EBioMedicine xxx (2019) xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

A transcriptomic model to predict increase in fibrous cap thickness in response to high-dose statin treatment: Validation by serial intracoronary OCT imaging

Kipp W. Johnson [a,b], Benjamin S. Glicksberg [c], Khader Shameer [d], Yuliya Vengrenyuk [e], Chayakrit Krittanawong [f], Adam J. Russak [a,f], Samin K. Sharma [e], Jagat N. Narula [e], Joel T. Dudley [a,b], Annapoorna S. Kini [e,*]

[a] Institute for Next Generation Healthcare, Mount Sinai Health System, New York, NY, United States of America
[b] Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America
[c] Bakar Computational Health Sciences Institute, The University of California, San Francisco, San Francisco, CA, United States of America
[d] Advanced Analytics Center, AstraZeneca, Gaithersburg, MD, United States of America
[e] Mount Sinai Heart, Mount Sinai Health System, New York, NY, United States of America
[f] Department of Internal Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, United States of America

ARTICLE INFO

Article history:
Received 15 February 2019
Received in revised form 15 April 2019
Accepted 3 May 2019
Available online xxxx

Keywords:
Statin
Optical coherence tomography
Personalized medicine
Predictive modeling

ABSTRACT

Background: Fibrous cap thickness (FCT), best measured by intravascular optical coherence tomography (OCT), is the most important determinant of plaque rupture in the coronary arteries. Statin treatment increases FCT and thus reduces the likelihood of acute coronary events. However, substantial statin-related FCT increase occurs in only a subset of patients. Currently, there are no methods to predict which patients will benefit. We use transcriptomic data from a clinical trial of rosuvastatin to predict if a patient's FCT will increase in response to statin therapy.
Methods: FCT was measured using OCT in 69 patients at (1) baseline and (2) after 8–10 weeks of 40 mg rosuvastatin. Peripheral blood mononuclear cells were assayed via microarray. We constructed machine learning models with baseline gene expression data to predict change in FCT. Finally, we ascertained the biological functions of the most predictive transcriptomic markers.
Findings: Machine learning models were able to predict FCT responders using baseline gene expression with high fidelity (Classification AUC = 0.969 and 0.972). The first model (elastic net) using 73 genes had an accuracy of 92.8%, sensitivity of 94.1%, and specificity of 91.4%. The second model (KTSP) using 18 genes has an accuracy of 95.7%, sensitivity of 94.3%, and specificity of 97.1%. We found 58 enriched gene ontology terms, including many involved with immune cell function and cholesterol biometabolism.
Interpretation: In this pilot study, transcriptomic models could predict if FCT increased following 8–10 weeks of rosuvastatin. These findings may have significance for therapy selection and could supplement invasive imaging modalities.

1. Introduction

Acute coronary events most commonly result from atherosclerotic plaque rupture with luminal thrombosis and vessel occlusion. The plaques vulnerable to rupture demonstrate several distinct histomorphological signatures: Fibrous cap thickness (FCT) is the most important determinant of plaque vulnerability, followed by measures of the extent of plaque inflammation and necrotic core size [1,2]. FCT is best assessed in vivo by optical coherence tomography (OCT) [3–5]. However, since OCT is an invasive intravascular imaging test, it is not routinely performed. FCT is known to increase in response to statin (HMG-CoA reductase inhibitor) therapy. This increase in FCT is indicative of plaque stabilization. However, observable increases in FCT occur in only a subset of patients following statin therapy, and it would be important to identify these subjects without an invasive imaging modality [6].

Plaque morphology and composition cannot generally be identified by traditional coronary artery angiography, and numerous imaging modalities, such as intravascular ultrasound (IVUS) [7], OCT [8], and near infrared spectroscopy (NIRS) have been employed in clinical practice [9,10]. Of all the imaging strategies, only OCT imaging is capable of

Abbreviations: FCT, fibrous cap thickness; OCT, optical coherence tomography; Hmg-CoA, 3-hydroxy-3-methyl-glutaryl-coenzyme A; KTSP, K top-scoring pairs.

* Corresponding author.
E-mail address: annapoorna.kini@mountsinai.org (A.S. Kini).

https://doi.org/10.1016/j.ebiom.2019.05.007

Research in context

Evidence before this study

Acute myocardial infarction ("heart attack") results from the rupture of atherosclerotic plaques in the coronary arteries, with subsequent aggregation of platelets and formation of a clot that blocks the flow of blood through the artery, ultimately leading to tissue death in the heart.

One of the key determinants of plaque rupture in the coronary arteries is the thickness of the fibrous cap that forms on top of plaque. When fibrous cap thickness (FCT) is larger, the plaque is more stable and less likely to rupture, and thus less likely to cause myocardial infarction. Drugs such as the statins function to increase FCT and thus reduce the risk of cardiovascular events. However, it is known that statins only increase fibrous cap thickness in a subset of patients. Not every patient prescribed a statin will have a beneficial increase in FCT.

Unfortunately, accurate measurement of FCT is difficult in living patients. The best method to measure FCT is called optical coherence tomography, which requires threading an imaging wire into the coronary arteries to directly image the atherosclerotic plaque. Thus, it would be beneficial if a test could be performed that would be able to predict whether a patient's FCT will respond to statins without performing this invasive procedure.

Added value of this study

In this study, we demonstrate that gene expression data from patient's peripheral blood mononuclear cells can be used with high accuracy to demonstrate whether patients' fibrous caps will thicken in response to statin therapy. We do this by using two machine learning algorithms on the gene expression data, comparing their findings to findings obtained through invasive OCT imaging studies done on the same patients. We also demonstrate that clinical variables alone are not sufficient for predicting FCT response, nor do they add further utility to the gene expression based predictive models.

Implications of all the available evidence

This study demonstrates that change in FCT status in response to high-intensity rosuvastatin therapy can be predicted with very high fidelity in our sample of patients. After further study and validation, models such as ours may allow patients to have their FCT response approximated without undergoing invasive imaging. Before this test could be used in the clinic, it would need to successfully be shown of benefit through randomized, controlled clinical trials and the gene expression panel would need to be commercialized for clinical use.

measuring FCT. Noninvasive approaches such as computed tomography angiography [11] are able to identify the extent of necrotic core and expansive remodeling of the lesion sites, and magnetic resonance imaging is able to identify lipid-rich plaques [12]. However, the resolution of the non-invasive imaging modalities is not sufficient for the assessment of FCT. It is therefore important to develop biomarker surrogates of FCT that could distinguish patients who would or would not benefit from statin therapy.

Lipid lowering agents, most prominently statins, increase the FCT of thin-cap fibroatheromas [13–15]. Statins primarily inhibit the rate-limiting enzyme in cholesterol biosynthesis, HMG-CoA reductase, and also demonstrate pleiotropic effects [6,16]. The degree of FCT thickening (and therefore protection) caused by statins is positively correlated with the statin dose [17]. Unfortunately, not all patients respond equivalently to statin therapy, and the poor responders experience greater atheroma progression and higher rates of adverse cardiovascular outcomes [18,19].

Using OCT imaging, we also observed that only a subset of patients increased the FCT with statin therapy [6]. Usually the statin response is measured by changes in serum lipid levels following the initiation of statin therapy. However, even though there is likely some correlation, serum lipid levels do not directly predict the change in plaque FCT [20]. It is therefore important to non-invasively distinguish responders from non-responders [21], so that non-responders may seek other therapeutic options. We present a predictive model for FCT response to statin therapy utilizing imaging, clinical data, and gene expression (transcriptomics) from the previously described patients included in the Yellow II clinical trial [6].

2. Materials and methods

2.1. Ethics statement

Written informed consent was obtained from all patients. This study was approved by the Institutional Review Board of the Icahn School of Medicine at Mount Sinai.

2.2. Study design and procedure

We analyzed patient data from the Yellow II clinical trial which has been previously described [6]. Briefly, 85 patients with stable multivessel coronary artery disease underwent percutaneous coronary intervention (PCI) for an obstructive lesion. Lesions had greater than or equal to 70% stenosis as measured via angiographic visual assessment and fractional flow reserve. In the same procedure, a second "non-culprit" lesion was imaged with OCT imaging, and peripheral blood mononuclear cell (PBMC) gene expression was characterized by microarray. Trial patients were then assigned to 40 mg daily rosuvastatin therapy. Full trial inclusion criteria and details are deposited at clinicaltrials.org under the clinical trial identifier NCT01567826. Notably, 69 patients had previously been exposed to lower-intensity statin therapy, primarily atorvastatin ($n = 29$) and simvastatin ($n = 23$). At eight to ten weeks post-enrollment the patients underwent a staged PCI, the non-culprit lesion was re-imaged with OCT, and PBMC gene expression was again characterized by microarray. We ultimately obtained complete gene expression and OCT imaging data for 69 of these patients.

2.3. Microarray preparation and gene expression data

PBMCs were isolated from fasting blood samples by density gradient centrifugation using Ficollpaque, and mRNA was isolated using the TRIzol protocol. We used the Illumina HT-12 v4 microarray for gene expression profiling. We chose PBMCs because they are readily obtainable and gene expression signatures are robust across tissues [22]. All data preparation was conducted in R version 3.3.3 [23]. Quality control and normalization was accomplished with the lumi Bioconductor package [24,25] with a variance-stabilizing transformation applied to background-adjusted probe intensities [26]. Normalization was performed with the quantile method. We corrected for batch effects resulting from staged microarray preparation using the ComBat parametric empirical Bayes technique in the R package sva [27]. Probe-to-gene mapping was conducted with the nuID mapping scheme implemented in lumi. The gene expression values are deposited in the NIH Gene Expression Omnibus (GEO) database with ID GSE86216.

2.4. FCT imaging by OCT

OCT images were obtained with the C7-XR OCT device (St. Jude Medical, Minneapolis, Minnesota) with continuous intracoronary contrast injection. The OCT catheter was placed at least 10 mm distal to

Please cite this article as: K.W. Johnson, B.S. Glicksberg, K. Shameer, et al., A transcriptomic model to predict increase in fibrous cap thickness in response to high-dose statin ..., EBioMedicine, https://doi.org/10.1016/j.ebiom.2019.05.007

the target lesion, after which each OCT pullback imaged 54 mm of the vessel. OCT images were evaluated offline by an independent core laboratory (Cardiovascular Research Foundation, New York, New York) and analyzed at 1-mm intervals according to previously validated criteria for plaques characterization with the St. Jude Medical review workstation. The FCT of each lipid-rich plaque was measured three times at each time point at its thinnest point (6 measurements in total). The average value was then calculated for both baseline and follow-up time points for each patient (mean baseline and follow FCT). Finally, we calculated the change in FCT between baseline and follow-up timepoints by taking the absolute value of the difference. Compared to baseline, the location of the thinnest portion of the fibrous cap at follow-up remained the same in 65 of 69 patients (94.2%).

2.5. Responder identification

Each patient was identified as either a statin responder or non-responder depending upon whether FCT increased or did not increase from baseline time point to follow-up time point (Fig. 1b). We dichotomized this measurement because it is clinically meaningful to ascertain whether a patient's FCT will change in response to statin therapy, and thus whether statin therapy benefits a given patient. For example, in

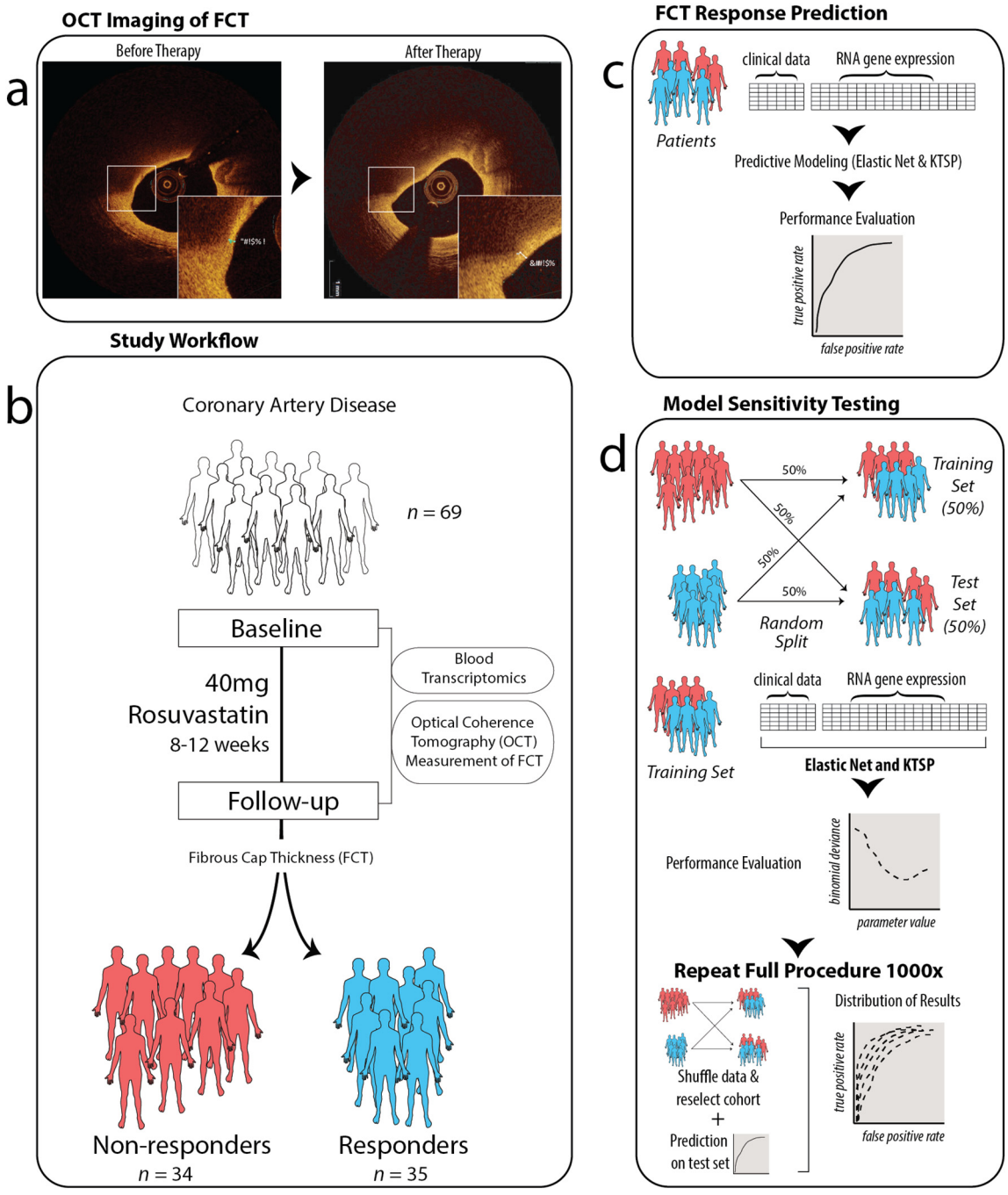

Fig. 1. Central figure illustrating the components of study. (a) An example OCT image of an atherosclerotic plaque, before and after 8–12 weeks of high intensity rosuvastatin therapy. (b) The study workflow. Blood transcriptomics and OCT imaging were performed at 69 patients at baseline and follow-up periods. 35 patients had increased FCT (responders), and 34 patients did not (non-responders). (c) Predictive modeling of FCT response. We combined clinical variables and transcriptomic data and used two machine learning methods to predict responder type. (d) Graphical explanation of extensive sensitivity testing. We iteratively performed the strategy depicted in panel (c) upon randomly selected subsets of our dataset in order to understand the variability of the results.

Please cite this article as: K.W. Johnson, B.S. Glicksberg, K. Shameer, et al., A transcriptomic model to predict increase in fibrous cap thickness in response to high-dose statin ..., EBioMedicine, https://doi.org/10.1016/j.ebiom.2019.05.007

2014, Kim et al. published a machine learning model using the expression levels of 100 genes to predict how LDL cholesterol levels in lymphoblastic cell lines changes following statin therapy [21]. This model for LDL cholesterol was able to classify the highest 15% of responders from the lowest 15% of responders with area under the ROC curve (AUC) up to 0.98 [28]. However, the model's utility diminished as they attempted to differentiate more moderate responders from moderate non-responders, with an AUC only modestly better than a random classifier (AUC < 0.6) for fully dichotomized data. Thus, for this initial step we chose to predict the binary outcome of response to statin therapy with increased FCT or not, instead of attempting to model the quantitative increase in cap thickness.

2.6. Statistical analysis

We used the glmnet [29] package in R [23] version 3.3.3 to fit elastic net regularized generalized linear models using a logistic link function via penalized maximum likelihood to transcriptomic data. The elastic net mixes the regression penalty of ridge regression with that of the least absolute shrinkage and selection operator (LASSO) regression. Elastic net coefficient estimates are computed by minimizing Eq. (1). Logistic regression in glmnet specifically solves the problem given in Eq. (2) for a sequence of λ values at a given α value. We search the mixing parameter α sequence in steps of 0.01 from $\alpha = 0$ (full ridge regression) to $\alpha = 1$ (full LASSO regression) for the value which minimizes the binomial deviance of the model, computed via leave-one-out cross validation (LOOCV) (Supplemental Material).

$$\beta = \operatorname{argmin}\left(||y - X\beta||^2 + \lambda_2||\beta||^2 + \lambda_2||\beta||_1\right) \tag{1}$$

$$\min(\beta_0, \beta) - \left[\frac{1}{N}\sum_{i=1}^{N} y_i(\beta_0 + x_i^T)\right) - \log\left(1 + e^{\beta_0 + x_i^T\beta}\right)\Big)\Big] + \lambda\left((1-\alpha)||\beta||_2^2/2 + a||\beta||_1\right) \tag{2}$$

To fit K top scoring pair (KTSP) classifiers, we used the switchbox [30] Bioconductor R package. KTSP is an extension of the original Top Scoring Pairs (TSP) classification algorithm [31]. Briefly, given a matrix of expression values *X* with dimension genes x samples where columns 1 to N_1 are samples belonging to outcome class 1 and columns N_{1+1} to N_2 are samples of class 2, KTSP finds the best pair of genes *i* and *j* which maximize the value Δ_{ij}, as given in Eq. (3).

$$\Delta_{ij} = | P_{ij}(1) - P_{ij}(2) | \tag{3}$$

where $P_{ij}(1)$ and $P_{ij}(2)$ are given by Eqs. (4) and (5):

$$P_{ij}(1) = \Pr(x_i < x_j | \text{class} = 1) = \frac{1}{N_1}\sum_{n=1}^{N_1} I(x_{in} < x_{jn}) \tag{4}$$

$$P_{ij}(2) = \Pr(x_i < x_j | \text{class} = 2) = \frac{1}{N_2}\sum_{n=N_1+1}^{N_2} I(x_{in} < x_{jn}) \tag{5}$$

For the extension to K pairs of genes, the next highest top-scoring pairs of genes are added to the classifier for a given range of K. The addition of genes is penalized by choosing the value of K which produces a classifier that maximizes $\tau = \delta_i/\sigma_i$, where δ_i is sum of K pairwise gene scores as given in Eq. (3) and σ_i is the maximum-likelihood estimator of the variance [32].

2.7. Incorporation of clinical variables into predictive models

To assess whether inclusion of clinical variables may help to alter or improve predictive performance, we conducted two additional analyses. First, we built an elastic net regression model as described above composed of only clinical variables. Second, we built another elastic net regression model as described above including both clinical variables as well as the transcriptomic data from the original elastic net model. A full list of clinical variables is included in the Supplemental Material.

2.8. Gene function analysis

We sought to ascertain the biological function of genes found to be predictive of FCT response by characterizing their Gene Ontology Enrichments. We used the *TopGO* R package to perform weighted Fisher Exact tests. We used the *TopGO* "weight01" algorithm to account for the hierarchical nature of Gene Ontology functional group assignments and to protect against false discovery from multiple hypothesis testing.

3. Results

3.1. FCT responder prediction

Clinical values for our patient cohort are available in Table 1. The mean increase in FCT for rosuvastatin-responders was 36.9 ± 69.8 μm. The mean change in FCT for rosuvastatin non-responders was −4.41 ± 7.05 μm. A plot demonstrating the distribution of FCT values is available in the Supplementary Material. LDL cholesterol and total cholesterol levels were not significantly different between responder and non-responders, either at baseline, follow-up, or when examining change in lipid levels from baseline to follow-up (Supplementary Material, Table S3). Using transcriptomic data to predict FCT response, we obtained a final model with leave-one-out-cross-validation (LOOCV) area under the receiver operating characteristic curve of 0.975. We thus could classify individuals as FCT statin responders or non-responders with high fidelity. The elastic net model using 73 genes had an accuracy of 92.8%, sensitivity of 94.1%, and specificity of 91.4%. Similarly, the KTSP classifier could discriminate between responders and non-responders with high performance, obtaining LOOCV accuracy of 95.7%, sensitivity of 94.3%, and specificity of 97.1% (Fig. 2a). Notably, this classifier required only 18 genes. Fig. 3 provides a visual demonstration of how well this small number of genes divide responders and non-responders.

Table 1
Clinical variables of individuals in dataset, stratified by Responder/Non-responder type.

	Non-responder ($n = 34$)	Responder ($n = 35$)	Total ($n = 69$)	*P*-Value*
Gender				
Female	9 (26.5%)	13 (37.1%)	22 (31.9%)	0.489
Male	25 (73.5%)	22 (62.9%)	47 (68.1%)	
Age at event (Years)				
Mean (SD)	67.1 (9.58)	62.9 (11.2)	65.0 (10.6)	0.11
Weight (kg)				
Mean (SD)	87.3 (17.0)	79.0 (13.6)	83.1 (15.8)	0.03
Smoking				
Current/Former	16 (47.1%)	15 (42.9%)	31 (44.9%)	0.913
Never	18 (52.9%)	20 (57.1%)	38 (55.1%)	
Systolic BP (mmHg)				
Mean (SD)	145 (22.2)	137 (24.8)	141 (23.7)	0.17
Diastolic BP (mmHg)				
Mean (SD)	73.3 (12.6)	69.5 (10.2)	71.4 (11.5)	0.17
Total Cholesterol (mg/dL)				
Mean (SD)	147 (42.0)	146 (35.6)	147 (38.6)	0.93
HDL Cholesterol (mg/dL)				
Mean (SD)	39.9 (10.6)	41.9 (13.9)	40.9 (12.3)	0.52
LDL Cholesterol (mg/dL)				
Mean (SD)	83.3 (39.6)	80.8 (32.3)	82.1 (35.8)	0.78
Hemoglobin A1c (%)				
Mean (SD)	7.06 (1.39)	6.80 (1.58)	6.93 (1.48)	0.48
hs-CRP (mg/L)				
Mean (SD)	3.42 (0.643)	3.52 (0.522)	3.47 (0.582)	0.48

* *P* values for continuous variables computed with the two-sample *t*-test. *P* values for categorical variables computed with the Chi-square test of independence.

Please cite this article as: K.W. Johnson, B.S. Glicksberg, K. Shameer, et al., A transcriptomic model to predict increase in fibrous cap thickness in response to high-dose statin ..., EBioMedicine, https://doi.org/10.1016/j.ebiom.2019.05.007

3.2. Sensitivity testing

When constructing predictive models, the potential for overfitting the training dataset is of high concern, especially when sample sizes are small. We conducted extensive cross validation and sensitivity testing to characterize the stability of our predictive models to patient sampling (Fig. 1d). Briefly, our strategy for sensitivity testing was to (1) randomly split the data in half; (2) build a predictive model on half of the data, using LOOCV to select the most robust model; (3) test this model on the remaining held out 50% of data to obtain a true test-set validation of the model. We then repeated steps 1–3 1000 times to gain insight into the sensitivity of this model-building procedure and distributions of the model statistics (Fig. 2b and c).

We could predict FCT responder status in a held-out test set with high discrimination. The median elastic net AUC was 0.969, and the median KTSP AUC was 0.972. Our sensitivity analysis revealed that even the lower-performing models still performed with high accuracy on the held-out testing set: 97% of the elastic net models has AUCs of better than 0.886, and 97% of the KTSP models had AUCs better than 0.946.

3.3. Incorporation of clinical variables into predictive models

After discovering that the transcriptome-only model provided excellent results for predictive models, we explored whether clinical variables could either augment or replace our transcriptome-based predictive models. We included 49 different clinical and laboratory parameters. A full table of clinical variables is given in the supplementary material Table S2. We followed identical training procedures to those described above for the transcriptome-only model. An elastic net trained on transcriptome plus clinical data produced identical results to the elastic net trained on transcriptome data alone. The elastic net training algorithm did not select any clinical variables for inclusion in the final predictive model, instead choosing the same 73 genes. An elastic net trained on the clinical variables only and then used to predict FCT response performed very poorly (AUC = 0.16, Accuracy = 0.33, Sensitivity = 0.09, Specificity = 0.57) (Supplementary Fig. 3). In fact, the very poor AUC of this model mean that is performed well below random chance (AU = 0.5). Theoretically, simply flipping each of this model's predictions would have made it perform moderately well (AUC 0.83). However, it would be statistically unfair to reverse model predictions after looking at the outcome of the trained model on the test set.

3.4. Gene ontology term enrichment

Our elastic net regression model utilized 73 gene expression values (Table 2). The final KTSP model used only 18 genes (nine pairs of top-scoring genes) (Table 3). 12 of the 18 genes from the KTSP algorithm were present in the elastic net regression selected genes (Fig. 4). We performed Gene Ontology (GO) biological process enrichment testing upon significant elastic net covariates to understand biological processes which may regulate the fibrous cap thickening response to statin therapy. GO terms are a controlled vocabulary of curated genes and gene products mapped to specific biological processes, molecular functions, and cellular components. We found 58 significant GO term enrichments at $p < .05$, including many involved with (1) immune cell function (GO terms: locomotion involved in locomotory behavior; T cell homeostasis; positive regulation of MHC class I biosynthesis; positive regulation of NK T cell activation; negative regulation of T cell mediated cytotoxicity; monocyte activation), and (2) cholesterol metabolism (cholesterol catabolic process; cholesterol import; C21-steroid hormone metabolic process; SREBP signaling pathway), and (3) protein modification (protein phosphorylation; protein adenylylation; protein K11-linked ubiquitination; N-terminal

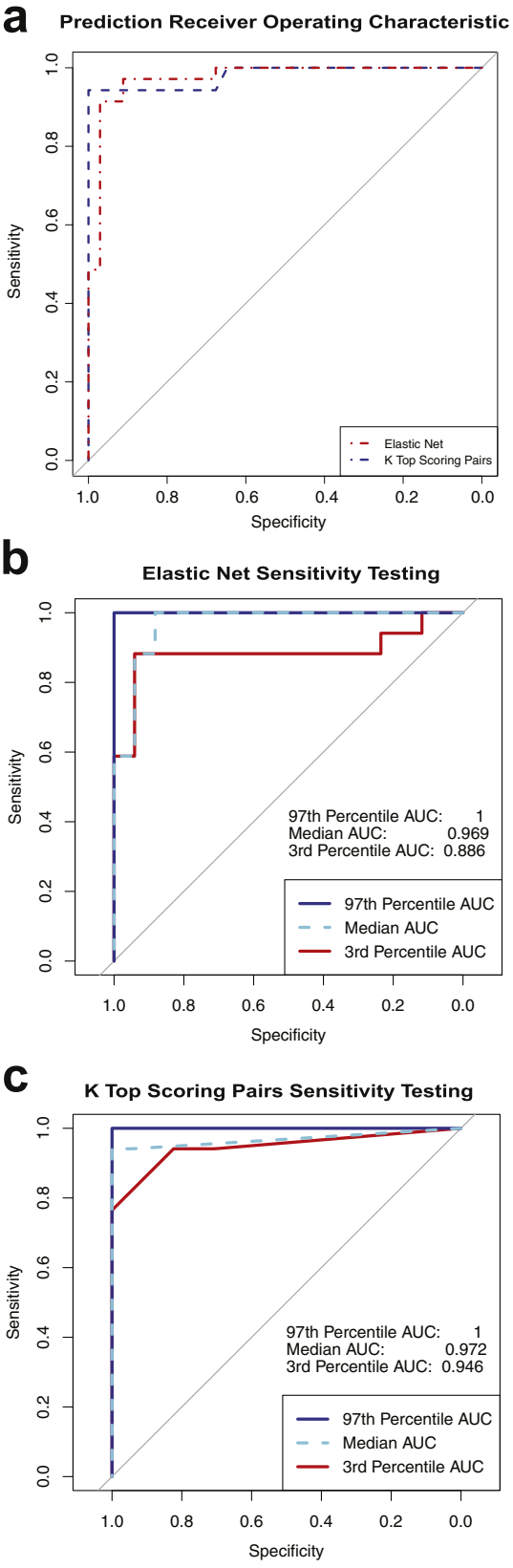

Fig. 2. Predictive Model Receiver Operating Characteristic Curves. The receiver operating characteristic (ROC) curves for the elastic net and K top scoring pairs predictive models are shown in (a). ROC scores were computed for KTSP by dividing the number of votes by number of potential votes (i.e. gene pairs) in the classifier as the predicted probability. Sensitivity testing using elastic net (b) and KTSP (c) showed performance is highly robust to sampling error.

Please cite this article as: K.W. Johnson, B.S. Glicksberg, K. Shameer, et al., A transcriptomic model to predict increase in fibrous cap thickness in response to high-dose statin ..., EBioMedicine, https://doi.org/10.1016/j.ebiom.2019.05.007

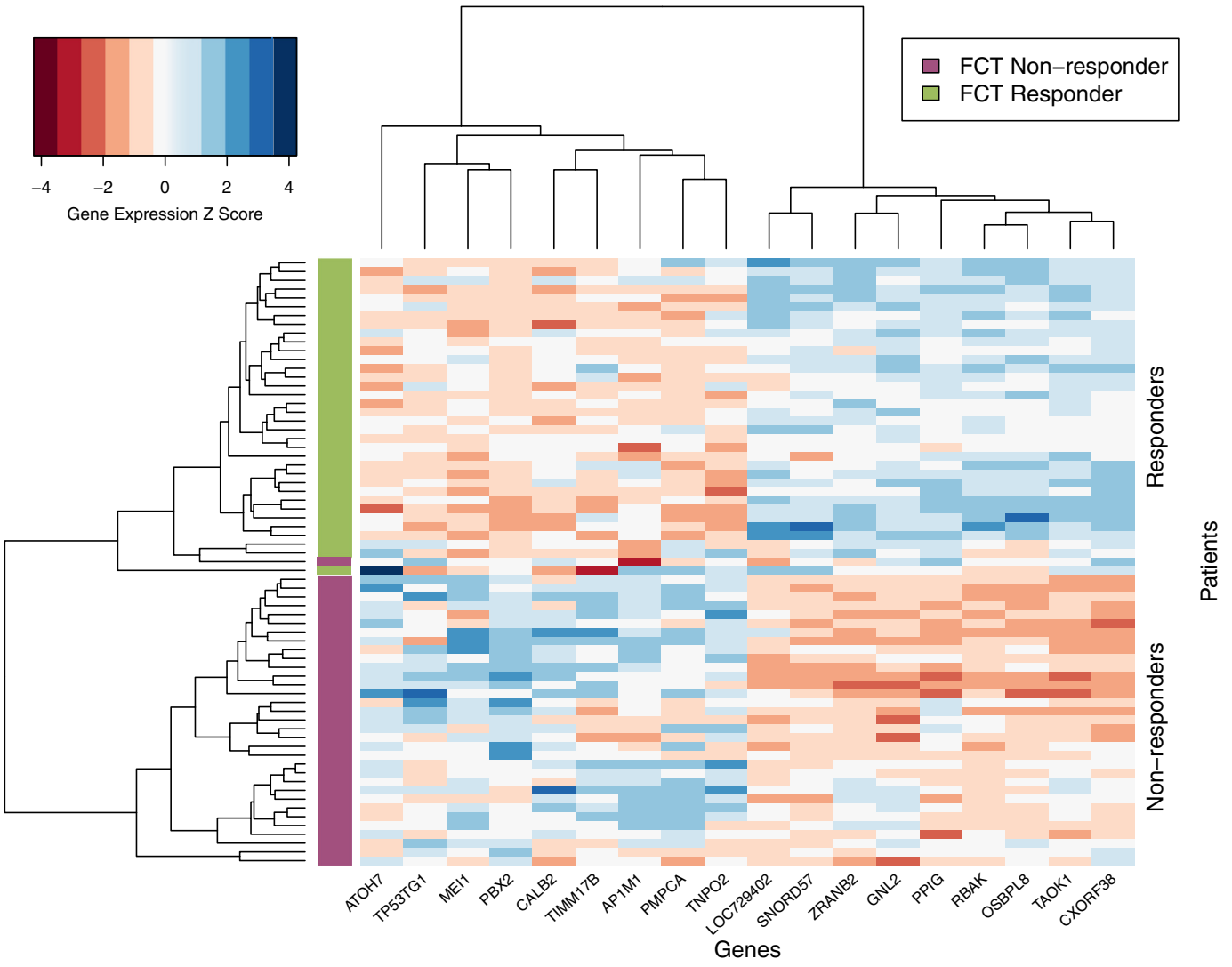

Fig. 3. Heatmap of 18 Genes Selected by K-Top-Scoring-Pairs Algorithm (KTSP). Patient samples and genes were grouped using hierarchical clustering. Gene expression values were normalized for plotting by dividing the gene's microarray signal intensity minus the mean signal intensity for that gene by the standard deviation of signal intensity for that gene (Z score).

peptidyl-lysine acetylation; protein K48-linked ubiquitination). A full table of GO terms with associated *p* values is available in the Supplementary Material, Table S4.

4. Discussion

In this post-hoc analysis of the Yellow-II study, we demonstrate that gene expression (when characterized via microarray with statistical models capable of dealing with high-dimensional gene expression data) can be used to reliably predict whether a given patient's fibrous cap thickness will increase with statin treatment. Our study is one of the first attempts to use predictive modeling techniques applied to transcriptomic data to predict plaque response to lipid lowering therapy. Transcriptomic profiling has been used for many years in cardiovascular research to identify genes differentially expressed between cases and controls in a variety of diseases [33], or somewhat more recently to analyze putative biological interactions using methods such as weighted gene coexpression networks [34,35]. We attempted to, instead of simply identifying statistically differentially expressed genes across a condition, use gene expression data to predict outcomes. In fact, statistically significant variables (e.g., differentially expressed genes) are often poorly predictive variables [36], and often explain only a fraction of the variance in the outcome. To accurately predict an outcome, one must instead explain a significant fraction of the outcome variance, which is a more difficult task. In the present study, we explain much of the variance in outcome as demonstrated by our high prediction statistics (AUCs on the order of 0.95). For clinically applicable personalized medicine, a physician would be interested in a test which is predictive of an outcome for an individual patient, such as the test we present here.

Second, we demonstrate that the genes which are most predictive of FCT response to statin therapy are those broadly involved in immune cell function, cholesterol biosynthesis and metabolism, and cellular protein modification processes such as phosphorylation and adenylation. A number of researchers have linked fibrous cap production and degradation to inflammatory processes [37]. The same statin used in the Yellow II trial—rosuvastatin—was demonstrated to reduce major cardiovascular events in the JUPITER study with elevated levels of the nonspecific inflammatory marker c-reactive protein (CRP) [38] but presumably normal cholesterol levels. Indeed, we have previously demonstrated in this cohort an inverse relationship between CRP levels and FCT [6,39]. Through MRI-PET/CT and mathematical fluid-structure-interaction modeling it was observed that in mice that cap inflammation can lead to increased cap strain, especially in the presence of thin-caps and hypertension [40]. The CANTOS trial of the Interleukin-1β (IL-1β) monoclonal antibody successfully demonstrated that selective anti-

Table 2
Elastic Net covariates.

Covariate	Coefficient	Covariate	Coefficient	Covariate	Coefficient
(Intercept)	−90.652	ZSCAN12L1	−0.047	WDR23	0.322
RNF113A	−1.155	GSDM1	−0.045	CBX5	0.365
MEI1	−1.042	RNU6–15	−0.041	XRCC5	0.397
PBX2	−0.711	DTX3L	−0.026	EIF2A	0.401
STARD5	−0.587	SPATA13	0.006	NOL7	0.424
TIMM17B	−0.568	HSPH1	0.012	FMO2	0.477
RNU4ATAC	−0.557	CLK1	0.016	C6ORF27	0.496
FICD	−0.455	SMAP1	0.089	MEF2A	0.549
HCST	−0.408	TINP1	0.104	CSE1L	0.577
CALB2	−0.401	TMEM189-UBE2V1	0.112	SNORD57	0.633
TP53TG1	−0.399	SLC46A3	0.118	HSPA9	0.635
DHX37	−0.393	KIAA0020	0.118	AQP12B	0.656
USP48	−0.393	PPP3CB	0.126	TMEM183A	0.661
FOXO4	−0.299	XCR1	0.13	TAOK1	0.73
HNRNPUL2	−0.29	UBE2D3	0.148	C14ORF68	0.74
MRPS16	−0.268	CXORF38	0.157	CREBBP	0.801
MT1G	−0.26	C19ORF12	0.181	SFRS2	0.829
RNU6-1	−0.222	NR2C2	0.204	MBNL3	0.837
AURKAIP1	−0.204	SR140	0.223	YARS2	0.878
FNIP2	−0.133	GTF2A1	0.236	CLN6	0.981
UBE2B	−0.129	ZRANB2	0.236	RBAK	1.012
SLC39A3	−0.108	SEC31A	0.253	KCTD7	1.069
LOC148413	−0.074	ZKSCAN1	0.273	LOC729402	1.118
AKR1D1	−0.066	TP53RK	0.307	IRX6	2.525
ZNF264	−0.053	GNL2	0.313		

inflammatory therapy can reduce the incidence of adverse cardiovascular events [41]. Other recent work has focused upon the role of IL-1β-modulated neutrophils in the atherosclerotic plaque [42], or the roles of neutrophil extracellular traps [43,44]. Another recent paper has demonstrated that neutrophil reprogramming can modulate chronic inflammation in atherosclerosis [45]. We note that two important proteins related to neutrophil extravasation, *XCR1* and *XRCC5*, were important in our elastic net predictive model. On the other hand, at least one recent paper has experimentally demonstrated in mice the unexpected finding that IL-1β may actually reduce protective remodeling of atherosclerotic fibrous caps [46]. Whatever the ultimate effect of this particular inflammatory mediator, it is evident that inflammatory processes play an important role in the development of atherosclerosis, and it is thus not surprising that these genes also can be used for prediction of FCT response. Similarly, OCT imaging and intravascular ultrasound imaging have demonstrated that synthesis and absorption of cholesterol were related to plaque vulnerability and FCT, and also that HDL-C was an independent marker for FCT; these findings echo previous results of the YELLOW-II study [6,47–50]. A number of other genes used by both the elastic net and KTSP algorithms may be of interest for further investigation of their role in the formation of atherosclerotic fibrous caps. For example, *TAOK1* is a ubiquitously expressed kinase with important identified roles in modulation of interleukin-17 [51]. Interestingly, *TAOK1* has been shown in vascular smooth muscle to be differentially expressed following the administration of captopril and is also involved in the angiotensin-II-mediated mitogen-activated protein kinsase signaling pathway [52]. The gene *ZRANB2* has shown to regulate the process of vasculogenic mimicry in neoplams [53] and also to be involved in the immunologic response to gram-negative bacteria [54]. The intracellular calcium-binding protein calbindin 2 (*CALB2*) plays a role in a number of cellular functions, such as the buffering of intracellular calcium, and has been implicated in vascular smooth muscle cell adhesion [55], cardiogenic differentiation [56], cellular response to ischemia [57], steroidogenesis [58]. Finally, we were interested in the negative correlation between expression of steroidogenic acute regulator-related lipid transfer domain protein (*STARD5*) and FCT response. This family of proteins, which modulate intracellular lipid and cholesterol transport, have been investigated for their roles in cholesterol and lipid metabolism [59,60] as well as their function in atherogenic lipid phenotypes [61] and fibroblast function [62]. This gene also has been identified as an emerging cardiovascular disease risk biomarker [63].

Table 3
K Top-Scoring Pairs (KTSP) Gene Pairs.

Gene Pair	Score
TAOK1, MEI1	0.884
ZRANB2, PMPCA	0.884
GNL2, AP1M1	0.884
PPIG, PBX2	0.855
ATOH7, RBAK	0.826
LOC729402, CALB2	0.797
CXORF38, TNPO2	0.797
TIMM17B, OSBPL8	0.796
TP53TG1, SNORD57	0.796

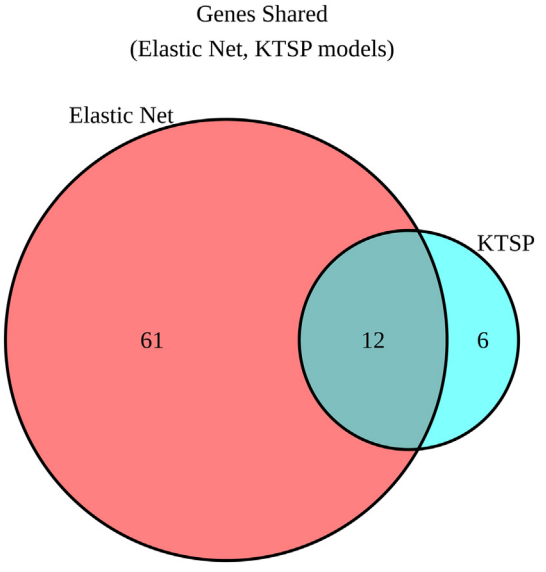

Fig. 4. Genes Shared Between Elastic Net and KTSP predictive models. Venn diagram showing the overlap of genes included in the elastic net and KTSP algorithms. 12 of 18 KTSP genes were also selected by the elastic net algorithm.

Please cite this article as: K.W. Johnson, B.S. Glicksberg, K. Shameer, et al., A transcriptomic model to predict increase in fibrous cap thickness in response to high-dose statin ..., EBioMedicine, https://doi.org/10.1016/j.ebiom.2019.05.007

4.1. Study limitations

Our study must be interpreted in the light of its limitations. The major limitation to this study is the limited sample size, essentially due to the invasive nature of repeated OCT imaging. We are exploring options for external validation, which would require a second clinical trial. However, we are comforted by the extensive cross validation efforts we undertook to test that we are not overfitting our model to the dataset. Furthermore, the gene expression results have been previously validated in the Yellow II study [6]. Second, there are currently few transcriptome-based tests utilized in cardiovascular medicine, and likely none used in routine clinical cardiology. In this measure, cardiovascular disease has lagged behind other fields such as oncology, which have rapidly embraced transcriptome characterization via microarray and RNA sequencing [64]. However, there is no fundamental reason that transcriptomic medicine will not ultimately prove to be useful in cardiovascular medicine, especially as future efforts toward precision cardiology increase [65–67]. However, we caution that this pilot study requires validation in a randomized, controlled clinical trial before its results could be applied clinically.

5. Conclusions

Taken altogether, we believe our results represent a step toward enabling statin therapy recommendation based upon the patient's transcriptome in chronic cardiovascular disease. OCT is a precise and highly useful technique with a large body of existing literature, and it would require further validation studies before a transcriptomic approach could be used with confidence in the clinic. However, our findings here are promising: a non-invasive, blood-based test appears to detect with high sensitivity and specificity whether an individual patient in our sample will benefit from high intensity statin therapy. Since FCT is associated with systemic cardiovascular disease outcomes and risk of myocardial infarction, development of a non-invasive diagnostic test for patient response is potentially of high clinical interest. We envision that predicted non-responders to statin therapy may benefit from additional intervention, such as earlier introduction of PCSK9 inhibitors.

Funding sources

No specific funding sources exist for this project. Kipp Johnson is partially supported by the Medical Scientist Training Program grant 2T32GM007280-41 to the Icahn School of Medicine at Mount Sinai (National Institute of General Medical Sciences (NIGMS), National Institute of Health (NIH), United States). No funders had any role in study design, data collection, data analysis, interpretation, writing of the report.

Declaration of interests

Dr. Annapoorna Kini received grants from AstraZeneca (Gaithersburg, MD) to conduct the initial Yellow II clinical trial of high-intensity rosuvastatin therapy. Dr. Shameer Khader is an employee of AstraZeneca. No other authors have any financial conflicts of interest to disclose.

Author contributions

All authors contributed to study design and data analysis. Data collection was performed by Yuliya Vengrenyuk, Samin Sharma, Jagat Narula, and Annapoorna Kini. The first draft of the article was written by Kipp Johnson, Benjamin Glicksberg, Khader Shameer, Yuliya Vengrenyuk, Jagat Narula, Joel Dudley, and Annapoorna Kini. All authors contributed to subsequent revisions and drafts of the paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.05.007.

References

[1] Narula J, Nakano M, Virmani R, et al. Histopathologic characteristics of atherosclerotic coronary disease and implications of the findings for the invasive and noninvasive detection of vulnerable plaques. J Am Coll Cardiol 2013;61(10):1041–51.
[2] Burke AP, Farb A, Malcom GT, Liang YH, Smialek J, Virmani R. Coronary risk factors and plaque morphology in men with coronary disease who died suddenly. N Engl J Med 1997;336(18):1276–82.
[3] Otsuka F, Joner M, Prati F, Virmani R, Narula J. Clinical classification of plaque morphology in coronary disease. Nat Rev Cardiol 2014;11(7):379–89.
[4] Feldman MD, Milner TE. Improving plaque classification with optical coherence tomography. JACC Cardiovasc Imaging 2017;11(11):1677–8. https://doi.org/10.1016/j.jcmg.2017.11.007.
[5] Kini AS, Vengrenyuk Y, Yoshimura T, et al. Fibrous cap thickness by optical coherence tomography in vivo. J Am Coll Cardiol 2017;69(6):644–57.
[6] Kini AS, Vengrenyuk Y, Shameer K, et al. Intracoronary imaging, cholesterol efflux, and transcriptomes after intensive statin treatment: the YELLOW II study. J Am Coll Cardiol 2017;69(6):628–40.
[7] Yamagishi M, Terashima M, Awano K, et al. Morphology of vulnerable coronary plaque: insights from follow-up of patients examined by intravascular ultrasound before an acute coronary syndrome. J Am Coll Cardiol 2000;35(1):106–11.
[8] Jang IK, Bouma BE, Kang DH, et al. Visualization of coronary atherosclerotic plaques in patients using optical coherence tomography: comparison with intravascular ultrasound. J Am Coll Cardiol 2002;39(4):604–9.
[9] Gardner CM, Tan H, Hull EL, et al. Detection of lipid core coronary plaques in autopsy specimens with a novel catheter-based near-infrared spectroscopy system. JACC Cardiovasc Imaging 2008;1(5):638–48.
[10] Kini AS, Baber U, Kovacic JC, et al. Changes in plaque lipid content after short-term intensive versus standard statin therapy: the YELLOW trial (reduction in yellow plaque by aggressive lipid-lowering therapy). J Am Coll Cardiol 2013;62(1):21–9.
[11] Sato A, Hoshi T, Kakefuda Y, et al. In vivo evaluation of fibrous cap thickness by optical coherence tomography for positive remodeling and low-attenuation plaques assessed by computed tomography angiography. Int J Cardiol 2015;182:419–25.
[12] Sun J, Zhao XQ, Balu N, et al. Carotid plaque lipid content and fibrous cap status predict systemic CV outcomes: the MRI substudy in AIM-HIGH. JACC Cardiovasc Imaging 2017;10(3):241–9.
[13] Eken SM, Jin H, Chernogubova E, et al. MicroRNA-210 enhances fibrous cap stability in advanced atherosclerotic lesions. Circ Res 2017;120(4):633–44.
[14] Wang Z, Cho YS, Soeda T, et al. Three-dimensional morphological response of lipid-rich coronary plaques to statin therapy: a serial optical coherence tomography study. Coron Artery Dis 2016;27(5):350–6.
[15] Park SJ, Kang SJ, Ahn JM, et al. Effect of statin treatment on modifying plaque composition: a double-blind, randomized study. J Am Coll Cardiol 2016;67(15):1772–83.
[16] Oesterle A, Laufs U, Liao JK. Pleiotropic effects of statins on the cardiovascular system. Circ Res 2017;120(1):229–43.
[17] Komukai K, Kubo T, Kitabata H, et al. Effect of atorvastatin therapy on fibrous cap thickness in coronary atherosclerotic plaque as assessed by optical coherence tomography: the EASY-FIT study. J Am Coll Cardiol 2014;64(21):2207–17.
[18] Kataoka Y, St John J, Wolski K, et al. Atheroma progression in hyporesponders to statin therapy. Arterioscler Thromb Vasc Biol 2015;35(4):990–5.
[19] Wei KK, Zhang LR. Interactions between CYP3A5*3 and POR*28 polymorphisms and lipid lowering response with atorvastatin. Clin Drug Investig 2015;35(9):583–91.
[20] Kataoka Y, Hammadah M, Puri R, et al. Plaque microstructures in patients with coronary artery disease who achieved very low low-density lipoprotein cholesterol levels. Atherosclerosis 2015;242(2):490–5.
[21] Trompet S, Postmus I, Slagboom PE, et al. Non-response to (statin) therapy: the importance of distinguishing non-responders from non-adherers in pharmacogenetic studies. Eur J Clin Pharmacol 2016;72(4):431–7.
[22] Dudley JT, Tibshirani R, Deshpande T, Butte AJ. Disease signatures are robust across tissues and experiments. Mol Syst Biol 2009;5:307.
[23] R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
[24] Du P, Kibbe WA, Lin SM. nuID: a universal naming scheme of oligonucleotides for illumina, affymetrix, and other microarrays. Biol Direct 2007;2:16.
[25] Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24(13):1547–8.
[26] Lin SM, Du P, Huber W, Kibbe WA. Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic Acids Res 2008;36(2):e11.
[27] Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012;28(6):882–3.
[28] Kim K, Bolotin E, Theusch E, Huang H, Medina MW, Krauss RM. Prediction of LDL cholesterol response to statin using transcriptomic and genetic variation. Genome Biol 2014;15(9):460.
[29] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1–22.
[30] Afsari B, Fertig EJ, Geman D, Marchionni L. switchBox: an R package for k-top scoring pairs classifier development. Bioinformatics 2015;31(2):273–4.

[31] Geman D, d'Avignon C, Naiman DQ, Winslow RL. Classifying gene expression profiles from pairwise mRNA comparisons. Stat Appl Genet Mol Biol 2004;3:19.
[32] Afsari B, Braga-Neto UM, Geman D. Rank discriminants for predicting phenotypes from Rna expression. Ann Appl Stat 2014;8(3):1469–91.
[33] Mendelson MM, Marioni RE, Joehanes R, et al. Association of body mass index with DNA methylation and gene expression in blood cells and relations to cardiometabolic disease: a Mendelian randomization approach. PLoS Med 2017;14(1):e1002215.
[34] Miao L, Yin RX, Pan SL, Yang S, Yang DZ, Lin WX. Weighted gene co-expression network analysis identifies specific modules and hub genes related to hyperlipidemia. Cell Physiol Biochem 2018;48(3):1151–63.
[35] Chen J, Yu L, Zhang S, Chen X. Network analysis-based approach for exploring the potential diagnostic biomarkers of acute myocardial infarction. Front Physiol 2016;7:615.
[36] Lo A, Chernoff H, Zheng T, Lo SH. Why significant variables aren't automatically good predictors. Proc Natl Acad Sci U S A 2015;112(45):13892–7.
[37] Ridker PM, Narula J. Will reducing inflammation reduce vascular event rates? JACC Cardiovasc Imaging 2018;11(2 Pt 2):317–9.
[38] Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. N Engl J Med 2008;359(21):2195–207.
[39] Chamaria S, Johnson KW, Vengrenyuk Y, et al. Intracoronary imaging, cholesterol efflux, and transcriptomics after intensive statin treatment in diabetes. Sci Rep 2017;7(1):7001.
[40] Tang D, Yang C, Huang S, et al. Cap inflammation leads to higher plaque cap strain and lower cap stress: an MRI-PET/CT-based FSI modeling approach. J Biomech 2017;50:121–9.
[41] Ridker PM, Everett BM, Thuren T, et al. Antiinflammatory therapy with canakinumab for atherosclerotic disease. N Engl J Med 2017;377(12):1119–31.
[42] Tall AR, Westerterp M. Inflammasomes, neutrophil extracellular traps, and cholesterol. J Lipid Res 2019;60(4):721–7. https://doi.org/10.1194/jlr.S091280.
[43] Soehnlein O, Steffens S, Hidalgo A, Weber C. Neutrophils as protagonists and targets in chronic inflammation. Nat Rev Immunol 2017;17(4):248–61.
[44] Doring Y, Soehnlein O, Weber C. Neutrophil extracellular traps in atherosclerosis and atherothrombosis. Circ Res 2017;120(4):736–43.
[45] Geng S, Zhang Y, Lee C, Li L. Novel reprogramming of neutrophils modulates inflammation resolution during atherosclerosis. Sci Adv 2019;5(2):eaav2309.
[46] Gomez D, Baylis RA, Durgin BG, et al. Interleukin-1beta has atheroprotective effects in advanced atherosclerotic lesions of mice. Nat Med 2018;24(9):1418–29.
[47] Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol 2014;63(25 Pt B):2889–934.
[48] Nasu K, Terashima M, Habara M, et al. Impact of cholesterol metabolism on coronary plaque vulnerability of target vessels: a combined analysis of virtual histology intravascular ultrasound and optical coherence tomography. JACC Cardiovasc Interv 2013;6(7):746–55.
[49] Ozaki Y, Tanaka A, Komukai K, et al. High-density lipoprotein cholesterol level is associated with fibrous cap thickness in acute coronary syndrome. Circ J 2013;77(12):2982–9.
[50] Chhatriwalla AK, Rader DJ. Intracoronary imaging, reverse cholesterol transport, and transcriptomics: precision medicine in CAD? J Am Coll Cardiol 2017;69(6):641–3.
[51] Zhang Z, Tang Z, Ma X, et al. TAOK1 negatively regulates IL-17-mediated signaling and inflammation. Cell Mol Immunol 2018;15(8):794–802.
[52] Chen FC, Brozovich FV. Gene expression profiles of vascular smooth muscle show differential expression of mitogen-activated protein kinase pathways during captopril therapy of heart failure. J Vasc Res 2008;45(5):445–54.
[53] Li X, Xue Y, Liu X, et al. ZRANB2/SNHG20/FOXK1 axis regulates vasculogenic mimicry formation in glioma. J Exp Clin Cancer Res 2019;38(1):68.
[54] Wang X, Du X, Li H, Zhang S. Identification of the zinc finger protein ZRANB2 as a novel maternal lipopolysaccharide-binding protein that protects embryos of Zebrafish against gram-negative bacterial infections. J Biol Chem 2016;291(8):4019–34.
[55] Prakash SK, LeMaire SA, Guo DC, et al. Rare copy number variants disrupt genes regulating vascular smooth muscle cell adhesion and contractility in sporadic thoracic aortic aneurysms and dissections. Am J Hum Genet 2010;87(6):743–56.
[56] Singhal P, Luk A, Rao V, Butany J. Molecular basis of cardiac myxomas. Int J Mol Sci 2014;15(1):1315–37.
[57] Chen JH, Kuo HC, Lee KF, Tsai TH. Global proteomic analysis of brain tissues in transient ischemia brain damage in rats. Int J Mol Sci 2015;16(6):11873–91.
[58] Xu W, Zhu Q, Liu S, et al. Calretinin participates in regulating steroidogenesis by PLC-Ca(2+)-PKC pathway in Leydig cells. Sci Rep 2018;8(1):7403.
[59] Calderon-Dominguez M, Gil G, Medina MA, Pandak WM, Rodriguez-Agudo D. The StarD4 subfamily of steroidogenic acute regulatory-related lipid transfer (START) domain proteins: new players in cholesterol metabolism. Int J Biochem Cell Biol 2014;49:64–8.
[60] Riegelhaupt JJ, Waase MP, Garbarino J, Cruz DE, Breslow JL. Targeted disruption of steroidogenic acute regulatory protein D4 leads to modest weight reduction and minor alterations in lipid metabolism. J Lipid Res 2010;51(5):1134–43.
[61] Borthwick F, Allen AM, Taylor JM, Graham A. Overexpression of STARD3 in human monocyte/macrophages induces an anti-atherogenic lipid phenotype. Clin Sci (Lond) 2010;119(7):265–72.
[62] Rodriguez-Agudo D, Calderon-Dominguez M, Ren S, et al. Subcellular localization and regulation of StarD4 protein in macrophages and fibroblasts. Biochim Biophys Acta 2011;1811(10):597–606.
[63] Upadhyay RK. Emerging risk biomarkers in cardiovascular diseases and disorders. J Lipids 2015;2015(97):1453.
[64] Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat Rev Genet 2016;17(5):257–71.
[65] Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial intelligence in cardiology. J Am Coll Cardiol 2018;71(23):2668–79.
[66] Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? Heart 2018;104(14):1156–64.
[67] Johnson KW, Shameer K, Glicksberg BS, et al. Enabling precision cardiology through multiscale biology and systems medicine. JACC Basic Transl Sci 2017;2(3):311–27.

Please cite this article as: K.W. Johnson, B.S. Glicksberg, K. Shameer, et al., A transcriptomic model to predict increase in fibrous cap thickness in response to high-dose statin ..., EBioMedicine, https://doi.org/10.1016/j.ebiom.2019.05.007