplemental Figure 4.6: By subsetting female patients only, we found that XIST is specifically over-represented in the chemotherapy group across all cell types. However, note that we only have one female patient in the normal group versus three in the chemotherapy group.



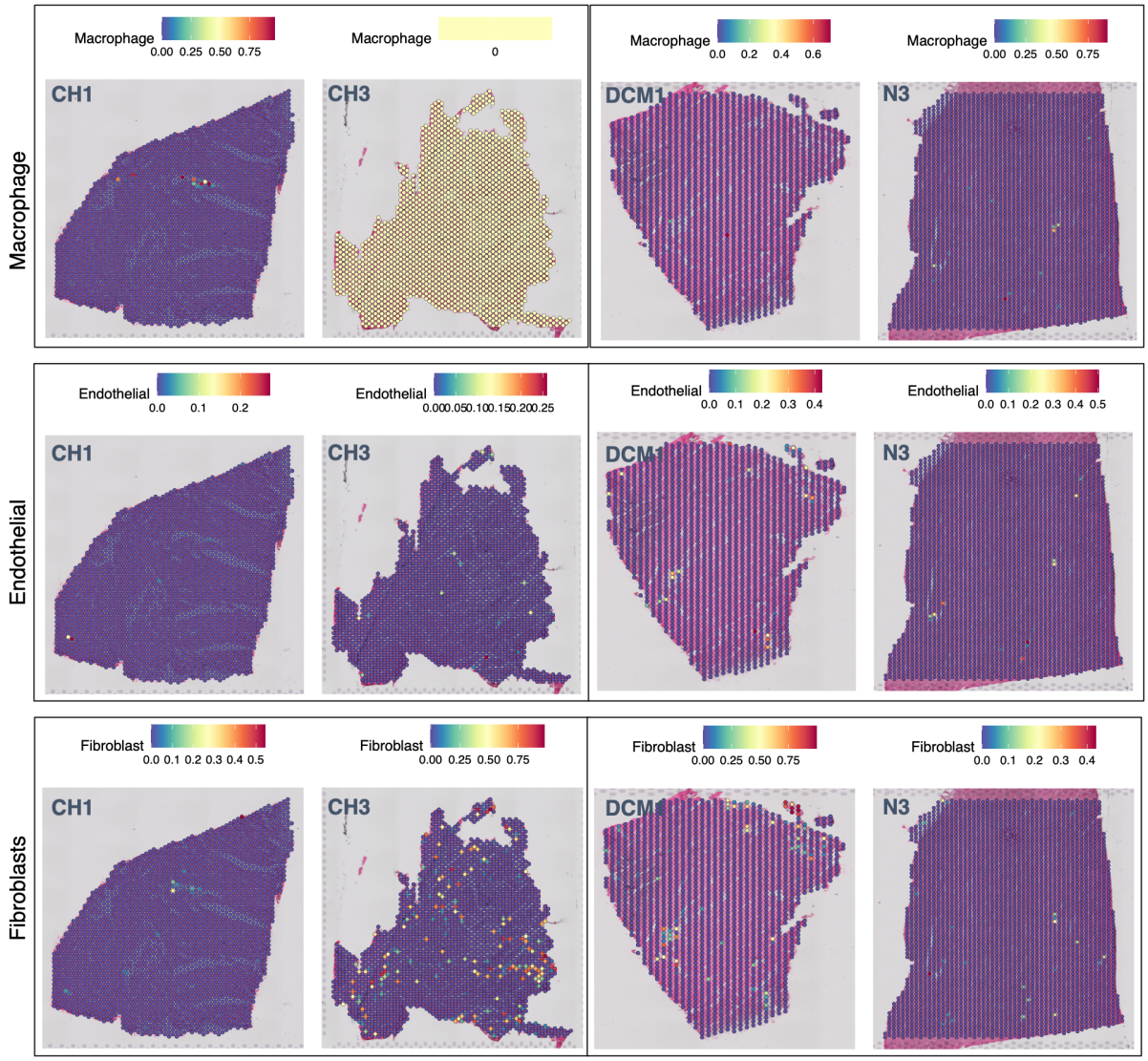
Supplemental Figure 4.7: Endothelial dysfunction could be linked to RHO and GPCR signaling pathways. (A) clusterprofiler result using the reactomePA database and (B) Gene Ontology database.



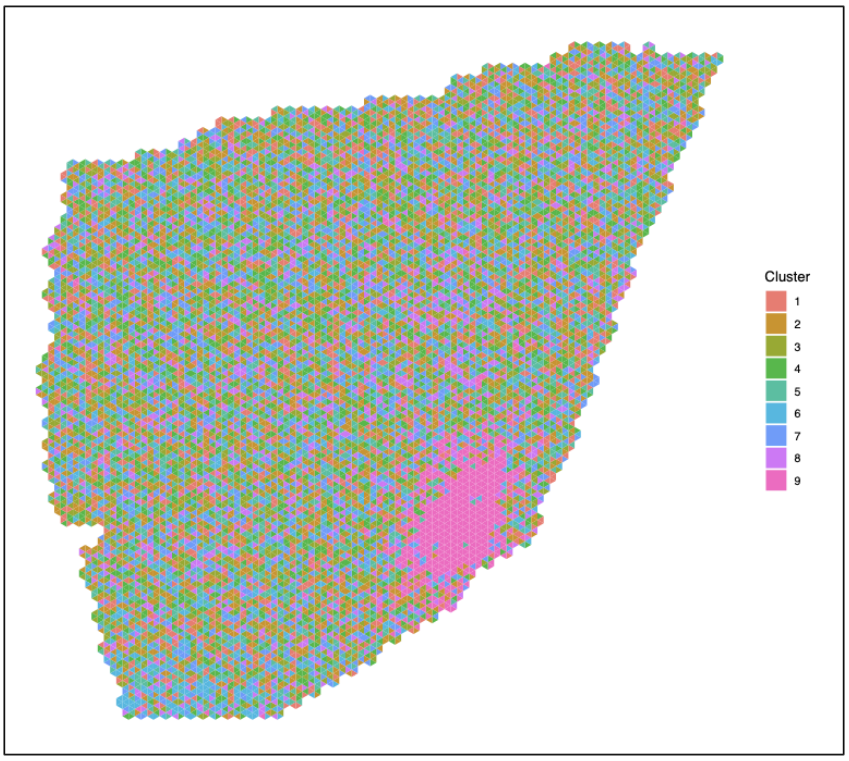
Supplemental Figure 4.8: Strong batch effect was observed (Right) in the snATAC-seq data. However, the two batches of data converged in only two iterations in Harmony (Left).



Supplemental Figure 4.9: open chromatin regions were observed as peaks in the gene region XIST for cardiomyocytes and several other cell types. Unfortunately, we were unable to predict co-accessibility sites (links) or potential drivers for XIST.



Supplemental Figure 4.10: Seurat integration of the spatial RNA-seq data showed no major skewing of major cell types. We note that CH3 exhibited a high number of voxel (spots) of fibroblasts but no focal infiltration was present, suggesting that CH3 is actively and generally inflamed.



Supplemental Figure 4.11: BayesSpace enhancement of voxel resolution in CH1 tissue did not show clear clustering except for one area. Unfortunately, this is likely due to an artifact of sequencing as earlier Seurat clustering did not show any focal areas of cell subsets.

Chapter 5: Future Directions

Role of XIST

Many future directions were presented by the foundational study presented in Chapter 4. In terms of the chemotherapy-induced cardiotoxicity project, one of the most pertinent points of research is the role of *XIST*.

The observation that *XIST* is elevated in chemotherapy LV tissue is equivalent to the problem of the chicken or the egg– which one came first? Is *XIST* a predisposing factor or a response to chemotherapy? Furthermore, what is the role of XIST in the sex-differences of susceptibility? My working hypothesis is that overexpression of *XIST* may sensitize iPSCs for chemotherapy-induced toxicity phenotype.

To test for the former, *XIST* can be first modulated in *in vitro* models of iPSCs, where *XIST* can be transfected with agents like lipofectamine or lentiviral vectors. Alternatively, *XIST* expression can be permanently altered by means of CRISPR technologies. Then, different chemotherapy such as DOX can be used to test for physiologic responses in *XIST*-overexpressed and *XIST*-repressed cells. Simultaneously, the control cells should be measured for *XIST* via qPCR before and after exposure to DOX. Furthermore, *XIST* can be tested in cells with an XY chromosome arrangement to see if overexpression of *XIST* is protective or deleterious to the chemotherapy response. Although single-cell sequencing was used in our study, the modality of validation and sequencing (i.e. bulk RNA-seq, qPCR, or snRNA-seq) should be considered in terms of cost and end points.

In follow up experiments, while mouse models exist to measure chemotherapy toxicity, the best known models are pigs and larger animals, since their cardiovascular system mimics those of humans. In the future, large animal studies can be performed if *in vitro* models show significant results.

Role of GPCR and FA Metabolism

GPCR and FA metabolism pathways were consistently elevated in various cell types in the chemotherapy group. However, the interplay between these pathways within different cells have not been measured. Unfortunately, this is an extremely complicated interaction that is hard to mimic in the *in vitro* settings, even with the advent of cardiac stem cells and cardiac organoids due the fact that multiple cell types need to be present.

One of the most straightforward points of examination would be to inhibit fatty acid metabolism in the heart with chemotherapy challenge. I hypothesize that inhibition of FA metabolism could rescue chemotherapy-induced toxicity phenotype in iPSCs and animal models. Drugs that inhibit fatty acid metabolism in the heart, such as ranolazine or trimetazidine, have been used to treat angina (183). Using the closest animal model and iPSCs available, three arms of experiment can be made: chemotherapy alone, chemotherapy with ranolazine (antioxidant), and ranolazine alone. The last arm should be included because the effect of ranolazine in said animal model may not be fully characterized. End-points for the study should include but not limited to (for cells) electrophysiology, (for animals) ejection fraction, heart histology, and snRNA-seq of the left ventricle.

Effects of Heart-Failure Drugs

Currently, heart failure (HF) from chemotherapy is treated essentially the same as for other kinds of HF. In our analysis in Chapter 4, we showed that common HF drugs may have differential interactions on the heart, depending on how the cardiomyopathy was acquired. Elucidating the specific effects of these HF drugs can help differentiate and stratify HF drug treatment regimes to prioritize HF drugs that are most beneficial in the specific disease context.

While individualized animal model testing is great when dealing with few drugs, HF treatments can be complicated and involve many modalities of therapy, including diuretics, drugs that act on the RAAS system, and others (see Chapter 4 for details). To test for a large number of drugs, CRISPR perturbations in combination with iPSCs would be the most efficient method to first screen out false discoveries presented in Figure 4.14 (58,164,165,197–199). In 96-well plates, differentiated iPSCs can be targeted with control gRNA or gRNAs that knock-out the list of potential targets in Figure 4.14. These cells can then be treated with either control or DOX. Primary functional tests should include measures of apoptosis, ROS generation, mitochondrial dysfunction, calcium handling, and contractility.

An enhanced (but difficult) version of the aforementioned experiment would be to curate isogenic cells from cancer patients. This way, we know that the iPSC cells would inherit the same genetic background (although not the active transcriptome) and may provide a starting point for predictive, personalized medicine.

Final Thoughts on Cardiac Single-Cell Sequencing

At the time of experiment, we did not know if snATAC-seq, snRNA-seq, or Visium spatial RNA-seq would perform better, since the tissues have been aged and kept at -80°C for more than five years. In this study, we demonstrate that snRNA-seq outperforms other modality and this method should be used for subsequent follow up sequence experiments.

For snATAC-seq, we were able to recapitulate the same major cell types, but on its own, the data were extremely sparse and imputed gene scores were difficult to be used to find cell identities, when used alone. Computational calculations with Seurat (75,93) or others like LIGER (200) were required to fully define the identities

of each cluster. Furthermore, the number of nuclei input vs. nuclei that passed QC filter was poor, since we input the maximum allowed nuclei (16,000 per sample), it seemed that the nuclei have been too degraded or not suitable for this kind of multimodal analysis.

Spatial RNA-seq was our backup modality because we did not know whether the nuclei quality was going to be sufficient for single-nucleus sequencing. Theoretically, spatial RNA-seq can provide near-single-cell resolution sequencing data, even if the nuclei/tissue were too degraded for nuclei-based sequencing, since no tissue dissociation is needed. However, we found that although we were able to obtain decent quality sequencing, we were unable to determine spatially variable genes. This could be due to several factors: 1) we did not capture focal changes of inflammation or fibrosis in the small section of LV that we were able to obtain, 2) the changes to the heart from chemotherapy is broad and not focal, and 3) the LV naturally is not spatially variable, unlike many spatial atlases currently published, e.g., on the brain. In summary, at least for follow up experiments, the modality of choice is snRNA-seq, unless macroscopic focal changes in the tissue were observed.

Chapter 6: References

1. Bak S, Gaist D, Sindrup SH, Skytthe A, Christensen K. Genetic Liability in *Stroke*. Stroke (2002) 33:769–774. doi: 10.1161/hs0302.103619

2. Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low- and middle-income countries. *Curr Prob Cardiology* (2010) 35:72-- 115. doi: 10.1016/j.cpcardiol.2009.10.002

3. Control C for D, Kochanek KD, Xu J, Arias E. Mortality in the United States, 2019. Center for Disease Control (2020). https://www.cdc.gov/nchs/products/databriefs/db395.htm [Accessed April 17, 2021]

4. Control C for D. Mortality in the United States, 2020. (2021) doi: 10.15620/cdc:112079

5. Wilson PW. Overview of established risk factors for cardiovascular disease - *UpToDate*. *UptoDate* (2022) https://www.uptodate.com/contents/overview-of-established-risk-factors-for-cardiovascular-disease?search=cvd&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1

6. Amini M, Zayeri F, Salehi M. Trend analysis of cardiovascular disease mortality, incidence, and mortality-to-incidence ratio: results from global burden of disease study 2017. *Bmc Public Health* (2021) 21:401. doi: 10.1186/s12889-021-10429-0

7. Peter W. Overview of established risk factors for cardiovascular disease - UpToDate. *03/30/2022* (2022) https://www.uptodate.com/contents/overview-of-established-risk-factors-for-cardiovascular-disease?search=atherosclerosis%20risk%20factors&source=search_result&selectedTitle=2~150&usage_type=default&display_rank=2#H3 [Accessed September 17, 2022]

8. Cardiovascular diseases (CVDs).

https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [Accessed September 15, 2022]

9. Zuhlke L. Clinical manifestations and diagnosis of rheumatic heart disease - UpToDate. (2022) https://www.uptodate.com/contents/clinical-manifestations-and-diagnosis-of-rheumatic-heart-disease?search=rheumatic%20heart%20disease&source=search_result&selectedTitle=1~126&usage_type=default&display_rank=1 [Accessed November 30, 2022]

10. Buiatti A, Merlo M, Pinamonti B, Biasio MD, Bussani R, Sinagra G. Clinical presentation and long-term follow-up of perimyocarditis. *J Cardiovasc Med* (2013) 14:235–241. doi: 10.2459/jcm.0b013e328351da6e

11. Caforio ALP, Marcolongo R, Jahns R, Fu M, Felix SB, Iliceto S. Immune-mediated and autoimmune myocarditis: clinical presentation, diagnosis and management. *Heart Fail Rev* (2013) 18:715–732. doi: 10.1007/s10741-012-9364-5

12. Doyle T, Kavanaugh-McHugh A. Pathophysiology, clinical features, and diagnosis of tetralogy of Fallot - UpToDate. UptoDate https://www.uptodate.com/contents/pathophysiology-clinical-features-and-diagnosis-of-tetralogy-of-fallot?search=tof&source=search_result&selectedTitle=1~68&usage_type=default&display_rank=1 [Accessed November 30, 2022]

13. Puri K, Allen HD, Qureshi AM. Congenital Heart Disease. *Pediatr Rev* (2017) 38:471–486. doi: 10.1542/pir.2017-0032

14. Libby P, Buring JE, Badimon L, Hansson GK, Deanfield J, Bittencourt MS, Tokgözoğlu L, Lewis EF. *Atherosclerosis*. *Nat Rev Dis Primers* (2019) 5:56. doi: 10.1038/s41572-019-0106-z

15. Zhao X-Q. *Pathogenesis of atherosclerosis*. Kask JC, Libby P, editors. (2022). https://www.uptodate.com/contents/pathogenesis-of-atherosclerosis?search=atherosclerosis&source=search_result&selectedTitle=1~150&usage_type=default&displa

y_rank=1 [Accessed August 1, 2022]

16. Tuzcu EM, Kapadia SR, Tutar E, Ziada KM, Hobbs RE, McCarthy PM, Young JB, Nissen SE. High Prevalence of Coronary Atherosclerosis in Asymptomatic Teenagers and Young Adults. *Circulation* (2001) 103:2705–2710. doi: 10.1161/01.cir.103.22.2705

17. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, Kundu R, Nagao M, Coller J, Koyano TK, et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med* (2019) 25:1280--1289. doi: 10.1038/s41591-019-0512-5

18. Cifu AS, Davis AM. Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults. *Jama* (2017) 318:2132. doi: 10.1001/jama.2017.18706

19. Mega JL, Stitziel NO, Smith JG, Chasman DI, Caulfield MJ, Devlin JJ, Nordio F, Hyde CL, Cannon CP, Sacks FM, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet* (2015) 385:2264–2271. doi: 10.1016/s0140-6736(14)61730-x

20. Jackson R, Lawes CM, Bennett DA, Milne RJ, Rodgers A. Treatment with drugs to lower blood pressure and blood cholesterol based on an individual's absolute cardiovascular risk. Lancet (2005) 365:434–441. doi: 10.1016/s0140-6736(05)17833-7

21. Wong D, Turner AW, Miller CL. Genetic Insights Into Smooth Muscle *Cell* Contributions to Coronary Artery Disease. *Arteriosclerosis Thrombosis Vasc Biology* (2019) 39:1006–1017. doi: 10.1161/atvbaha.119.312141

22. Turner AW, Hu SS, Mosquera JV, Ma WF, Hodonsky CJ, Wong D, Auguste G, Song Y, Sol-Church K, Farber E, et al. Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat Genet* (2022)1–13. doi: 10.1038/s41588-022-01069-0

23. Dobnikar L, Taylor AL, Chappell J, Oldach P, Harman JL, Oerton E, Dzierzak E,

Bennett MR, Spivakov M, Jørgensen HF. Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nat Commun* (2018) 9:4567. doi: 10.1038/s41467-018-06891-x

24. Wang Y, Nanda V, Direnzo D, Ye J, Xiao S, Kojima Y, Howe KL, Jarr K-U, Flores AM, Tsantilas P, et al. Clonally expanding smooth muscle cells promote atherosclerosis by escaping efferocytosis and activating the complement cascade. *P Natl Acad Sci Usa* (2020) 117:15818–15826. doi: 10.1073/pnas.2006348117

25. Motterle A, Pu X, Wood H, Xiao Q, Gor S, Ng FL, Chan K, Cross F, Shohreh B, Poston RN, et al. Functional analyses of coronary artery disease associated variation on chromosome 9p21 in vascular smooth muscle cells. *Hum Mol Genet* (2012) 21:4021–4029. doi: 10.1093/hmg/dds224

26. Wolf D, Ley K. Immunity and Inflammation in Atherosclerosis. *Circ Res* (2019) 124:315–327. doi: 10.1161/circresaha.118.313591

27. Ma WF, Hodonsky CJ, Turner AW, Wong D, Song Y, Mosquera JV, Ligay AV, Slenders L, Gancayco C, Pan H, et al. Enhanced single-cell RNA-seq workflow reveals coronary artery disease cellular cross-talk and candidate drug targets. Atherosclerosis (2021) 340:12–22. doi: 10.1016/j.atherosclerosis.2021.11.025

28. Baylis RA, Gomez D, Mallat Z, Pasterkamp G, Owens GK. The CANTOS Trial. Arteriosclerosis Thrombosis Vasc Biology (2018) 37:e174–e177. doi: 10.1161/atvbaha.117.310097

29. Ma WF, Boudreau HE, Leto TL. Pan-Cancer Analysis Shows TP53 Mutations Modulate the Association of NOX4 with Genetic Programs of Cancer Progression and Clinical Outcome. *Antioxidants* (2021) 10:235. doi: 10.3390/antiox10020235

30. Tapia-Vieyra JV, Delgado-Coello B, Mas-Oliva J. Atherosclerosis and Cancer; A Resemblance with Far-reaching Implications. *Arch Med Res* (2017) 48:12–26. doi: 10.1016/j.arcmed.2017.03.005

31. Zhao G, Lu H, Liu Y, Zhao Y, Zhu T, Garcia-Barrio MT, Chen YE, Zhang J. Single-Cell Transcriptomics Reveals Endothelial Plasticity During Diabetic

Atherogenesis. *Frontiers Cell Dev Biology* (2021) 9:689469. doi: 10.3389/fcell.2021.689469

32. Wilson P. Overview of possible risk factors for cardiovascular disease - UpToDate. UpToDate (2022) https://www.uptodate.com/contents/overview-of-possible-risk-factors-for-cardiovascular-disease?search=genetics%20of%20cvd&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1#H1267867 [Accessed September 17, 2022]

33. Erbilgin A, Civelek M, Romanoski CE, Pan C, Hagopian R, Berliner JA, Lusis AJ. Identification of CAD candidate genes in GWAS loci and their expression in vascular cells. *J Lipid Res* (2013) 54:1894–1905. doi: 10.1194/jlr.m037085

34. Vattikuti S, Guo J, Chow CC. Heritability and Genetic Correlations Explained by Common SNPs for Metabolic Syndrome Traits. *Plos Genet* (2012) 8:e1002637. doi: 10.1371/journal.pgen.1002637

35. Barbalic M, Reiner AP, Wu C, Hixson JE, Franceschini N, Eaton CB, Heiss G, Couper D, Mosley T, Boerwinkle E. Genome-Wide Association Analysis of Incident Coronary Heart Disease (CHD) in African Americans: A Short Report. Plos Genet (2011) 7:e1002199. doi: 10.1371/journal.pgen.1002199

36. Schunkert H, Scheidt M von, Kessler T, Stiller B, Zeng L, Vilne B. Genetics of coronary artery disease in the light of genome-wide association studies. *Clin Res Cardiol Official J Ger Cardiac Soc* (2018) 107:2--9. doi: 10.1007/s00392-018-1324-1

37. Krause MD, Huang R-T, Wu D, Shentu T-P, Harrison DL, Whalen MB, Stolze LK, Rienzo AD, Moskowitz IP, Civelek M, et al. Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics. P Natl Acad Sci Usa (2018) 115:201810568. doi: 10.1073/pnas.1810568115

38. Howson JMM, Zhao W, Barnes DR, Ho W-K, Young R, Paul DS, Waite LL, Freitag DF, Fauman EB, Salfati EL, et al. Fifteen new risk loci for coronary artery

disease highlight arterial-wall-specific mechanisms. Nat Genet (2017) 49:1113--1119. doi: 10.1038/ng.3874

39. Mullen M, Jim J. Management of asymptomatic extracranial carotid atherosclerotic disease - UpToDate. (2022) https://www.uptodate.com/contents/management-of-asymptomatic-extracranial-carotid-atherosclerotic-disease?search=atherosclerosis%20treatment&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1 [Accessed September 26, 2022]

40. Rosenson R. Mechanisms of benefit of lipid-lowering drugs in patients with coronary heart disease - UpToDate. (2022) https://www.uptodate.com/contents/mechanisms-of-benefit-of-lipid-lowering-drugs-in-patients-with-coronary-heart-disease?search=atherosclerosis%20treatment&source=search_result&selectedTitle=2~150&usage_type=default&display_rank=2 [Accessed September 26, 2022]

41. Sirtori CR. The pharmacology of statins. *Pharmacol Res* (2014) 88:3–11. doi: 10.1016/j.phrs.2014.03.002

42. Crisby M, Nordin-Fredriksson G, Shah PK, Yano J, Zhu J, Nilsson J. Pravastatin Treatment Increases Collagen Content and Decreases Lipid Content, Inflammation, Metalloproteinases, and Cell Death in Human Carotid Plaques. Circulation (2001) 103:926–933. doi: 10.1161/01.cir.103.7.926

43. John S, Schlaich M, Langenfeld M, Weihprecht H, Schmitz G, Weidinger G, Schmieder RE. Increased Bioavailability of Nitric Oxide After Lipid-Lowering Therapy in Hypercholesterolemic Patients. Circulation (1998) 98:211–216. doi: 10.1161/01.cir.98.3.211

44. Deedwania P. Statin therapy in patients with heart failure - UpToDate. (2022) https://www.uptodate.com/contents/statin-therapy-in-patients-with-heart-failure?sectionName=POTENTIAL%20BENEFITS%20AND%20HARMS%20OF%20STATIN%20THERAPY%20IN%20HF&search=atherosclerosis%20treatment&topicRef=4546&

anchor=H3&source=see_link#H3 [Accessed September 26, 2022]

45. Basile J, Bloch M. Overview of hypertension in adults - UpToDate. (2022) https://www.uptodate.com/contents/overview-of-hypertension-in-adults?sectionName=TREATMENT&search=atherosclerosis%20treatment&topicRef=1113&anchor=H24&source=see_link#H28 [Accessed September 26, 2022]

46. Colucci WS. Prognosis of heart failure - UpToDate. (2022) https://www.uptodate.com/contents/prognosis-of-heart-failure?search=mortality%20heart%20failure&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1#H10440925 [Accessed September 26, 2022]

47. Group* TCTS. Effects of Enalapril on Mortality in Severe Congestive Heart Failure. *New Engl J Medicine* (1987) 316:1429–1435. doi: 10.1056/nejm198706043162301

48. Gorre F, Vandekerckhove H. Beta-blockers: focus on mechanism of action Which beta-blocker, when and why? *Acta Cardiol* (2017) 65:565–570. doi: 10.1080/ac.65.5.2056244

49. Zaha VG, Meijers WC, Moslehi J. Cardio-Immuno-Oncology. Circulation (2020) 141:87–89. doi: 10.1161/circulationaha.119.042276

50. Peretto G, Lazzeroni D, Sartorio CL, Camici PG. Cardiotoxicity in oncology and coronary microcirculation future challenges in theranostics nbsp. *Front Biosci* (2017) 22:1760–1773. doi: 10.2741/4570

51. Han X, Zhou Y, Liu W. Precision cardio-oncology: understanding the cardiotoxicity of cancer therapy. *Npj Precis Oncol* (2017) 1:31. doi: 10.1038/s41698-017-0034-x

52. Herrmann J. Adverse cardiac effects of cancer therapies: cardiotoxicity and arrhythmia. *Nat Rev Cardiol* (2020) 17:474–502. doi: 10.1038/s41569-020-0348-1

53. Curigliano G, Cardinale D, Dent S, Criscitiello C, Aseyev O, Lenihan D, Cipolla CM. Cardiotoxicity of anticancer treatments: Epidemiology, detection, and management. *Ca Cancer J Clin* (2016) 66:309–325. doi: 10.3322/caac.21341

54. Moslehi JJ. Cardiovascular Toxic Effects of Targeted Cancer Therapies. New Engl J Medicine (2016) 375:1457–1467. doi: 10.1056/nejmra1100265

55. Schultheiss H-P, Fairweather D, Caforio ALP, Escher F, Hershberger RE, Lipshultz SE, Liu PP, Matsumori A, Mazzanti A, McMurray J, et al. Dilated cardiomyopathy. Nat Rev Dis Primers (2019) 5:32. doi: 10.1038/s41572-019-0084-1

56. Liu Y, Zhang W, Hu T, Ni J, Xu B, Huang W. A Doxorubicin-Induced Murine Model of Dilated Cardiomyopathy In Vivo. *J Vis Exp* (2020) doi: 10.3791/61158

57. Alimadadi A, Munroe PB, Joe B, Cheng X. Meta-Analysis of Dilated Cardiomyopathy Using Cardiac RNA-Seq Transcriptomic Datasets. *Genes-basel* (2020) 11:60. doi: 10.3390/genes11010060

58. Huang KM, Thomas MZ, Magdy T, Eisenmann ED, Uddin ME, DiGiacomo DF, Pan A, Keiser M, Otter M, Xia SH, et al. Targeting OCT3 attenuates doxorubicin-induced cardiac injury. *Proc National Acad Sci* (2021) 118:e2020168118. doi: 10.1073/pnas.2020168118

59. Higgins AY, O'Halloran TD, Chang JD. Chemotherapy-induced cardiomyopathy. Heart Fail Rev (2015) 20:721–730. doi: 10.1007/s10741-015-9502-y

60. Oliveira GH, Qattan MY, Al-Kindi S, Park SJ. Advanced Heart Failure Therapies for Patients With Chemotherapy-Induced Cardiomyopathy. *Circulation Hear Fail* (2014) 7:1050–1058. doi: 10.1161/circheartfailure.114.001292

61. Herrmann J, Yang EH, Iliescu CA, Cilingiroglu M, Charitakis K, Hakeem A, Toutouzas K, Leesar MA, Grines CL, Marmagkiolis K. Vascular Toxicities of Cancer Therapies. Circulation (2016) 133:1272–1289. doi: 10.1161/circulationaha.115.018347

62. Neilan TG, Asnani A, Tripathy D, Scherrer-Crosbie M. Prevention and management of anthracycline cardiotoxicity - UpToDate. UpToDate (2021) https://www.uptodate.com/contents/prevention-and-management-of-anthracycline-cardiotoxicity?search=cardiotoxicity%20chemotherapy&source=search_result&selectedTitle=3~150&usage_type=default&display_rank=3 [Accessed May 23, 2021]

63. Marks LB, Constine L, Adams MJ. *Cardiotoxicity of radiation therapy for Hodgkin lymphoma and pediatric malignancies* - UpToDate. Cardiotoxicity of radiation therapy for Hodgkin lymphoma and pediatric malignancies https://www.uptodate.com/contents/cardiotoxicity-of-radiation-therapy-for-hodgkin-lymphoma-and-pediatric-malignancies?search=cardiotoxicity%20children&source=search_result&selectedTitle=4~150&usage_type=default&display_rank=4#H2335092975 [Accessed June 4, 2021]

64. Aissa AF, Islam ABMMK, Ariss MM, Go CC, Rader AE, Conrardy RD, Gajda AM, Rubio-Perez C, Valyi-Nagy K, Pasquinelli M, et al. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nat Commun (2021) 12:1628. doi: 10.1038/s41467-021-21884-z

65. Gawel DR, Serra-Musach J, Lilja S, Aagesen J, Arenas A, Asking B, Bengnér M, Björkander J, Biggs S, Ernerudh J, et al. A validated single-cell-based strategy to identify diagnostic and therapeutic targets in complex diseases. *Genome Med* (2019) 11:47. doi: 10.1186/s13073-019-0657-3

66. Heath JR, Ribas A, Mischel PS. Single-cell analysis tools for drug discovery and development. *Nat Rev Drug Discov* (2016) 15:204–216. doi: 10.1038/nrd.2015.16

67. Ravindra N, Sehanobish A, Pappalardo JL, Hafler DA, Dijk D van. Disease State Prediction From Single-Cell Data Using Graph Attention Networks. *Proc Acm Conf Heal Inference Learn* (2020)121–130. doi: 10.1145/3368555.3384449

68. Chaffin M, Papangeli I, Simonson B, Akkad A-D, Hill MC, Arduini A, Fleming SJ, Melanson M, Hayat S, Kost-Alimova M, et al. Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. *Nature* (2022) 608:174–180. doi: 10.1038/s41586-022-04817-8

69. Nicin L, Schroeter SM, Glaser SF, Schulze-Brüning R, Pham M-D, Hille SS, Yekelchyk M, Kattih B, Abplanalp WT, Tombor L, et al. A human cell atlas of the pressure-induced hypertrophic heart. *Nat Cardiovasc Res* (2022) 1:174–185. doi:

10.1038/s44161-022-00019-7

70. Reichart D, Lindberg EL, Maatz H, Miranda AMA, Viveiros A, Shvetsov N, Gärtner A, Nadelmann ER, Lee M, Kanemaru K, et al. Pathogenic variants damage cell composition and single cell transcription in cardiomyopathies. *Science* (2022) 377:eabo1984. doi: 10.1126/science.abo1984

71. Koenig AL, Shchukina I, Amrute J, Andhey PS, Zaitsev K, Lai L, Bajpai G, Bredemeyer A, Smith G, Jones C, et al. Single-cell transcriptomics reveals cell-type-specific diversification in human heart failure. Nat Cardiovasc Res (2022) 1:263–280. doi: 10.1038/s44161-022-00028-6

72. Litviňuková M, Talavera-López C, Maatz H, Reichart D, Worth CL, Lindberg EL, Kanda M, Polanski K, Heinig M, Lee M, et al. Cells of the adult human heart. Nature (2020)1–10. doi: 10.1038/s41586-020-2797-4

73. Genet G, Guilbeau-Frugier C, Honton B, Dague E, Schneider MD, Coatrieux C, Calise D, Cardin C, Nieto C, Payré B, et al. Ephrin-B1 Is a Novel Specific Component of the Lateral Membrane of the Cardiomyocyte and Is Essential for the Stability of Cardiac Tissue Architecture Cohesion. Circ Res (2012) 110:688–700. doi: 10.1161/circresaha.111.262451

74. Zappia L, Phipson B, Oshlack A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *Plos Comput Biol* (2018) 14:e1006245. doi: 10.1371/journal.pcbi.1006245

75. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* (2018) 36:411--420. doi: 10.1038/nbt.4096

76. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Biorxiv* (2019)820936. doi: 10.1101/820936

77. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al. Integrative Analysis of Complex Cancer

Genomics and Clinical Profiles Using the cBioPortal. *Sci Signal* (2013) 6:pl1–pl1. doi: 10.1126/scisignal.2004088

78. Basatemur GL, Jørgensen HF, Clarke MCH, Bennett MR, Mallat Z. Vascular smooth muscle cells in atherosclerosis. Nat Rev Cardiol (2019) 16:727–744. doi: 10.1038/s41569-019-0227-9

79. Argraves WS, Tanaka A, Smith EP, Twal WO, Argraves KM, Fan D, Haudenschild CC. Fibulin-1 and fibrinogen in human atherosclerotic lesions. *Histochem Cell Biol* (2009) 132:559–565. doi: 10.1007/s00418-009-0628-7

80. Bennett MR, Sinha S, Owens GK. Vascular Smooth Muscle Cells in Atherosclerosis. Circ Res (2016) 118:692-- 702. doi: 10.1161/circresaha.115.306361

81. Pan H, Xue C, Auerbach BJ, Fan J, Bashore AC, Cui J, Yang DY, Trignano SB, Liu W, Shi J, et al. Single-Cell Genomics Reveals a Novel Cell State During Smooth Muscle Cell Phenotypic Switching and Potential Therapeutic Targets for Atherosclerosis in Mouse and Human. Circulation (2020) doi: 10.1161/circulationaha.120.048378

82. Yifan C, Fan Y, Jun P. Visualization of cardiovascular development, physiology and disease at the single-cell level: Opportunities and future challenges. *J Mol Cell Cardiol* (2020) 142:80–92. doi: 10.1016/j.yjmcc.2020.03.005

83. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* (2019) 20:163–172. doi: 10.1038/s41590-018-0276-y

84. DiRenzo D, Owens GK, Leeper NJ. “Attack of the Clones.” Circ Res (2017) 120:624--626. doi: 10.1161/circresaha.116.310091

85. Pliner HA, Shendure J, Trapnell C. Supervised classification enables rapid annotation of cell atlases. *Nat Methods* (2019) 16:983–986. doi: 10.1038/s41592-019-0535-3

86. Shao X, Liao J, Lu X, Xue R, Ai N, Fan X. scCATCH: automatic annotation on

cell-types of clusters from single-cell RNA-seq data. *Iscience* (2020) 23:100882. doi: 10.1016/j.isci.2020.100882

87. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zagar M, et al. Integrated analysis of multimodal single-cell data. Biorxiv (2020)2020.10.12.335331. doi: 10.1101/2020.10.12.335331

88. Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, Zager MA, Aldinger KA, Blecher-Gonen R, Zhang F, et al. A human cell atlas of fetal gene expression. Science (2020) 370:eaba7721. doi: 10.1126/science.aba7721

89. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature (2019) 566:496–502. doi: 10.1038/s41586-019-0969-x

90. Manno GL, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A, et al. RNA velocity of single cells. Nature (2018) 560:494--498. doi: 10.1038/s41586-018-0414-6

91. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, Debnath O, Thürmann L, Kurth F, Völker MT, et al. COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. Nat Biotechnol (2020)1–10. doi: 10.1038/s41587-020-0602-4

92. Alencar GF, Owsiany KM, K S, Sukhavasi K, Mocci G, Nguyen A, Williams CM, Shamsuzzaman S, Mokry M, Henderson CA, et al. The Stem Cell Pluripotency Genes Klf4 and Oct4 Regulate Complex SMC Phenotypic Changes Critical in Late-Stage Atherosclerotic Lesion Pathogenesis. Circulation (2020) doi: 10.1161/circulationaha.120.046672

93. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell (2019) 177:1888-1902.e21. doi: 10.1016/j.cell.2019.05.031

94. Consortium TTS, Quake SR. The Tabula Sapiens: a single cell transcriptomic

atlas of multiple organs from individual human donors. Biorxiv (2021)2021.07.19.452956. doi: 10.1101/2021.07.19.452956

95. Huang Q, Liu Y, Du Y, Garmire LX. Evaluation of Cell Type Annotation R Packages on Single-cell RNA-seq Data. *Genom Proteom Bioinform* (2020) doi: 10.1016/j.gpb.2020.07.004

96. Perisic L, Aldi S, Sun Y, Folkersen L, Razuvaev A, Roy J, Lengquist M, Åkesson S, Wheelock CE, Maegdefessel L, et al. Gene expression signatures, pathways and networks in carotid atherosclerosis. *J Intern Med* (2016) 279:293-- 308. doi: 10.1111/joim.12448

97. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *Bmc Genomics* (2018) 19:477. doi: 10.1186/s12864-018-4772-0

98. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, Rajewsky N, Simon L, Theis FJ. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* (2019) 20:59. doi: 10.1186/s13059-019-1663-x

99. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol (2019) 37:547–554. doi: 10.1038/s41587-019-0071-9

100. Cannoodt R, Saelens W, Sichien D, Tavernier S, Janssens S, Guilliams M, Lambrecht B, Preter KD, Saeys Y. SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. Biorxiv (2016)079509. doi: 10.1101/079509

101. Shankman LS, Gomez D, Cherepanova OA, Salmon M, Alencar GF, Haskins RM, Swiatlowska P, Newman AAC, Greene ES, Straub AC, et al. KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. Nat Med (2015) 21:628-- 637. doi: 10.1038/nm.3866

102. Carter S, Lemieux I, Li Z, Alméras N, Tremblay A, Bergeron J, Poirier P,

Després J-P, Picard F. Changes in IGFBP-2 levels following a one-year lifestyle modification program are independently related to improvements in plasma apo B and LDL apo B levels. Atherosclerosis (2018) 281:89–97. doi: 10.1016/j.atherosclerosis.2018.12.016

103. Carter AM. Complement Activation: An Emerging Player in the Pathogenesis of Cardiovascular Disease. Sci (2012) 2012:1–14. doi: 10.6064/2012/402783

104. Döring Y, Vorst EPC van der, Duchene J, Jansen Y, Gencer S, Bidzhekov K, Atzler D, Santovito D, Rader DJ, Saleheen D, et al. CXCL12 Derived From Endothelial Cells Promotes Atherosclerosis to Drive Coronary Artery Disease. Circulation (2019) 139:1338–1340. doi: 10.1161/circulationaha.118.037953

105. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res* (2017) 45:D362–D368. doi: 10.1093/nar/gkw937

106. Farbehi N, Patrick R, Dorison A, Xaymardan M, Janbandhu V, Wystub-Lis K, Ho JW, Nordon RE, Harvey RP. Single-cell expression profiling reveals dynamic flux of cardiac stromal, vascular and immune cells in health and injury. *Elife* (2019) 8:e43882. doi: 10.7554/elife.43882

107. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan C-H, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. Nat Commun (2021) 12:1088. doi: 10.1038/s41467-021-21246-9

108. Yang T-L, Lee P-L, Lee D-Y, Wang W-L, Wei S-Y, Lee C-I, Chiu J-J. Differential regulations of fibronectin and laminin in Smad2 activation in vascular endothelial cells in response to disturbed flow. *J Biomed Sci* (2018) 25:1. doi: 10.1186/s12929-017-0402-4

109. Xaus J, Comalada M, Barrachina M, Herrero C, Goñalons E, Soler C, Lloberas J, Celada A. The Expression of MHC Class II Genes in Macrophages Is Cell Cycle Dependent. *J Immunol* (2000) 165:6364–6371. doi: 10.4049/jimmunol.165.11.6364

110. Xu B-F, Liu R, Huang C-X, He B-S, Li G-Y, Sun H-S, Feng Z-P, Bao M-H. Identification of key genes in ruptured atherosclerotic plaques by weighted gene correlation network analysis. *Sci Rep-uk* (2020) 10:10847. doi: 10.1038/s41598-020-67114-2

111. Cotto KC, Wagner AH, Feng Y-Y, Kiwala S, Coffman AC, Spies G, Wollam A, Spies NC, Griffith OL, Griffith M. DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. Nucleic Acids Res (2017) 46:D1068–D1073. doi: 10.1093/nar/gkx1143

112. Buono C, Come CE, Witztum JL, Maguire GF, Connelly PW, Carroll M, Lichtman AH. Influence of C3 Deficiency on Atherosclerosis. Circulation (2002) 105:3025–3031. doi: 10.1161/01.cir.0000019584.04929.83

113. Mastaglio S, Ruggeri A, Risitano AM, Angelillo P, Yancopoulou D, Mastellos DC, Huber-Lang M, Piemontese S, Assanelli A, Garlanda C, et al. The first case of COVID-19 treated with the complement C3 inhibitor AMY-101. *Clin Immunol* (2020) 215:108450. doi: 10.1016/j.clim.2020.108450

114. Java A, Apicelli AJ, Liszewski MK, Coler-Reilly A, Atkinson JP, Kim AHJ, Kulkarni HS. The complement system in COVID-19: friend and foe? *Jci Insight* (2020) 5: doi: 10.1172/jci.insight.140711

115. Zhang R, Song B, Hong X, Shen Z, Sui L, Wang S. microRNA-9 Inhibits Vulnerable Plaque Formation and Vascular Remodeling via Suppression of the SDC2-Dependent FAK/ERK Signaling Pathway in Mice With Atherosclerosis. *Front Physiol* (2020) 11:804. doi: 10.3389/fphys.2020.00804

116. Kuzuya M, Nakamura K, Sasaki T, Cheng XW, Itohara S, Iguchi A. Effect of MMP-2 Deficiency on Atherosclerotic Lesion Formation in ApoE-Deficient Mice. Arteriosclerosis Thrombosis Vasc Biology (2006) 26:1120–1125. doi: 10.1161/01.atv.0000218496.60097.e0

117. Camaré C, Pucelle M, Nègre-Salvayre A, Salvayre R. Angiogenesis in the atherosclerotic plaque. *Redox Biol* (2017) 12:18-- 34. doi:

10.1016/j.redox.2017.01.007

118. Libby P, Hansson GK. Taming Immune and Inflammatory Responses to Treat Atherosclerosis ∗. *J Am Coll Cardiol* (2018) 71:173–176. doi: 10.1016/j.jacc.2017.10.081

119. Mabbott NA, Baillie JK, Brown H, Freeman TC, Hume DA. An expression atlas of human primary cells: inference of gene function from coexpression networks. Bmc Genomics (2013) 14:632. doi: 10.1186/1471-2164-14-632

120. Liu M, Gomez D. Smooth Muscle Cell Phenotypic Diversity: At the Crossroads of Lineage Tracing and Single-Cell Transcriptomics. Arteriosclerosis Thrombosis Vasc Biology (2019) 39:1715–1723. doi: 10.1161/atvbaha.119.312131

121. Lareau CA, Ludwig LS, Muus C, Gohil SH, Zhao T, Chiang Z, Pelka K, Verboon JM, Luo W, Christian E, et al. Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. Nat Biotechnol (2020)1–11. doi: 10.1038/s41587-020-0645-6

122. Xu J, Nuno K, Litzenburger UM, Qi Y, Corces MR, Majeti R, Chang HY. Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA. Elife (2019) 8:e45105. doi: 10.7554/elife.45105

123. Khomtchouk BB, Tran D-T, Vand KA, Might M, Gozani O, Assimes TL. Cardioinformatics: the nexus of bioinformatics and precision cardiology. *Brief Bioinform* (2019) 21:2031–2051. doi: 10.1093/bib/bbz119

124. Khomtchouk BB, Nelson CS, Vand KA, Palmisano S, Grossman RL. HeartBioPortal2.0: new developments and updates for genetic ancestry and cardiometabolic quantitative traits in diverse human populations. *Database* (2020) 2020:baaa115-. doi: 10.1093/database/baaa115

125. Khomtchouk BB, Vand KA, Koehler WC, Tran D-T, Middlebrook K, Sudhakaran S, Nelson CS, Gozani O, Assimes TL. HeartBioPortal: An Internet-of-Omics for Human Cardiovascular Disease Data. *Circulation Genom Precis Medicine* (2019) 12:e002426. doi: 10.1161/circgen.118.002426

126. Skinnider MA, Squair JW, Courtine G. Enabling reproducible re-analysis of single-cell data. Genome Biol (2021) 22:215. doi: 10.1186/s13059-021-02422-y

127. Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, Koval J, Das I, Callaway MB, Eldred JM, et al. DGIdb: mining the druggable genome. Nat Methods (2013) 10:1209–1210. doi: 10.1038/nmeth.2689

128. Canver MC, Bauer DE, Maeda T, Pinello L. DrugThatGene: integrative analysis to streamline the identification of druggable genes, pathways and protein complexes from CRISPR screens. *Bioinformatics* (2018) 35:1981–1984. doi: 10.1093/bioinformatics/bty913

129. Method of the Year 2013. Nat Methods (2014) 11:1–1. doi: 10.1038/nmeth.2801

130. Method of the Year 2019: Single-cell multimodal omics. Nat Methods (2020) 17:1–1. doi: 10.1038/s41592-019-0703-5

131. Zappia L, Theis FJ. Over 1000 tools reveal trends in the single-cell RNA-seq analysis landscape. Genome Biol (2021) 22:301. doi: 10.1186/s13059-021-02519-4

132. Puntambekar S, Hesselberth JR, Riemondy KA, Fu R. Cell-level metadata are indispensable for documenting single-cell sequencing datasets. *Plos Biol* (2021) 19:e3001077. doi: 10.1371/journal.pbio.3001077

133. Li F, Yan K, Wu L, Zheng Z, Du Y, Liu Z, Zhao L, Li W, Sheng Y, Ren L, et al. Single-cell RNA-seq reveals cellular heterogeneity of mouse carotid artery under disturbed flow. *Cell Death Discov* (2021) 7:180. doi: 10.1038/s41420-021-00567-0

134. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol (2020) 38:1408–1414. doi: 10.1038/s41587-020-0591-3

135. Li M, Zhang X, Ang KS, Ling J, Sethi R, Lee NYS, Ginhoux F, Chen J. DISCO: a database of Deeply Integrated human Single-Cell Omics data. Nucleic Acids Res (2021) 50:gkab1020-. doi: 10.1093/nar/gkab1020

136. Li Y, Ren P, Dawson A, Vasquez HG, Ageedi W, Zhang C, Luo W, Chen R, Li Y,

Kim S, et al. Single-Cell Transcriptome Analysis Reveals Dynamic Cell Populations and Differential Gene Expression Patterns in Control and Aneurysmal Human Aortic Tissue. Circulation (2020) 142:1374–1388. doi: 10.1161/circulationaha.120.046528

137. Xu K, Xie S, Huang Y, Zhou T, Liu M, Zhu P, Wang C, Shi J, Li F, Sellke FW, et al. Cell-Type Transcriptome Atlas of Human Aortic Valves Reveal Cell Heterogeneity and Endothelial to Mesenchymal Transition Involved in Calcific Aortic Valve Disease. Arteriosclerosis Thrombosis Vasc Biology (2020) 40:2910–2921. doi: 10.1161/atvbaha.120.314789

138. Andueza A, Kumar S, Kim J, Kang D-W, Mumme HL, Perez JI, Villa-Roel N, Jo H. Endothelial Reprogramming by Disturbed Flow Revealed by Single-Cell RNA and Chromatin Accessibility Study. *Cell Reports* (2020) 33:108491. doi: 10.1016/j.celrep.2020.108491

139. Gu W, Ni Z, Tan Y-Q, Deng J, Zhang S-J, Lv Z-C, Wang X-J, Chen T, Zhang Z, Hu Y, et al. Adventitial Cell Atlas of wt (Wild Type) and ApoE (Apolipoprotein E)-Deficient Mice Defined by Single-Cell RNA Sequencing. Arteriosclerosis Thrombosis Vasc Biology (2019) 39:1055–1071. doi: 10.1161/atvbaha.119.312399

140. Delorey TM, Ziegler CGK, Heimberg G, Normand R, Yang Y, Segerstolpe A, Abbondanza D, Fleming SJ, Subramanian A, Montoro DT, et al. A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-CoV-2. Biorxiv (2021)2021.02.25.430130. doi: 10.1101/2021.02.25.430130

141. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Anagnostopoulos A, Asabor R, Baldarelli RM, Beal JS, Bello SM, et al. Mouse Genome Database (MGD) 2019. Nucleic Acids Res (2019) 47:D801–D806. doi: 10.1093/nar/gky1056

142. DiStefano MT, Goehringer S, Babb L, Alkuraya FS, Amberger J, Amin M, Austin-Tse C, Balzotti M, Berg JS, Birney E, et al. The Gene Curation Coalition: A global effort to harmonize gene–disease evidence resources. *Genet Med* (2022) doi: 10.1016/j.gim.2022.04.017

143. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z,

Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res (2016) 44:W90–W97. doi: 10.1093/nar/gkw377

144. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *Bmc Bioinformatics* (2013) 14:128. doi: 10.1186/1471-2105-14-128

145. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, et al. Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* (2021) 1:e90. doi: 10.1002/cpz1.90

146. Schürmann C, Rezende F, Kruse C, Yasar Y, Löwe O, Fork C, Sluis B van de, Bremer R, Weissmann N, Shah AM, et al. The NADPH oxidase Nox4 has anti-atherosclerotic functions. *Eur Heart J* (2015) 36:3447–3456. doi: 10.1093/eurheartj/ehv460

147. Abdelaal T, Michielsen L, Cats D, Hoogduin D, Mei H, Reinders MJT, Mahfouz A. A comparison of automatic cell identification methods for single-cell RNA sequencing data. Genome Biol (2019) 20:194. doi: 10.1186/s13059-019-1795-z

148. Consortium* TS, Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, Salzman J, Yosef N, Bulthaup B, Brown P, et al. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science (2022) 376:eabl4896. doi: 10.1126/science.abl4896

149. Ekiz HA, Conley CJ, Stephens WZ, O'Connell RM. CIPR: a web-based R/shiny app and R package to annotate cell clusters in single cell RNA sequencing experiments. Bmc Bioinformatics (2020) 21:191. doi: 10.1186/s12859-020-3538-2

150. Tucker NR, Chaffin M, Fleming SJ, Hall AW, Parsons VA, Bedi KC, Akkad A-D, Herndon CN, Arduini A, Papangeli I, et al. Transcriptional and Cellular Diversity of the Human Heart. Circulation (2020) 142:466–482. doi: 10.1161/circulationaha.119.045401

151. Burstein B, Libby E, Calderone A, Nattel S. Differential Behaviors of Atrial

Versus Ventricular Fibroblasts. Circulation (2008) 117:1630–1641. doi: 10.1161/circulationaha.107.748053

152. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol (2014) 32:381--386. doi: 10.1038/nbt.2859

153. Alsaigh T, Evans D, Frankel D, Torkamani A. Decoding the transcriptome of atherosclerotic plaque at single-cell resolution. Biorxiv (2020)2020.03.03.968123. doi: 10.1101/2020.03.03.968123

154. Jia G, Preussner J, Chen X, Guenther S, Yuan X, Yekelchyk M, Kuenne C, Looso M, Zhou Y, Teichmann S, et al. Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. Nat Commun (2018) 9:4877. doi: 10.1038/s41467-018-07307-6

155. Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, Griffith M, Griffith OL, Wagner AH. Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res (2020) 49:gkaa1084-. doi: 10.1093/nar/gkaa1084

156. Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, Darmanis S, Batson J, Botvinnik O, Chen MB, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature (2018) 562:367–372. doi: 10.1038/s41586-018-0590-4

157. Li G, Luan C, Dong Y, Xie Y, Zentz SC, Zelt R, Roach J, Liu J, Qian L, Li Y, et al. ExpressHeart: Web Portal to Visualize Transcriptome Profiles of Non-Cardiomyocyte Cells. *Int J Mol Sci* (2021) 22:8943. doi: 10.3390/ijms22168943

158. Fornaro A, Castelli G, Ciaccheri M, Tomberli B, Olivotto I, Gensini GF. Improving survival of chemotherapy-induced cardiomyopathy in the modern heart failure therapy era. Eur Heart J (2013) 34:P2968–P2968. doi: 10.1093/eurheartj/eht309.p2968

159. Podyacheva EY, Kushnareva EA, Karpov AA, Toropova YG. Analysis of Models of Doxorubicin-Induced Cardiomyopathy in Rats and Mice. A Modern View From the Perspective of the Pathophysiologist and the Clinician. *Front Pharmacol* (2021) 12:670479. doi: 10.3389/fphar.2021.670479

160. Santos DS dos, Goldenberg RC dos S. Cardiotoxicity. (2018) doi: 10.5772/intechopen.79588

161. Scheidt M von, Zhao Y, Kurt Z, Pan C, Zeng L, Yang X, Schunkert H, Lusis AJ. Applications and Limitations of Mouse Models for Understanding Human Atherosclerosis. *Cell Metab* (2017) 25:248–261. doi: 10.1016/j.cmet.2016.11.001

162. Fernandez DM, Giannarelli C. Immune cell profiling in atherosclerosis: role in research and precision medicine. Nat Rev Cardiol (2021)1–16. doi: 10.1038/s41569-021-00589-2

163. Oppi S, Lüscher TF, Stein S. Mouse Models for Atherosclerosis Research—Which Is My Line? *Frontiers Cardiovasc Medicine* (2019) 6:46. doi: 10.3389/fcvm.2019.00046

164. Magdy T, Burridge PW. Use of hiPSC to explicate genomic predisposition to anthracycline-induced cardiotoxicity. *Pharmacogenomics* (2021) 22:41–54. doi: 10.2217/pgs-2020-0104

165. Fonoudi H, Burridge PW. Cellular model systems to study cardiovascular injury from chemotherapy. *J Thromb Thrombolys* (2021) 51:890–896. doi: 10.1007/s11239-020-02299-x

166. Colucci W. *Overview of the management of heart failure with reduced ejection fraction in adults* - UpToDate. Overview of the management of heart failure with reduced ejection fraction in adults (2022) https://www.uptodate.com/contents/overview-of-the-management-of-heart-failure-with-reduced-ejection-fraction-in-adults?search=heart%20failure%20management&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1# [Accessed November 25, 2022]

167. Gladka MM, Molenaar B, Ruiter H de, Elst S van der, Tsui H, Versteeg D, Lacraz GPA, Huibers MMH, Oudenaarden A van, Rooij E van. Single-Cell Sequencing of the Healthy and Diseased Heart Reveals Cytoskeleton-Associated Protein 4 as a New Modulator of Fibroblasts Activation. Circulation (2018) 138:166–180. doi: 10.1161/circulationaha.117.030742

168. Nicin L, Abplanalp WT, Schänzer A, Sprengel A, John D, Mellentin H, Tombor L, Keuper M, Ullrich E, Klingel K, et al. Single Nuclei Sequencing Reveals Novel Insights Into the Regulation of Cellular Signatures in Children With Dilated Cardiomyopathy. Circulation (2021) 143:1704–1719. doi: 10.1161/circulationaha.120.051391

169. Wang L-B, Karpova A, Gritsenko MA, Kyle JE, Cao S, Li Y, Rykunov D, Colaprico A, Rothstein JH, Hong R, et al. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* (2021) 39:509-528.e20. doi: 10.1016/j.ccell.2021.01.006

170. Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, Hughes TK, Wadsworth MH, Burks T, Nguyen LT, et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat Biotechnol (2020) 38:737–746. doi: 10.1038/s41587-020-0465-8

171. Slyper M, Porter CBM, Ashenberg O, Waldman J, Drokhlyansky E, Wakiro I, Smillie C, Smith-Rosario G, Wu J, Dionne D, et al. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nat Med (2020) 26:792--802. doi: 10.1038/s41591-020-0844-1

172. Wang Y, Fan JL, Melms JC, Amin AD, Georgis Y, Ho P, Tagore S, Abril-Rodríguez G, Biermann J, Hofree M, et al. Multi-modal single-cell and whole-genome sequencing of minute, frozen specimens to propel clinical applications. Biorxiv (2022)2022.02.13.480272. doi: 10.1101/2022.02.13.480272

173. Ma Y, Iyer RP, Jung M, Czubryt MP, Lindsey ML. Cardiac Fibroblast Activation Post-Myocardial Infarction: Current Knowledge Gaps. *Trends Pharmacol Sci* (2017)

38:448–458. doi: 10.1016/j.tips.2017.03.001

174. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol (2019) 37:38–44. doi: 10.1038/nbt.4314

175. Hilger D, Masureel M, Kobilka BK. Structure and dynamics of GPCR signaling complexes. *Nat Struct Mol Biol* (2018) 25:4–12. doi: 10.1038/s41594-017-0011-7

176. Wang J, Gareri C, Rockman HA. G-Protein–Coupled Receptors in Heart Disease. Circ Res (2018) 123:716–735. doi: 10.1161/circresaha.118.311403

177. Morita H, Rehm HL, Menesses A, McDonough B, Roberts AE, Kucherlapati R, Towbin JA, Seidman JG, Seidman CE. Shared Genetic Causes of Cardiac Hypertrophy in Children and Adults. New Engl J Medicine (2008) 358:1899–1908. doi: 10.1056/nejmoa075463

178. Sweeney HL, Bowman BF, Stull JT. Myosin light chain phosphorylation in vertebrate striated muscle: regulation and function. *Am J Physiol-cell Ph* (1993) 264:C1085–C1095. doi: 10.1152/ajpcell.1993.264.5.c1085

179. Linari M, Piazzesi G, Pertici I, Dantzig JA, Goldman YE, Lombardi V. Straightening Out the Elasticity of Myosin Cross-Bridges. *Biophys J* (2020) 118:994–1002. doi: 10.1016/j.bpj.2020.01.002

180. Zheng P-F, Chen L-Z, Liu P, Pan H-W. A novel lncRNA-miRNA-mRNA triple network identifies lncRNA XIST as a biomarker for acute myocardial infarction. *Aging Albany Ny* (2022) 14:4085–4106. doi: 10.18632/aging.204075

181. Xiao L, Gu Y, Sun Y, Chen J, Wang X, Zhang Y, Gao L, Li L. The long noncoding RNA XIST regulates cardiac hypertrophy by targeting miR-101. *J Cell Physiol* (2019) 234:13680–13692. doi: 10.1002/jcp.28047

182. D'Amario D, Camilli M, Migliaro S, Canonico F, Galli M, Arcudi A, Montone RA, Borovac JA, Crea F, Savarese G. Sex-Related Differences in Dilated Cardiomyopathy with a Focus on Cardiac Dysfunction in Oncology. *Curr Cardiol Rep* (2020) 22:102. doi: 10.1007/s11886-020-01377-z

183. Stanley WC. Partial fatty acid oxidation inhibitors for stable angina. *Expert Opin Inv Drug* (2002) 11:615–629. doi: 10.1517/13543784.11.5.615

184. Allahverdian S, Chehroudi AC, McManus BM, Abraham T, Francis GA. Contribution of intimal smooth muscle cells to cholesterol accumulation and macrophage-like cells in human atherosclerosis. Circulation (2014) 129:1551--1559. doi: 10.1161/circulationaha.113.005015

185. Winkels H, Ehinger E, Vassallo M, Buscher K, Dinh HQ, Kobiyama K, Hamers AAJ, Cochain C, Vafadarnejad E, Saliba A-E, et al. Atlas of the Immune Cell Repertoire in Mouse Atherosclerosis Defined by Single-Cell RNA-Sequencing and Mass Cytometry. Circ Res (2018) 122:1675–1688. doi: 10.1161/circresaha.117.312513

186. Kim K, Shim D, Lee JS, Zaitsev K, Williams JW, Kim K-W, Jang M-Y, Jang HS, Yun TJ, Lee SH, et al. Transcriptome Analysis Reveals Nonfoamy Rather Than Foamy Plaque Macrophages Are Proinflammatory in Atherosclerotic Murine Models. Circ Res (2018) 123:1127–1142. doi: 10.1161/circresaha.118.312804

187. Thorp EB. Macrophage Metabolic Signaling during Ischemic Injury and Cardiac Repair. *Immunometabolism* (2021) 3:e210018. doi: 10.20900/immunometab20210018

188. Yonezawa T, Kurata R, Yoshida K, Murayama M, Cui X, Hasegawa A. Free Fatty Acids-Sensing G Protein-Coupled Receptors in Drug Targeting and Therapeutics. *Curr Med Chem* (2013) 20:3855–3871. doi: 10.2174/09298673113209990168

189. Morbidelli L, Donnini S, Ziche M. Targeting endothelial cell metabolism for cardio-protection from the toxicity of antitumor agents. *Cardio-oncology* (2016) 2:3. doi: 10.1186/s40959-016-0010-6

190. Szczepaniak P, Siedlinski M, Hodorowicz-Zaniewska D, Nosalski R, Mikolajczyk TP, Dobosz AM, Dikalova A, Dikalov S, Streb J, Gara K, et al. Breast cancer chemotherapy induces vascular dysfunction and hypertension through

NOX4 dependent mechanism. *J Clin Invest* (2022) 132:e149117. doi: 10.1172/jci149117

191. Flentje A, Kalsi R, Monahan TS. Small GTPases and Their Role in Vascular Disease. Int J Mol Sci (2019) 20:917. doi: 10.3390/ijms20040917

192. Parri M, Chiarugi P. Rac and Rho GTPases in cancer cell motility control. *Cell Commun Signal* (2010) 8:23. doi: 10.1186/1478-811x-8-23

193. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141

194. Dries R, Chen J, Rossi N del, Khan MM, Sistig A, Yuan G-C. Advances in spatial transcriptomic data analysis. *Genome Res* (2021) 31:1706–1718. doi: 10.1101/gr.275224.121

195. Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, Williams SR, Uytingco CR, Taylor SEB, Nghiem P, et al. Spatial transcriptomics at subspot resolution with BayesSpace. Nat Biotechnol (2021)1–10. doi: 10.1038/s41587-021-00935-2

196. Elosua-Bayes M, Nieto P, Mereu E, Gut I, Heyn H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. Nucleic Acids Res (2021) doi: 10.1093/nar/gkab043

197. Csöbönyeiová M, Polák Š, Danišovič L. Toxicity testing and drug screening using iPSC-derived hepatocytes, cardiomyocytes, and neural cells. *Can J Physiol Pharm* (2016) 94:687–694. doi: 10.1139/cjpp-2015-0459

198. Collins C, Osborne LD, Guilluy C, Chen Z, O'Brien ET, Reader JS, Burridge K, Superfine R, Tzima E. Haemodynamic and extracellular matrix cues regulate the mechanical phenotype and stiffness of aortic endothelial cells. Nat Commun (2014) 5:3984. doi: 10.1038/ncomms4984

199. Gintant G, Burridge P, Gepstein L, Harding S, Herron T, Hong C, Jalife J, Wu JC, Sciences on behalf of the AHAC on BC. Use of Human Induced Pluripotent

Stem Cell–Derived Cardiomyocytes in Preclinical Cancer Drug Cardiotoxicity Testing: A Scientific Statement From the American Heart Association. Circ Res (2019) 125:e75–e92. doi: 10.1161/res.0000000000000291

200. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. Cell (2019) 177:1873-1887.e17. doi: 10.1016/j.cell.2019.05.006