Comprehensive Bioinformatics Meta-Analysis of Coronary Artery Disease and Myocardial Infarction

Rachana Gurudu

Abstract

Coronary Artery Disease (CAD) and Myocardial Infarction (MI) are the leading causes of mortality in the United States. Certain genes have been shown through bioinformatics analysis to be related to CAD and MI, but are not incorporated into USPSTF Guidelines. In this study, we aimed to create a model of diagnosis and prognosis in adult patients with coronary artery disease and myocardial infarction from 10 diverse datasets, and also aimed to identify biological pathways, immune cells, and drugs related to CAD and MI to improve biological understanding of these conditions. Using the R MetaIntegrator Package and multiple datasets including the Gene Expression Omnibus, Library of Integrated Network-Based Cellular Signatures, and Reactome, we analyzed data from 10 datasets and found a high predictive value in predicting patients with either CAD or MI. We also used the gene signatures generated through our original meta-analysis to identify significant biological pathways, which included oncogene induced senescence ($p < 0.05$) and neutrophil degranulation ($p = 0.000275$), significant drugs that could be potential treatments for CAD or MI, which included enzalutamide ($r = -0.5464015$, $p = 3.043318e-08$, FDR = 3.800495e-05) and ibutilide ($r = -0.5209267$, $p = 1.663780e-07$, FDR = 9.198588e-05), and significant immune cells using immune cell deconvolution, which included natural killer cells. We conclude that through these results, we have created a more biologically heterogeneous gene signature and meta-analysis to predict diagnosis and prognosis in patients with CAD and MI, and found new biological pathways, drugs, and immune cells that can be used to improve understanding of the conditions and treat them.

Introduction

Annually, more than 3 million people in the United States alone are diagnosed with Coronary Artery Disease (CAD), the leading cause of mortality in the United States[1]. Acute myocardial infarction (MI) is the leading cause of mortality in developed countries, reaching 1 million deaths annually in the United States[2]. Certain genes have been identified as predictors of coronary artery disease in patients through gene expression analysis in recent studies[3,4]. However, these genes have not been incorporated into screening measures via USPSTF guidelines, and are not consistently used. Finding consistent and replicable genetic markers is critical to advancing treatment and biological understanding of CAD that is more patient-specific and comprehensive in order to eventually reduce mortality of CAD and MI.

As the leading cause of mortality, CAD possesses a large burden on patients and families. CAD is caused by atherosclerosis, or arterial plaque buildup, and can eventually cause cardiac ischemia and myocardial infarction by blocking off blood flow to a region of the heart entirely[5]. CAD is caused by multiple factors, including hyperlipidemia, diabetes, family history, tobacco use, and obesity, and is considered the leading cause of heart attacks. Current diagnostic measures include electrocardiograph tests, exercise stress tests, echocardiograms, and blood tests that measure levels of cholesterol and other indicators of atherosclerosis. During MI, which is most commonly caused by CAD, symptoms of angina and shortness of breath are typically assessed through EKG and cardiac enzyme levels, such as troponin and

myoglobin, which are elevated in the presence of cardiac damage. From there, MI is treated with anticoagulants and coronary intervention or surgery to relieve the blockage and symptoms[6].

Many studies have been done on specific patient populations to examine gene expression data in relation to CAD and MI. Indeed, genome wide association studies have identified single nucleotide polymorphisms related to CAD[7], and inflammatory and RNA transport genes have been identified as significant to MI[8]. However, most of these studies lack integrated analysis of different populations and a predictive model which is necessary to improve biological understanding of CAD and MI and allow for faster diagnosis and prognosis determination. As a result, machine learning models are often skewed in favor of the populations they trained on, and are not accurate for all populations and demographics. Meta-analysis allows for this integration of different populations through varied datasets, offering more specific and comprehensive results regarding gene expression in relation to CAD and MI. In this study, we will use meta-analysis and machine learning algorithms leveraging gene expression data from 10 datasets from different technologies, sources, and ethnicities to create a model of diagnosis and prognosis in adult patients with coronary artery disease and myocardial infarction. We aim to predict diagnosis and prognosis in adults with coronary artery disease and myocardial infarction using this machine learning model at a higher accuracy than previous models. Furthermore, we aim to identify biological pathways, immune cells, and drugs related to CAD and MI to improve biological understanding of these conditions.

Materials and Methods

Gene expression data sets comparing CAD and MI patients to healthy patients with complete gene signatures were accessed through the NCBI GEO Database[9]. The GEO Accession numbers are GSE141512[10], GSE98583[11], GSE34822[12], GSE42148[13], GSE12288[14], GSE90074[4], GSE29111[15], GSE123342[16,17], GSE97320[18], GSE61144[19], GSE62646[20], and GSE34198[21] (Table 1). In total, these 12 datasets contained 968 peripheral blood samples, which were either control samples (N = 207) depending on the comparison or case samples (N = 761) that either represented CAD or MI. In some datasets, samples were further distinguished further based on timepoints.

All datasets were downloaded in R and processed by labeling samples as case/control and filtering data based on time points (if relevant) using the MetaIntegrator package from CRAN[22,23]. Next, we used statistical meta-analysis techniques[24] to determine what genes were statistically different between cases and controls across datasets, with the MetaIntegrator package to create ROC plots, gene heatmaps, and violin plots to compare genes across datasets and groups of similar datasets to larger, validation datasets (GSE12288 and GSE34198).

$$g = J\frac{\overline{X}_1 - \overline{X}_0}{\sqrt{\frac{(n_1-1)S_1^2+(n_0-1)S_0^2}{n_1+n_0-2}}}$$

24

Specifically, the R package MetaIntegrator [25-28] was used to estimate gene effect size of all genes within the comparison dataset CAD v. MI or CAD v. Healthy, as shown in Table 1, as Hedge's g effect size, where X_1 and X_2 are the average case and control gene expressions, S is the pooled standard deviation for case and control samples, n is the number of cases or controls, and J is the correction factor, which is 1 - 3/(4df - 1) where df = degrees of freedom. An effect size was estimated for each gene within each individual dataset and then were pooled into a comparison-specific (CADvMI or CADvHealthy) effect size by performing an inverse weighting of the variance of each individual dataset's effect size. After dataset-wide effect sizes were determined, we corrected for multiple comparisons by adjusting all p-values using the Benjamini-Hochberg false discovery rate (FDR) correction[29]. We also used the forewardSearch algorithm in the R MetaIntegrator Package to further filter our genes to only those with the highest discriminatory power[30]. For the CADvHealthy Dataset, an effect size threshold of 0.3 and an FDR threshold of 0.3 filtered down the original 19,730 genes from the meta-analysis and 18 genes in total were identified as differentially expressed (15 upregulated, 3 downregulated). For the CADvMI Dataset, an effect size threshold of 0.8 and an FDR threshold of 0.1 filtered down the 26,146 original genes from the meta-analysis and 44 genes in total were identified as differentially expressed (35 upregulated, 9 downregulated).

We also performed an analysis to identify potential genes for drug-repurposing using the Library of Integrated Network-Based Cellular Signatures (LINCS) database to identify significant drugs that may have a significantly anti-correlated relationship with CAD or MI gene signatures. We used Level 5 differential gene expression data from the LINCS database to compare to our gene signatures from the CADvMI and CADvHealthy datasets and found genes that were significantly anti-correlated to the gene signatures previously identified. P-values were multiple-hypothesis adjusted using FDR[31].

We also performed immune cell deconvolution to attempt to identify cell types that are differentially expressed between our cases and controls. We used the same datasets from the gene analysis, including the same grouping for the comparison of CAD v. Healthy samples and CAD v. MI samples. First, we deconvolved the bulk gene-expression data on a dataset-level using immunoStates - turning our bulk gene expression data into differential cell expression data[32]. Following, we ran a meta-analysis on the deconvolved data to determine differentially expressed cell types across all datasets in a comparison. Following the meta-analysis, we filtered down the results to select differentially expressed cell types. For the CADvHealthyImmuno dataset, an effect size threshold of 0.1 and an FDR threshold of 0.8 generated 2 downregulated immune cells, and for the CADvMIImmuno dataset, an effect size threshold of 0.4 and an FDR threshold of 0.2 identified 1 significant immune cell.

Lastly, we performed pathway analysis on the CADvMI dataset and the CADvHealthy dataset using the Reactome database's gene sets in order to determine significant biological pathways that have a relationship with MI and CAD[33-36]. Fisher's exact test was used statistically after using clusterProfiler to find similarities between genes and pathways. We used a larger gene signature to determine which genes are differentially expressed using the CADvMI dataset, with an FDR threshold of 0.1 and an effect size threshold of 0.8, and the CADvHealthy dataset, with an FDR threshold of 0.3 and an effect size threshold of 0.3. Then, to explore the effects of these genes, we performed pathway analysis on the CADvMI dataset, with a p-value cutoff of 0.2 and a minimum g size of 10, and the CADvHealthy dataset, with a p-value cutoff of 0.2 and a minimum g size of 10 as well. We adjusted these p-values using the Benjamini-Hochberg false discovery rate (FDR) correction[29].

Table 1.

Accession	Platform	Tissue	Demographic	Control Samples	Case Samples	Total # of Samples	Location of Sample	Comparison Used
GSE141512 (Discovery)	GPL17586	PBMCs (Peripheral Blood Mono-nuclear cells)	Male patients with MI, with normal ECG and no history of cardiovascular diseases (CVD) and diabetes mellitus	6	6	12	European Russia	CAD v. MI
GSE98583 (Discovery)	GPL571	Blood	Non-diabetic male patients with stable CAD	6	12	18	North India	CAD v. Healthy
GSE34822 (Discovery)	GPL6480	Peripheral Blood	Patients with good risk factor control and progressive CAD	0	32(16 with continued inter-vention, 16 stable)	32	Germany	CAD v. Healthy
GSE42148 (Discovery)	GPL13607	Peripheral Blood	Patients with angiographically confirmed coronary artery disease (CAD) between ages 40 - 55 years	11	13(7 SA, 6 MI)	24	India	CAD v. Healthy
GSE12288 (Validation)	GPL96	Peripheral Blood	American patients with CAD undergoing coronary angiography(Different races/sexes)	112	110	222	USA/Switzerland	CAD v. Healthy(Validation)
GSE90074 (Discovery)	GPL6480 and GPL6801	Peripheral Blood	Individuals at least 65 years old and undergoing	0	249	249	USA (North Carolina)	CAD v. Healthy

			clinically indicated cardiac catheterization					
GSE123342 (Discovery)	GPL17586	Peripheral Blood	Individuals with either CAD diagnosis or MI	0	192(65 acute MI, 64 30 days post MI, 37 1 year post MI, 4 replicates	192	Belgium	CAD v. MI
GSE61144 (Discovery)	GPL6106	Peripheral Blood	Patients with ACS who visited emergency department within 4 hours after the onset of chest pain	10	14 (before and 7 days after primary intervention	24	South Korea	CAD v. MI
GSE62646 (Discovery)	GPL6244	PBMCs	28 patients with STEMI - blood taken on 1st day of MI, after 4-6 days, and 6 months	14(with stable CAD)	84(28 patients at each of the 3 timepoints)	98	Poland	CAD v. MI
GSE34198 (Validation)	GPL6102	Peripheral Blood	90 patients in 3 different disease groups: controls, immediately after MI, 6 months post MI	48	49	97	Czech Republic(Prague)	CAD v. MI(Validation)

Results

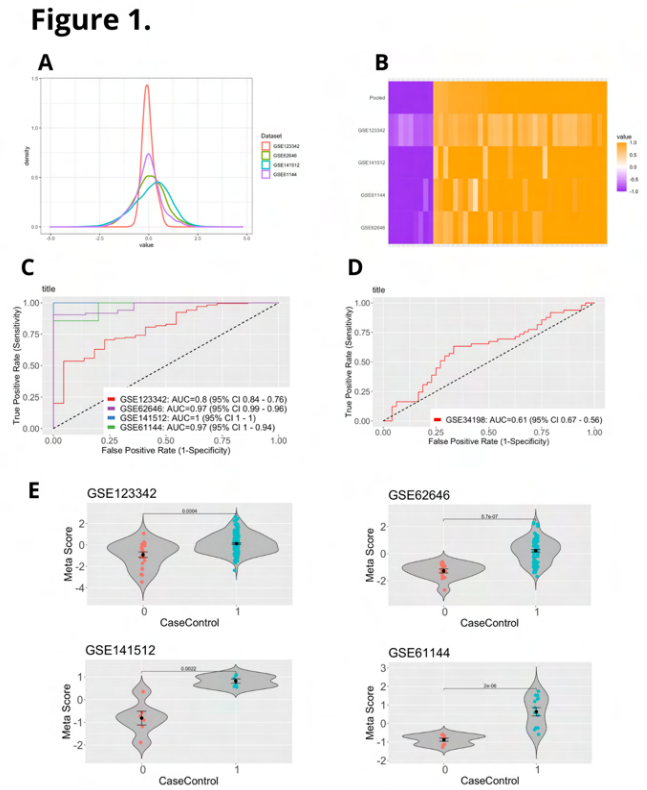

Figure 1: CADvMI dataset gene signature plots. (A) Dataset-wide distribution of effect sizes for all datasets in meta-analysis. (B) Heatmap of the selected gene signature for all 4 CADvMI datasets. (C) Summary ROC plot of gene signature of discovery datasets. (D) Summary ROC plot of validation dataset for CADvMI (GSE34198). (E) Individual violin plots for each CADvMI Dataset comparing gene signatures of case and control groups for discovery datasets.

We combined multiple independent gene datasets from the NCBI GEO database to allow for clinical and technical heterogeneity in our sample of patients, unlike existing literature on gene signatures of patients with CAD or MI that are typically isolated to one cohort. Using the GEO database, we used 10 datasets, from 9 different countries, that were divided up into two different datasets: CADvMI (GSE123342, GSE62646, GSE141512, GSE61144), comparing CAD patient gene signatures to MI patient gene signatures with a validation dataset (GSE34198) to test our model, and CADvHealthy (GSE42148, GSE90074, GSE98583, GSE34822), comparing CAD patient gene signatures to healthy patient gene signatures with another validation dataset (GSE12288) to test our model. The effect size threshold used on the meta-analysis of the CADvMI dataset was 0.8, the FDR threshold was 0.1, and the number of studies threshold was 4 (Figure 1). Importantly, the CADvMI analysis predicted classes well, with an AUC for GSE123342 of 0.8, an AUC for GSE62646 of 0.97, an AUC for GSE141512 of 1, and an AUC for GSE61144 of 0.97. The lower AUC for GSE123342 can likely be attributed to the fact that this dataset contained samples at multiple different timepoints post-MI, making CAD harder to differentiate. The model from the CADvMI dataset was tested using a larger validation dataset with similar data comparing patients. The CADvMI model, however, did not predict CAD or MI outcomes in patients from the validation dataset GSE34198 as well as the discovery datasets above, with an AUC for GSE34198 of 0.61. This is likely due to the fact that the control patients selected in GSE34198 did not have CAD and were healthy, as opposed to the patients in the discovery dataset that had CAD or MI.

We also analyzed the CADvMI dataset using biological pathway analysis from the Reactome database (Figure 3). Pathway analysis for the CADvMI dataset indicated a significant relationship ($p = 0.000275$) between CAD and neutrophil degranulation (gene ratio =~ 0.29). Neutrophils were identified during the immune cell deconvolution results below as upregulated during MI, which is likely because white blood cells increase following cardiac tissue damage

during an MI for repair. The neutrophil degranulation pathway causes the release of granules from neutrophils through exocytosis, and, when excessive, can cause inflammatory disorders such as rheumatoid arthritis and asthma[37]. Prior literature indicates a relationship between neutrophil degranulation and myocardial infarction due to the damage in cardiac tissue during MI that prompts inflammatory responses from white blood cells[38], which explains the low p-value and indicated significance between our gene signature and the neutrophil degranulation pathway.

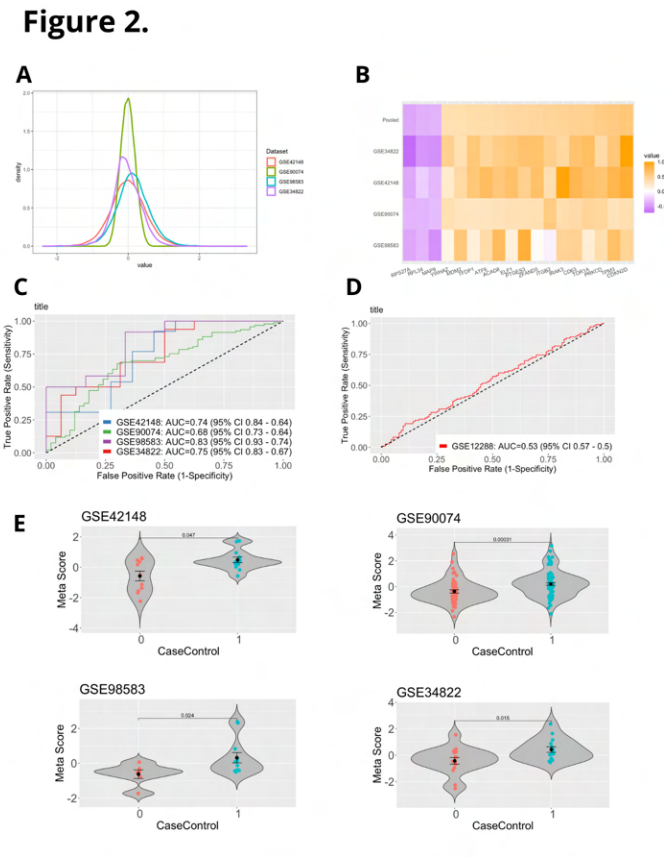

Figure 2: CADvHealthy dataset gene signature plots. (A) Dataset-wide distribution of effect sizes for all datasets in meta-analysis. (B) Heatmap of the selected gene signature for all 4 CADvHealthy datasets. (C) Summary ROC plot of gene signature of discovery datasets. (D) Summary ROC plot of validation dataset for CADvHealthy (GSE12288). (E) Individual violin plots for each CADvHealthy Dataset comparing gene signatures of case and control groups.

The CADvHealthy dataset mentioned above was less effective in differentiating CAD patients from healthy controls than the CADvMI dataset was, both with the discovery and validation datasets. The effect size threshold set for the CADvHealthy dataset was 0.3, the FDR threshold was also 0.3, and genes were limited to only those that appeared in all datasets. The CADvHealthy dataset produced an AUC for GSE42148 of 0.74, an AUC for GSE90074 of 0.68, an AUC for GSE98583 of 0.83, and an AUC for GSE34822 of 0.75 (Figure 2). Although these AUCs are higher than 0.5, indicating some differentiation of the model, they are not significantly high, likely due to the difficulty in distinguishing CAD from healthy patient populations in terms of gene signatures. Indeed, some datasets, such as GSE90074, used patients with a CAD severity index greater than 50%, calculated as the percentage of stenosis determined through coronary angiography[4], while other datasets, such as GSE98583, only looked at gene expression in patients with stable CAD, resulting in poor predictability of the model. The CADvHealthy model was tested on the validation dataset GSE12288 and an AUC of 0.53 was produced, likely attributable to the fact that patients in the GSE12288 were divided based on the coronary artery disease index, as opposed to angiographically confirmed diagnosis, resulting in data that is not comparable.

Pathway analysis for the CADvHealthy dataset indicated a significant relationship ($p < 0.05$) between CAD and oncogene induced senescence (gene ratio < 0.17). Interestingly,

oncogene induced senescence is involved in the process of tumor suppression and has been shown to reduce benign tumor growth and protect against malignancy[39]. Limited literature exists on the relationship between oncogene induced senescence and coronary artery disease, but cellular senescence has been shown to accelerate atherosclerosis, which causes CAD, and worsen outcomes after myocardial infarction by limiting tissue repair[40]. Future studies investigating oncogene induced senescence and CAD may provide more information about a reversed relationship due to the effects of this biological pathway on tumor suppression and control.

Figure 3.

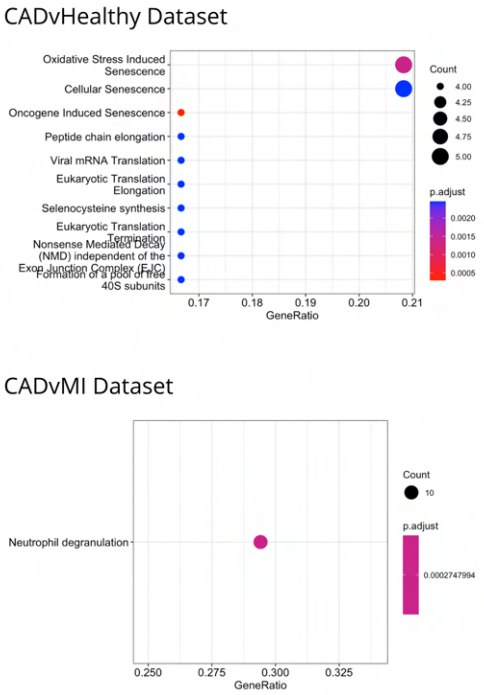

Figure 3: Dotplot enrichment maps for both the CADvHealthy and CADvMI gene signatures we performed meta-analysis on earlier.

Immune Cell Deconvolution Results

A different way we analyzed the gene signatures of CAD, MI, and healthy patients was creating a predictive model through immune cell deconvolution using immunoStates[32]. Using ImmunoStates, we deconvolved the bulk level gene expression using previously determined cell-type specific genes, to calculate an estimate of the levels of different cell types in the samples. Using the levels of these cell types, we performed a meta-analysis to find consistently upregulated or downregulated cell types across our datasets. We did a CAD v. Healthy meta-analysis, using an effect size threshold of 0.1, a FDR threshold of 0.8 and a number of studies threshold of 2 to filter the immune cells. The CADvHealthy immune cell signature produced an AUC for GSE42148 of 0.58, an AUC for GSE90074 of 0.5, an AUC for GSE98583 of 0.43, and an AUC for GSE34822 of 0.41. These AUC values produce an average AUC of 0.48, indicating that the model predicts at random (AUC = 0.5). This model was also tested on the same validation dataset as the gene set analysis, GSE12288, producing an AUC of 0.5. The CADvHealthy dataset, when tested with the immune cell deconvolution approach, had no strong effect, indicating that the gene signature is not correlated to upregulated or downregulated immune cells and that immune cell levels are not predictive of CAD patients when compared to healthy patients.

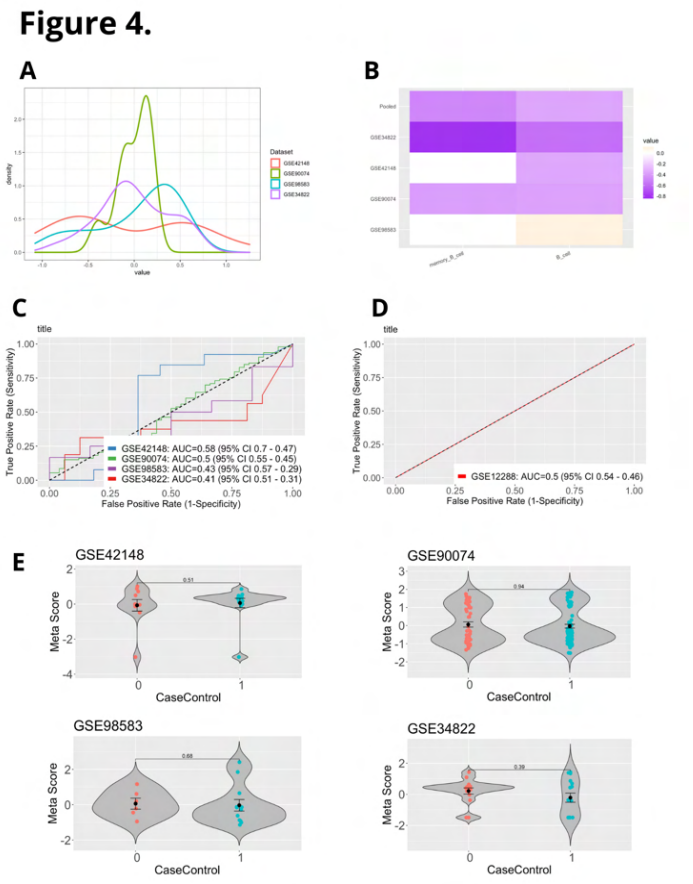

Figure 4: Performance plots for the CADvHealthy Immune Cell Deconvolution experiment. (A) Effect-size distribution for the meta-analysis of the immune cell signature plot of the 4 CADvHealthy Immune Cell datasets from GEO. (B) Heatmap of the cell type signature for all 4 CADvHealthy Immune datasets. (C) Summary ROC plot of immune cell signature. (D) Validation ROC plot of validation dataset for CADvHealthy Immune (GSE12288). (E) Individual violin plots for each CADvHealthy Dataset comparing immune cell deconvolution signatures of case and control groups.

The immune cells produced using this immune cell analysis for the CADvHealthy gene signature were memory B cells and B cells, both of which were downregulated. Little research has been done to examine the relationship between B cells and atherosclerosis and CAD, but B cells have been shown to result in heart failure and impaired function by increasing the progression of the disease[41].

We also performed immune cell deconvolution on the CADvMI dataset from the gene analysis above, filtering immune cell types with an effect size threshold of 0.4, an FDR threshold of 0.2, and a number of studies threshold of 2. The CADvMI immune cell signature produced an AUC for GSE123342 of 0.49, an AUC for GSE62646 of 0.93, an AUC for GSE141512 of 0.86, and an AUC for GSE61144 of 0.89. These AUC values indicate that the CADvMI model is more predictive than the CADvHealthy model for significant immune cells, and are significantly more predictive than the model at random. The validation dataset GSE34198 produced an AUC of 0.56, indicating that this model fails to generalize to our left-out validation set, however, likely attributable to the different timepoints after MI that samples were collected in patients. The immune cells produced for the CADvMI gene signature were natural killer cells, which were downregulated. Although limited research exists on the role of natural killer cells in CAD or MI, natural killer cells have been shown to have a protective effect against atherosclerosis[42], or the plaque buildup that causes CAD, which may explain why natural killer cells appear to be downregulated in CAD and MI patient samples in these datasets.

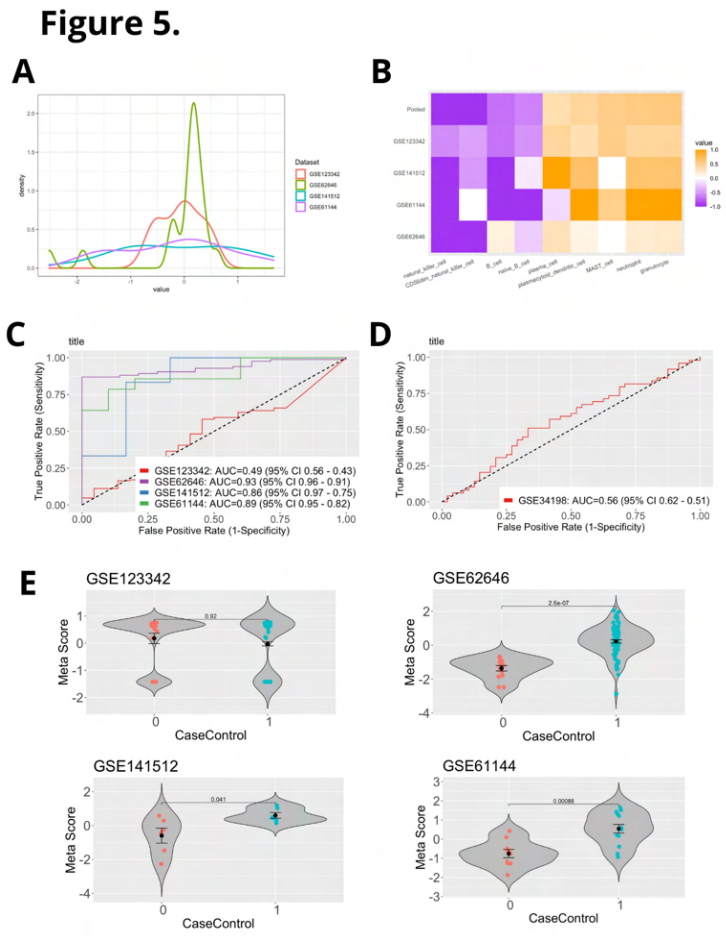

Figure 5: Performance plots for the CADvMI Immune Cell Deconvolution experiment. (A) Effect-size distribution for the meta-analysis of the immune cell signature plot of the 4 CADvMI Immune Cell datasets from GEO. Meta-analysis of immune cell signature plot of 4 CADvMI Immune Cell datasets. (B) Heatmap of the cell type signature for all 4 CADvMI Immune datasets. (C) Summary ROC plot of immune cell signature for discovery datasets. (D) Summary ROC plot of validation dataset for CADvMI Immune (GSE34198). (E) Individual violin plots for each CADvMI Dataset comparing immune cell deconvolution signatures of case and control groups.

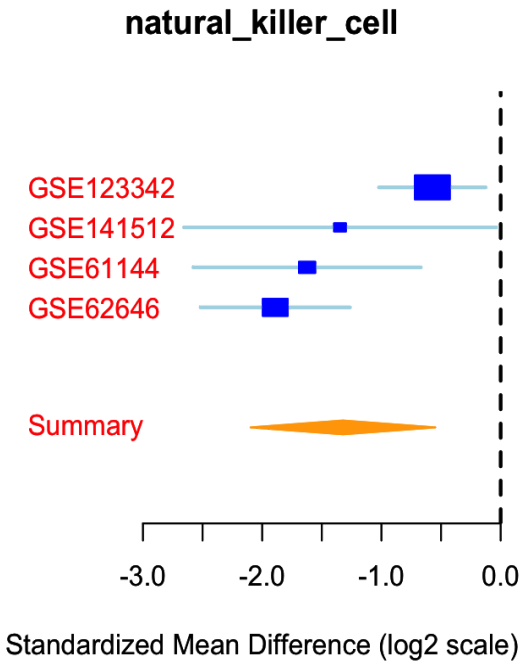

Figure 6: Forest plot for natural killer cells identified in immune cell deconvolution for the CADvMI gene signature.

Drug Results

We also examined the relationship between drug transcriptome profiles using the LINCS database and both CAD and MI to find possible drug treatments for CAD and MI that could aid in prevention or treatment of the disease, by reversing the gene signatures found above. The most common medications given for coronary artery disease currently include anticoagulants such as aspirin or clopidogrel, heparins, and beta-blockers[43]. Out of the 12,486 molecules statistically compared to both the CADvHealthy and CADvMI signatures, we

limited significant results to FDA approved drugs only as these medications are already approved for other uses and are considered safe for humans[44].

Out of the drugs compared in the LINCS database, we aimed to find drugs that would reverse the CADvHealthy gene signature by being anti-correlated with the gene signature. The two strongest correlations between CAD and the drug's usage were nifedipine (r = -0.5107063, p = 1.558371e-10, FDR = 1.946094e-06) and myriocin (r = -0.3375490, p = 5.149852e-05, FDR = 7.174960e-03). Nifedipine's strong correlation to the gene signature comparing CAD to healthy peripheral blood samples is understandable because nifedipine is a calcium channel blocker and anti-hypertensive drug that dilates the coronary artery, preventing severe CAD and MI outcomes in patients despite not being recommended for CAD by the FDA currently[45]. Myriocin, on the other hand, is a fungal-derived FDA approved drug for inflammatory conditions such as multiple sclerosis, but is not currently recommended for coronary artery disease or cardiac conditions[46]. There is existing literature, however, that in rats myriocin has been shown to have a protective effect against atherosclerosis, which primarily causes CAD, by reducing glycosphingolipids[47]. Another drug identified as significant by the LINCS database was an unspecified spleen tyrosine kinase (SYK) inhibitor (r = -0.3534175, p = 2.118706e-05, FDR = 5.187923e-03). The only FDA approved SYK inhibitor is fostamatinib for chronic immune thrombocytopenia, but this drug has shown adverse cardiovascular side effects such as hypertension, CAD, and MI in some patients[48,49]. Future studies are needed to examine the relationship between SYK inhibitors and cardiovascular disease, including other SYK inhibitors besides fostamatinib to determine if these side effects are isolated to this drug only.

We also aimed to find drugs in the LINCS database that would reverse the CADvMI gene signature by being anti-correlated with the gene signature. The two strongest correlations between the CADvMI gene signature and the drug's usage were enzalutamide (r = -0.5464015, p = 3.043318e-08, FDR = 3.800495e-05) and ibutilide (r = -0.5209267, p = 1.663780e-07, FDR = 9.198588e-05). Enzalutamide is an androgen-receptor inhibitor used in patients with metastatic prostate cancer to improve prognosis, but has been shown to cause an increase in cardiovascular morbidity in some patients. However, enzalutamide has shown to be a safer alternative over other standard hormone therapy treatments for metastatic prostate cancer, which may explain the high inverse correlation[50,51]. Ibutilide has a closer connection with cardiovascular disease as its primary indication is to reverse atrial fibrillation to normal sinus rhythm and prevent future atrial fibrillation[52]. Ibutilide is only indicated in patients with atrial fibrillation and thus little literature exists on its ability to prevent or improve the prognosis of CAD or MI, but prior literature has shown that ibutilide is more effective on patients with previous CAD or cardiovascular disease who present with atrial fibrillation in comparison to other patients[53].

Discussion

Specific literature exists examining gene expression profiles in relation to CAD and MI, but significant genes identified in prior literature have not been incorporated into USPSTF screening guidelines and these significant genes have not been tested across international datasets. In this study, we aimed to perform meta-analysis to predict prognosis and diagnosis of CAD and MI in patients using 10 datasets from 9 different countries that we performed meta-analysis and machine learning algorithms on to improve biological understanding through increasing heterogeneity of our gene signature and predict CAD and MI at a higher accuracy.

We divided our meta-analysis into CAD v. Healthy patient datasets and CAD v. MI patient datasets and used forward search algorithms to identify significant genes. The CAD v. MI gene

signature was highly predictive on our discovery datasets, with all AUCs greater than 0.8, and a validation AUC of 0.61, likely lower due to the healthier patients in the validation dataset in comparison to the discovery dataset, where patients either had CAD or MI. The CAD v. Healthy gene signature was less predictive than the CAD v. MI gene signature on our discovery datasets, with all AUCs greater than 0.69 and a validation AUC of 0.53, likely due to the different classification of CAD severity between datasets.

We then performed biological pathway analysis on both gene signatures and identified oncogene induced senescence as a statistically significant pathway in relation to CAD v. Healthy patients ($p < 0.05$, gene ratio < 0.17). Although cellular senescence has been shown in literature to accelerate atherosclerosis, which causes CAD, future literature looking at the relationship between CAD and the tumor-suppressing pathway of oncogene-induced senescence is key. Biological pathway analysis on the CAD v. MI gene signature identified neutrophil degranulation as significant ($p = 0.000275$, gene ratio = ~0.29), likely due to the role neutrophils play in tissue repair following cardiac damage from an MI.

We also performed an immune cell deconvolution algorithm on both the CADvHealthy and CADvMI gene signatures to identify significant immune cells and their role in CAD or MI. We used the immunoStates algorithm to create immune objects based on our original gene signatures, and identified the downregulation of B cells in the CADvHealthy gene signature, which have been shown in prior literature to accelerate cardiovascular damage, the upregulation of granulocytes and neutrophils in the CADvMI gene signature, which aid in cardiac damage following MI, and the downregulation of natural killer cells in the CADvMI gene signature, which have been shown in prior literature to have a protective effect against atherosclerosis and thus cause a reduction in CAD and MI.

Lastly, we used the LINCS database to identify drugs that have a significant inverse correlation with CAD and MI that could be repurposed for treatment of prevention in high risk patients. In the CADvHealthy gene signature, nifedipine ($r = -0.5107063$, $p = 1.558371e-10$, FDR = 1.946094e-06) and myriocin ($r = -0.3375490$, $p = 5.149852e-05$, FDR = 7.174960e-03) were identified as significant. Nifedipine is an anti-hypertensive drug currently not recommended for CAD but likely had an inverse correlation due to its protective effect against atherosclerosis, while Myriocin is an anti-inflammatory drug that has actually shown a reduction in atherosclerosis in rats through a reduction in glycosphingolipids. In the CADvMI gene signature, enzalutamide ($r = -0.5464015$, $p = 3.043318e-08$, FDR = 3.800495e-05) and Ibutilide ($r = -0.5209267$, $p = 1.663780e-07$, FDR = 9.198588e-05). Enzalutamide is a medication indicated for metastatic prostate cancer that has been shown to have adverse cardiovascular morbidity, while Ibutilide is a medication indicated for atrial fibrillation that prior literature has revealed is stronger on patients with a prior history of CAD or MI.

Creating this machine learning model predictive of CAD and MI in patients allowed for increased biological heterogeneity in our gene signature compared to other models, and also allowed for the identification or more accurate drugs that can be repurposed for treatment or prevention and immune cells that have a significant relationship with CAD and MI.

However, this study has some limitations. First, we used 10 datasets but using more datasets from different countries, including low income and rural hospitals, would allow for a more heterogeneous gene signature and a more representative model. Additionally, creating a standard for collection, based on specific timepoints following MI and standards for what control patients are (healthy or stable CAD) would allow for a stronger model that would be more predictive on validation datasets. Second, most data regarding MI was taken after patients

present to the hospital with symptoms. In an ideal world, we would have chronological data on gene signatures from CAD diagnosis to MI to determine where gene expression changes and what genes are uniquely involved in accelerating MI. Third, our samples from the 10 datasets came from peripheral blood samples and in the 5 CAD v. MI datasets, the MI patients had already been in the hospital for some time and likely received intervention. Thus, we cannot entirely attribute the gene signature in these patients to MI, as drugs they received may have also altered gene expression and immune cell expression. However, since we had a consistent gene signature and a highly predictive model across datasets in countries with different standards of practice and patient characteristics, it is likely that the genes filtered were a result of changes in expression from MI, not drugs given. Fourth, limited data exists on the relationship between drugs such as ibutilide and myriocin, as well as biological pathways such as oncogene induced senescence, and CAD, and future studies are necessary to examine these relationships specifically in patients with existing CAD rather than just measuring cardiovascular morbidity alongside the condition the drug is indicated for.

References

1. Brown JC, Gerhardt TE, Kwon E. Risk Factors For Coronary Artery Disease. In: StatPearls. Treasure Island (FL): StatPearls Publishing; June 5, 2021.
2. Mechanic OJ, Gavin M, Grossman SA. Acute Myocardial Infarction. In: StatPearls. StatPearls Publishing; 2022. Accessed September 22, 2022. http://www.ncbi.nlm.nih.gov/books/NBK459269/
3. Qi B, Chen JH, Tao L, et al. Integrated Weighted Gene Co-expression Network Analysis Identified That TLR2 and CD40 Are Related to Coronary Artery Disease. Front Genet. 2021;11:613744. Published 2021 Jan 26. doi:10.3389/fgene.2020.613744
4. Ravi S, Schuck RN, Hilliard E, et al. Clinical Evidence Supports a Protective Role for CXCL5 in Coronary Artery Disease. Am J Pathol. 2017;187(12):2895-2911. doi:10.1016/j.ajpath.2017.08.006
5. Coronary artery disease: Causes, symptoms, diagnosis & treatments. Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/16898-coronary-artery-disease. Accessed August 11, 2022.
6. Sweis RN, Jivan A. Acute myocardial infarction (MI) - cardiovascular disorders. Merck Manuals Professional Edition. https://www.merckmanuals.com/professional/cardiovascular-disorders/coronary-artery-disease/acute-myocardial-infarction-mi. Published June 2022. Accessed August 11, 2022.
7. Hsu J, Smith JD. Genome-wide studies of gene expression relevant to coronary artery disease. Curr Opin Cardiol. 2012 May;27(3):210-3. doi: 10.1097/HCO.0b013e3283522198. PMID: 22476029; PMCID: PMC3332306.
8. Kontou P, Pavlopoulou A, Braliou G, et al. Identification of gene expression profiles in myocardial infarction: a systematic review and meta-analysis. BMC Medical Genomics. 2018;11(1):109. doi:10.1186/s12920-018-0427-x
9. Geo accession viewer. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi. Published May 5, 2017. Accessed August 2, 2022.
10. Osmak G, Baulina N, Koshkin P, Favorova O. Collapsing the list of myocardial infarction-related differentially expressed genes into a diagnostic signature. J Transl Med. 2020;18(1):231. doi:10.1186/s12967-020-02400-1

11. GEO Accession viewer. Accessed September 4, 2022. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE98583
12. Nührenberg TG, Langwieser N, Binder H, et al. Transcriptome analysis in patients with progressive coronary artery disease: identification of differential gene expression in peripheral blood. J Cardiovasc Transl Res. 2013;6(1):81-93. doi:10.1007/s12265-012-9420-5
13. GEO Accession viewer. Accessed September 4, 2022. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42148
14. Sinnaeve PR, Donahue MP, Grass P, et al. Gene expression patterns in peripheral blood correlate with the extent of coronary artery disease. PLoS One. 2009;4(9):e7037. doi:10.1371/journal.pone.0007037
15. GEO Accession viewer. Accessed September 4, 2022. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE29111
16. Vanhaverbeke M, Vausort M, Veltman D, et al. Peripheral Blood RNA Levels of QSOX1 and PLBD1 Are New Independent Predictors of Left Ventricular Dysfunction After Acute Myocardial Infarction. Circ Genom Precis Med. 2019;12(12):e002656. doi:10.1161/CIRCGEN.119.002656
17. Veltman D, Wu M, Pokreisz P, et al. Clec4e-Receptor Signaling in Myocardial Repair After Ischemia-Reperfusion Injury. JACC Basic Transl Sci. 2021;6(8):631-646. doi:10.1016/j.jacbts.2021.07.001
18. GEO Accession viewer. Accessed September 4, 2022. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97320
19. Park HJ, Noh JH, Eun JW, et al. Assessment and diagnostic relevance of novel serum biomarkers for early decision of ST-elevation myocardial infarction. Oncotarget. 2015;6(15):12970-12983. doi:10.18632/oncotarget.4001
20. Kiliszek M, Burzynska B, Michalak M, et al. Altered gene expression pattern in peripheral blood mononuclear cells in patients with acute myocardial infarction. PLoS One. 2012;7(11):e50054. doi:10.1371/journal.pone.0050054
21. Valenta Z, Mazura I, Kolár M, et al. Determinants of Excess Genetic Risk of Acute Myocardial Infarction - A Matched Case-control Study. European Journal of Biomedical Informatics. 2012;8(1). doi:10.24105/ejbi.2012.08.1.6
22. D. Venet, F. Pecasse, C. Maenhaut, H. Bersini, Separation of samples into their constituents using gene expression data , Bioinformatics, Volume 17, Issue suppl_1, June 2001, Pages S279–S287, https://doi.org/10.1093/bioinformatics/17.suppl_1.S279
23. Haynes WA, Vallania F, Tomczak A, et al. MetaIntegrator: Meta-Analysis of Gene Expression Data. Published online February 26, 2020. Accessed August 13, 2022. https://CRAN.R-project.org/package=MetaIntegrator
24. Haynes WA, Vallania F, Liu C, et al. Empowering Multi-Cohort Gene Expression Analysis to Increase Reproducibility. Pac Symp Biocomput. 2016;22:144-153.
25. Khatri P, Roedder S, Kimura N, et al. A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. J Exp Med. 2013;210(11):2205-2221. doi:10.1084/jem.20122709
26. Andres-Terre M, McGuire HM, Pouliot Y, et al. Integrated, Multi-cohort Analysis Identifies Conserved Transcriptional Signatures across Multiple Respiratory Viruses. Immunity. 2015;43(6):1199-1211. doi:10.1016/j.immuni.2015.11.003

27. Sweeney TE, Braviak L, Tato CM, Khatri P. Genome-wide expression for diagnosis of pulmonary tuberculosis: a multicohort analysis. Lancet Respir Med. 2016;4(3):213-224. doi:10.1016/S2213-2600(16)00048-5
28. Sweeney TE, Wong HR, Khatri P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. Sci Transl Med. 2016;8(346):346ra91. doi:10.1126/scitranslmed.aaf7165
29. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
30. forwardSearch: Forward Search Function in MetaIntegrator: Meta-Analysis of Gene Expression Data. Accessed September 11, 2022. https://rdrr.io/cran/MetaIntegrator/man/forwardSearch.html
31. Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171(6):1437-1452.e17. doi:10.1016/j.cell.2017.10.049
32. Vallania F, Tam A, Lofgren S, et al. Leveraging heterogeneity across multiple datasets increases cell-mixture deconvolution accuracy and reduces biological and technical biases. Nat Commun. 2018;9(1):4735. doi:10.1038/s41467-018-07242-6
33. Khatri P, Sirota M, Butte AJ. Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. PLOS Computational Biology. 2012;8(2):e1002375. doi:10.1371/journal.pcbi.1002375
34. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-15550. doi:10.1073/pnas.0506580102
35. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260
36. Fabregat A, Jupe S, Matthews L, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018;46(D1):D649-D655. doi:10.1093/nar/gkx1132
37. Lacy P. Mechanisms of Degranulation in Neutrophils. Allergy Asthma Clin Immunol. 2006;2(3):98-108. doi:10.1186/1710-1492-2-3-98
38. Zhang N, Aiyasiding X, Li W jing, Liao H han, Tang Q zhu. Neutrophil degranulation and myocardial infarction. Cell Communication and Signaling. 2022;20(1):50. doi:10.1186/s12964-022-00824-4
39. Chandeck C, Mooi WJ. Oncogene-induced cellular senescence. Adv Anat Pathol. 2010;17(1):42-48. doi:10.1097/PAP.0b013e3181c66f4e
40. Yan C, Xu Z, Huang W. Cellular Senescence Affects Cardiac Regeneration and Repair in Ischemic Heart Disease. Aging Dis. 2021 Apr 1;12(2):552-569. doi: 10.14336/AD.2020.0811. PMID: 33815882; PMCID: PMC7990367.
41. García-Rivas G, Castillo EC, Gonzalez-Gil AM, Maravillas-Montero JL, Brunck M, Torres-Quintanilla A, Elizondo-Montemayor L, Torre-Amione G. The role of B cells in heart failure and implications for future immunomodulatory treatment strategies. ESC Heart Fail. 2020 Aug;7(4):1387-1399. doi: 10.1002/ehf2.12744. Epub 2020 Jun 13. PMID: 32533765; PMCID: PMC7373901.

42. Braun NA, Covarrubias R, Major AS. Natural killer T cells and atherosclerosis: form and function meet pathogenesis. J Innate Immun. 2010;2(4):316-24. doi: 10.1159/000296915. Epub 2010 Mar 17. PMID: 20375560; PMCID: PMC2895753.
43. Sweis RN, Jivan A. Drugs for acute coronary syndromes - cardiovascular disorders. Merck Manuals Professional Edition. https://www.merckmanuals.com/professional/cardiovascular-disorders/coronary-artery-disease/drugs-for-acute-coronary-syndromes. Published June 2022. Accessed August 13, 2022.
44. Bai L, Scott MKD, Steinberg E, et al. Computational drug repositioning of atorvastatin for ulcerative colitis. Journal of the American Medical Informatics Association. 2021;28(11):2325-2335. doi:10.1093/jamia/ocab165
45. Mueller HS, Antman EM, Ferst JA, Muller JE. Nifedipine in the treatment of cardiovascular disease. Pharmacotherapy. 1981;1(2):78-94. doi:10.1002/j.1875-9114.1981.tb03555.x
46. Homans C, Yalcin EB, Tong M, et al. Therapeutic Effects of Myriocin in Experimental Alcohol-Related Neurobehavioral Dysfunction and Frontal Lobe White Matter Biochemical Pathology. JBBS. 2022;12(02):23-42. doi:10.4236/jbbs.2022.122003
47. Glaros EN, Kim WS, Wu BJ, et al. Inhibition of atherosclerosis by the serine palmitoyl transferase inhibitor myriocin is associated with reduced plasma glycosphingolipid concentration. Biochem Pharmacol. 2007;73(9):1340-1346. doi:10.1016/j.bcp.2006.12.023
48. Mullard A. FDA approves first-in-class SYK inhibitor. Nature Reviews Drug Discovery. 2018;17(6):385-385. doi:10.1038/nrd.2018.96
49. Farooq MZ, Lingamaneni P, Farid S, et al. A Systematic Review and Meta-Analysis of Cardio-Vascular Side Effects with Fostamatinib a Spleen Tyrosine Kinase Inhibitor. Blood. 2019;134:3465. doi:10.1182/blood-2019-132151
50. Davis ID, Martin AJ, Stockler MR, et al. Enzalutamide with Standard First-Line Therapy in Metastatic Prostate Cancer. New England Journal of Medicine. 2019;381(2):121-131. doi:10.1056/NEJMoa1903835
51. Kulkarni AA, Rubin N, Tholkes A, et al. Risk for stroke and myocardial infarction with abiraterone versus enzalutamide in metastatic prostate cancer patients. ESMO Open. 2021;6(5):100261. doi:10.1016/j.esmoop.2021.100261
52. Szymanski MW, Cassagnol M. Ibutilide. In: StatPearls. StatPearls Publishing; 2022. Accessed August 13, 2022. http://www.ncbi.nlm.nih.gov/books/NBK526021/
53. Das MK, Cheriparambil K, Bedi A, et al. Cardioversion of atrial fibrillation with ibutilide: When is it most effective? Clin Cardiol. 2006;25(9):411-415. doi:10.1002/clc.4960250904