Accepted Manuscript

Aspirin Exposure Reveals Novel Genes Associated with Platelet Function and Cardiovascular Events

Deepak Voora, MD Derek Cyr, PhD Joseph Lucas, PhD Jen-Tsan Chi, MD, PhD Jennifer Dungan, PhD Timothy A. McCaffrey, PhD Richard Katz, MD L. Kristin Newby, MD, MHS William E. Kraus, MD Richard C. Becker, MD Thomas L. Ortel, MD PhD Geoffrey S. Ginsburg, MD PhD

PII: S0735-1097(13)02547-3

DOI: 10.1016/j.jacc.2013.05.073

Reference: JAC 19055

To appear in: *Journal of the American College of Cardiology*

Received Date: 24 February 2013

Revised Date: 23 April 2013

Accepted Date: 6 May 2013

Please cite this article as: Voora D, Cyr D, Lucas J, Chi J-T, Dungan J, McCaffrey TA, Katz R, Newby LK, Kraus WE, Becker RC, Ortel TL, Ginsburg GS, Aspirin Exposure Reveals Novel Genes Associated with Platelet Function and Cardiovascular Events, *Journal of the American College of Cardiology* (2013), doi: 10.1016/j.jacc.2013.05.073.

Aspirin Exposure Reveals Novel Genes Associated with Platelet Function and Cardiovascular Events

Running title: Platelet Function Genes and Cardiovascular Events

Manuscript word count: 4589

Deepak Voora MD[1,2], Derek Cyr, PhD[1], Joseph Lucas PhD[1], Jen-Tsan Chi, MD, PhD[1], Jennifer Dungan PhD[2], Timothy A. McCaffrey, PhD[3], Richard Katz, MD[4], L. Kristin Newby MD, MHS[2], William E Kraus MD[2], Richard C. Becker MD[2], Thomas L. Ortel MD PhD[2], and Geoffrey S. Ginsburg[1,2] MD PhD

Institute for Genome Sciences & Policy[1] and Department of Medicine[2], Duke University, Durham, NC. Department of Medicine, Division of Genomic Medicine[3] and Cardiology[4], The George Washington University/Medical Faculty Associates, Washington, DC.

Corresponding authors:
Deepak Voora, MD and Geoffrey S Ginsburg, MD, PhD
Institute for Genome Sciences & Policy
101 Science Drive
Duke University,
Durham, NC 27710
Telephone: 919-684-6266
Fax: 919-664-0900
deepak.voora@duke.edu and geoffrey.ginsburg.duke.edu

Funding: This study was funded by institutional funds provided by the Duke Institute for Genome Sciences & Policy, a National Institutes of Health (NIH) T32 Training grant (5T32HL007101 to DV), a grant (5UL1RR024128) from the National Center for Research Resources (NCRR), a component of the NIH, and NIH Roadmap for Medical Research, a grant (5RC1GM091083 to GSG) from the National Institutes of General Medical Sciences, grant (5U01DD000014 to TLO) from the Centers for Disease Control and Prevention, and the David H Murdock Research Institute.

Relationships with Industry:

- GSG (Consultant to United States Diagnostic Standards, Scientific Advisor to CardiDx, Pappas Ventures, and Universal Medicine, and Equity in CardioDx)
- TAM (Equity in Cellgenex)
- LKN (found online: https://www.dcri.org/about-us/conflict-of-interest/Newby-COI_2012-2013.pdf)
- The following authors (DV, GSG, JTC, JEL, RCB, TLO) have filed a provisional patent application regarding the Aspirin Response Signature

Acknowledgements: We would like to acknowledge Jason Rose, MD for collecting the platelet count data for the CATHGEN cohorts.

Abstract

Objectives: To develop RNA profiles that could serve as novel biomarkers for the response to aspirin. **Background:** Aspirin reduces death and myocardial infarction (MI) suggesting that aspirin interacts with biological pathways that may underlie these events**. Methods:** We administered aspirin, followed by whole blood RNA microarray profiling, in a discovery cohort of healthy volunteers (HV1,n=50), and two validation cohorts of volunteers (HV2,n=53) or outpatient cardiology patients (OPC, n=25). Platelet function was assessed by platelet function score (PFS; HV1/HV2) or VerifyNow Aspirin (OPC). Bayesian sparse factor analysis identified sets of coexpressed transcripts, which were examined for association with PFS in HV1 and validated in HV2 and OPC. Proteomic analysis confirmed the association of validated transcripts in platelet proteins. Validated gene sets were tested for association with death/MI in two patient cohorts (n=587, total) from RNA samples collected at cardiac catheterization. **Results:** A set of 60 co-expressed genes named the "aspirin response signature" (ARS) was associated with PFS in HV1 ($r = -0.31$, $p = 0.03$), HV2 ($r = -0.34$, Bonferroni $p = 0.03$), and OPC ($p = 0.046$). Corresponding proteins for 17 ARS genes were identified in the platelet proteome, of which, six were associated with PFS. The ARS was associated with death/MI in both patient cohorts (odds ratio = 1.2, $p = 0.01$ and hazard ratio = 1.5, $p = 0.001$), independent of cardiovascular risk factors. Compared with traditional risk factors, reclassification (net reclassification index = 31 - 37%, $p \leq 0.0002$) was improved by including the ARS or one of its genes, *ITGA2B*. **Conclusions:** RNA profiles of platelet-specific genes are novel biomarkers for identifying those do not response adequately to aspirin and who are at risk for death/MI.

Key words: aspirin, platelets, genes, myocardial infarction, biomarkers

Abbreviations

PFS = Platelet Function Score
ARS = Aspirin Response Signature
CAD = Coronary Artery Disease
MI = Myocardial Infarction
RT-PCR = Real-time Polymerase Chain Reaction
PCR = Polymerace Chain Reaction
RNA = Ribonucleic Acid
IRB = Institutional Review Board
DNA = Deoxyribonucleic Acid
LDLc = low density lipoprotein cholesterol

Introduction

Identification of novel biomarkers for individuals at risk for CAD mortality, primarily due to platelet-mediated cardiovascular events such as MI, is a priority for reducing the burden of cardiovascular disease. Although genome-wide surveys of genomic variation and gene expression can identify loci associated with CAD (1-3), few can serve as biomarkers for cardiovascular events(4).

Aspirin is prescribed for the prevention of cardiovascular events, suggesting that aspirin interacts with biological pathways that may underlie these events. Platelet function assays are a surrogate biomarker for the effects of aspirin and are associated with cardiovascular events.(5) However, platelet function testing is not widely available primarily due to technical complexity. In contrast, whole blood RNA profiling using PCR-based assays is currently a widely available diagnostic testing platform. (6,7) Therefore, we hypothesized that aspirin could be used as a probe in conjunction with whole blood RNA profiling to elucidate novel biomarkers for platelet function in response to aspirin and for cardiovascular outcomes.

Methods

Platelet Function Outcomes in Healthy Volunteers Cohorts at Duke University Medical Center (DUMC)

We previously described (8) discovery and validation healthy volunteer cohorts (HV1 and HV2, Supplemental Methods, Figure 1) and the platelet function score (PFS) - a composite metric of the following platelet function assays: PFA100 (collagen/epinephrine) closure time and the areas under the optical aggregometry curve induced by adenosine diphosphate (10, 5, 1uM), epinephrine (10, 1, 0,5 uM), and collagen (5 , 2 mg/ml). We measured the PFS and mean platelet volume (MPV) in HV1 (n = 50) after 2 weeks of dosing with 325 mg/day non-enteric

coated, immediate release aspirin and HV2 (n = 53) after 4 weeks of dosing with 325 mg/day aspirin. In both cohorts whole blood RNA was collected into PAXgene® Blood RNA tubes (Becton, Dickinson , NJ, USA) before after aspirin exposure and stored at -80 C until microarray profiling. Platelet count was measured in platelet rich plasma in HV1.

Because three subjects in HV2 had participated in HV1, these were dropped from HV2, leaving 50 unique HV2 subjects. DUMC IRB approved the study protocols.

Platelet Function Outcomes in Patients At Risk For Cardiovascular Events At George Washington University (GWU)

We previously described (9) an outpatient cardiology cohort (OPC, Supplemental Methods, Figure 1) treated with 81mg/day aspirin assessed with the VerifyNow Aspirin device and whole blood RNA microarray analysis.

Clinical Outcomes in DUMC Patients

CATHGEN biorepository

The Catheterization Genetics (CATHGEN) biorepository has banked, whole blood RNA in PAXgene® tubes from DUMC patients from the time of cardiac catheterization, baseline medical history, and follow up for all-cause death and MI.(10,11) Two cohorts had available microarray data (Supplemental Methods, Figure 2):

Observational cohort: 224 sequential samples were selected for RNA analysis, of which, 191 had sufficient RNA for microarray analysis.

Case:control cohort: A nested case:control cohort of participants who had experienced death or MI (n = 250) after their index catheterization and age-, sex-, and race-matched controls (n = 250) who were free of death/MI > 2 years after cardiac catheterization was identified.(12) 447 had

sufficient RNA for microarray analysis; 44 overlapped with the observational cohort and were dropped, leaving 403 subjects for analysis.

Follow-up for death/MI was ascertained in both cohorts in October 2011; the median follow-up was 3.8 years. Patients with incomplete follow-up were censored at the time of last contact. Patients who had a history of cardiac transplantation at the time of catheterization (n =5), died within seven days (n =1), or failed quality control (n =1) were excluded. The remaining datasets left 190 samples in the observational cohort (48 death/MI events) and 397 (202 death/MI events) in the case-control cohort.

RNA extraction, labeling, microarray hybridization, quality control, and normalization

See Supplementary Methods for full details. Two microarray platforms were utilized: Affymetrix U133A2 array (HV-1, pre-aspirin) and U133 plus 2.0 array (all others). The Robust Multichip Average (RMA) method was used for normalization.

Real-Time PCR

See Supplementary Methods. Forty-five transcripts were selected for verification in the original RNA samples based on two criteria: 1) the strength of correlation of the probe set with PFS and 2) the strength of membership between the probe set and the set of co-expressed genes of interest.

Platelet purification, Protein Sample Preparation, and Proteomics Analysis by LC-MS/MS

See Supplemental Methods.

Statistical Analysis

The raw and normalized microarray data are available in the Gene Expression Omnibus for the OPC cohort (GSE38511). The data for the HV1, HV2, and CATHGEN cohorts is available through the database of Genotypes and Phenotypes (phs000548.v1.p1 and

phs000551.v1.p1). Unless stated otherwise, all tests were two-sided and were performed in R (2.10.0) or Matlab (R2010b); a p-value of < 0.05 was considered significant.

Discovery of coexpressed gene sets associated with PFS - Factor Modeling

The HV1, post-aspirin RMA normalized data were nonspecifically filtered (i.e., without regard to PFS) to remove probes with mean expression less than 2.0 (i.e., the gene was not expressed in whole blood) or with variance less than 0.25 (i.e., the gene was homogenously expressed), resulting in 2,929 probe sets for subsequent analysis. To discover "Factors" or sets of coexpressed genes representative of biological pathways, we used Bayesian factor regression modeling (BFRM, http://www.isds.duke.edu/research/software/west/bfrm/)(13,14) in an unsupervised fashion (i.e., without regard to PFS). Each of the probe sets used to estimate a particular Factor can be interpreted as a measurement of the activity of some (potentially unknown) biological pathway. Each sample can then be assigned a "Factor score", which represents the aggregate expression of the transcripts within a Factor. The Factor scores can then be used for association with the phenotype of interest in subsequent analyses.

Factor projection, Gene membership within a Factor, Comparison of factor gene lists with selected gene sets, and Co-expression of transcripts represented by a Factor before and after aspirin exposure

See Supplemental Methods

Correlations between factor scores and platelet function

Pearson correlation was used to test for association between a Factor and PFS in HV1 and HV2. In the second validation cohort, OPC, we chose a one-sided t-test because we hypothesized a *lower* factor score in the aspirin-resistant vs. aspirin-sensitive groups.

Linear regression was used to assess the independent association of Factor scores and PFS after accounting for log-transformed MPV and/or platelet count.

Correction for multiple hypotheses testing

As HV1 was a hypothesis-generating pilot study we did not adjust p-values. In the first validation cohort, HV2, we adjusted p-values using Bonferonni correction. In the second validation cohort, we performed only one hypothesis test.

Analyses of RT-PCR data

The expression of each selected transcript relative to the three reference genes was expressed as Δ Cq (or "deltaCq", See Supplemental Methods) and correlated with the corresponding microarray probe set or platelet function score using Pearson tests of correlation.

Platelet proteomic dataset analysis

See Supplemental Methods.

Analyses of CATHGEN cohorts

Logistic or Cox proportional hazards regression models were created in the case:control or observational cohorts, respectively, to test for association between the Factor and death/MI. Each model tested the Factor alone as well as after controlling for baseline variables (Supplemental Data, Table 6) associated with the Factor of interest. The assumption of proportional hazards for each Cox model was met. Odds (or hazards) ratios, 95% confidence intervals, and p-values are reported.

To assess the independent association between the Factor and death/MI, logistic regression models were built on the combined CATHGEN cohorts by forcing Framingham risk factors (age, sex, smoking, diabetes, hypertension, hyperlipidemia), African-American [AA] race, cohort, platelet count, and the presence of CAD (defined as a CAD index(15) >32 or

history of coronary artery bypass surgery/MI/percutaneous coronary intervention), into the model and adding the Factor score or individual probe set gene expression. To assess the incremental prognostic value of gene expression we compared the performance of competing models (risk factors ± Factor/probe set expression), using the areas under the receiver operating characteristics curve (ROC) (16), the net reclassification index (NRI, using risk categories of < 10%, 10-20%, or > 20%(17) or category-free NRI (18), and the integrated discrimination improvement [IDI] (17)).

Results

Discovery and validation of a set of co-expressed genes in whole blood that correlate with platelet function on aspirin

In the discovery cohort (HV1) we identified 20 Factors (numbered 1 - 20, Supplemental Data, Table 1) representing sets of highly correlated, co-expressed genes. To test the hypothesis that one or more of these gene sets were associated with PFS on aspirin, we correlated each set with PFS in HV1 and identified "Factor 14" (Figure 1A) and "Factor 3" ($r = 0.27$, p-value = 0.05). In the first validation cohort (HV2), we found a significant association between Factor 14 and PFS, with the same strength and direction as observed in HV1 (Figure 1B, Bonferroni adjusted p-value = 0.03), thus validating this association, however Factor 3 was not associated with PFS in HV2. We further validated Factor 14 with VerifyNow test results in the OPC cohort (Figure 2). Thus Factor 14, which we named the "aspirin response signature" (ARS), was validated in two independent cohorts as a set of co-expressed genes associated with platelet function on aspirin.

To verify the microarray-based expression of the ARS transcripts, we selected 45 of the 60 genes (see Methods for selection criteria) for verification in whole blood RNA from the HV2

cohort. Using RT-PCR, 42/45 transcripts significantly correlated with their microarray-based expression with 16/42 transcripts, including *ITGA2B, TREML1, MYL9,* and *MPL,* strongly ($r > 0.80$) correlating with microarray based gene expression (Figure 3 and Supplemental Data, Table 2). For the majority of transcripts there was concordance between both the RT-PCR and microarray correlations with PFS (Supplemental Data, Table 3 and Figure 1). Therefore, RT-PCR assays validate the microarray-based expression associations with PFS for most ARS transcripts.

Aspirin response signature transcripts are primarily of platelet origin

We observed that the transcripts with the strongest correlation with PFS (Table 1) mapped to several well-known platelet transcripts: *ITGA2B, CLU, IGF2BP3, GP1BB,* and *SPARC*. Based on this observation we hypothesized that transcripts represented by the ARS were of platelet origin. To test this hypothesis, we examined the overlap and enrichment of the 60 genes represented by the ARS with pre-defined gene sets specific to various peripheral blood cell types. Up to 24 of the 60 ARS genes significantly overlapped with platelet- or megakaryocyte-specific genes, whereas none overlapped with non-platelet peripheral blood cell type genes (Supplemental Data, Tables 4 and 5). Further, in the CATHGEN cohorts, we found the strongest correlation between expression of the ARS and platelet count ($r = 0.41$, $p < 2e^{-16}$) with no strong, positive correlations with any other peripheral blood cell type counts: white blood cells ($r = -0.01$, $p = 0.87$), lymphocytes ($r = -0.25$, $p = 1.2e\text{-}05$), neutrophils ($r = 0.16$, $p = 0.01$), or monocytes ($r = 0.06$, $p = 0.27$).

To confirm the platelet origin of the ARS genes, we analyzed purified platelet lysates by label-free proteomics in the HV2 cohort. We identified 17 proteins from the ARS gene set in the proteomics dataset, of which, six were associated with PFS including *ITGA2B, ITGB3,* and

MYL9 (Table 2), all in the same direction their corresponding transcripts. Therefore, from these data we conclude that a large number of ARS transcripts originate in platelets and are thus reporting on a coexpressed pathway of platelet transcripts and proteins associated with platelet function on aspirin.

Because mean platelet volume (MPV) is associated with platelet function (19) and the platelet origin of ARS transcripts, we assessed the extent to which the association between ARS and PFS was confounded by platelet volume or count. After controlling for MPV, the ARS remained significantly (adjusted regression coefficient for ARS = -0.5, standard error = 0.2, and p-value = 0.05 for HV1; and -0.87, 0.4, and p-value = 0.03 for HV2) associated with PFS. Further, in HV1, where platelet count and volume were both measured, the ARS remained significantly (-0.5 ± 0.2, p = 0.04) associated with PFS after their inclusion. Therefore, the association between ARS and platelet function is independent of other readily available platelet parameters such as count and MPV.

Prior to the administration of aspirin, the aspirin response signature is <u>not</u> associated with platelet function

Because pre-aspirin platelet function is a strong predictor of post-aspirin platelet function(8), we tested the hypothesis that the aggregate expression of the ARS genes was correlated with native, pre-aspirin PFS. In neither HV1 nor HV2 did we observe a correlation between the ARS and pre-aspirin PFS (Figure 3). Despite the absence of a correlation with PFS prior to aspirin, the ARS genes were similarly co-expressed before and after aspirin exposure (Supplemental Data, Figure 2). Therefore, although the set of ARS genes are highly correlated with one another prior to aspirin exposure, their aggregate expression does not appear to

contribute to native, pre-aspirin platelet function. Instead, the expression of the ARS genes specifically reflects platelet function *on aspirin*.

The aspirin response signature is an independent prognostic biomarker for cardiovascular events

Because of the association of the ARS with platelet function on aspirin and aspirin's role in preventing cardiovascular events, we tested the hypothesis that the ARS was associated with the risk of death/MI in two independent patient cohorts. In both case-control and observational cohorts, the ARS was significantly associated with death/MI in univariate analyses (odds ratio [OR] =1.2, 95% confidence interval [CI] =1.04-1.4, p =0.04 and hazard ratio [HR] =1.4, [CI] = 1.1-1.7, p =0.002, respectively). The majority of the individual transcripts represented by the ARS were also associated with death/MI in both cohorts. (Supplemental Data, Table 7)

To determine the extent to which the ARS or an individual probe set for *ITGA2B* was an independent prognostic biomarker for events, we combined the CATHGEN cohorts and found that the ARS (OR =1.3, CI = [1.1, 1.5], p =0.001) or the microarray-based expression of *ITGA2B* (probe set = 206494_s_at, OR = 1.5, CI = [1.2, 1.8], p = 0.0001) were independently associated with death/MI after adjustment for Framingham risk factors(20), race, platelet count, and presence of angiographic CAD.

To further assess the potential use of the ARS as a risk biomarker we tested the hypothesis that the ARS or *ITGA2B* probe set expression would improve measures of discrimination. Compared with a model using clinical risk factors alone, the inclusion of the ARS improved most measures of risk discrimination (Table 3, Figure 5A). Inclusion of *ITGA2B* probe set expression significantly improved all measures of discrimination (Table 3, Figure 5B).

Thus, the ARS or the expression of an individual ARS transcript such as *ITGA2B* were independent prognostic biomarkers for risk of death/MI.

Discussion

We used aspirin as a probe to identify novel genes and biomarkers associated with platelet function and cardiovascular events. We hypothesized that administering aspirin while simultaneously assaying the blood transcriptome might identify sets of genes that are related to aspirin's cardioprotective effect. We identified a set of platelet-enriched, co-expressed genes and proteins, the "ARS", that was reproducibly associated with platelet function in response to aspirin. When tested as a prognostic biomarker, the ARS or an individual ARS transcript (e.g., *ITGA2B*), *independently* and *incrementally* predicted the risk of death/MI compared with traditional risk factors. Our data shows that 1) the genomic response to a pharmacologic "challenge" with aspirin can reveal genes that underlie platelet function on aspirin and mechanisms responsible for death/MI and 2) that whole blood RNA profiling may identify novel biomarkers that discriminate individuals at heightened risk for death/MI.

Transcripts associated with platelet function on aspirin are associated with cardiovascular events.

We found neither association between the ARS and the presence of CAD, nor overlap between ARS genes those previously associated with CAD.(1,6) Instead, we found that the ARS was associated with death/MI after controlling for CAD and CAD risk markers. These findings highlight a unique and novel role that the biologic pathway represented by ARS genes have in the development of cardiovascular events, independent of CAD. We conclude that the biology of aspirin is complex and involves additional mechanisms beyond inhibiting platelet COX-1 and some of these mechanisms underlie risk for cardiovascular events.

A novel and translatable biomarker of platelet function in response to aspirin and the risk for cardiovascular events

Clinicians currently need a readily available biomarker for the response to aspirin. Despite the availability of platelet function assays, their widespread use is severely constrained by the need for specialized equipment and trained personnel. Point-of-care tests are available, but require testing to be completed within hours of phlebotomy; thus, they are out of reach for the vast majority of outpatients on aspirin. Further, most patients taking aspirin for chronic prevention are outpatients where results at the point-of-care are not required. Instead, testing in central laboratories, as is common for LDLc for statins, would be sufficient for determining aspirin response in the outpatient setting. Because of the coexpressed nature of the ARS genes, several individual transcripts (Table 1) correlated best with platelet function. We demonstrated that PCR for individual transcripts could be used in lieu of microarrays (Figure 3 and Supplemental Data, Table 2) for many ARS genes, thus demonstrating the feasibility of a blood-based diagnostic test.

Whole blood RNA testing is a well-established testing diagnostic testing platform. For cardiac allograft rejection and CAD diagnosis, whole blood microarray analyses were both transitioned to a PCR-based platform(6,7): AlloMap® and Corus® CAD, respectively. AlloMap® has been approved by the FDA and both are covered by major insurances. Therefore, there is a feasible path for blood-based RNA biomarkers to clinical adoption, FDA approval and insurance coverage.

Peripheral blood gene expression profiling reveals co-expressed transcripts of platelet origin associated with platelet function in response to aspirin.

The genes underlying variable platelet function on aspirin have been difficult to identify(21) or explain a small portion of the observed variability(22). We hypothesized that whole blood RNA profiling, which *de facto* contains platelet transcripts, would yield biological pathways important for the response to aspirin. We demonstrated that the transcripts represented by the ARS are likely of platelet origin (Supplemental Data, Tables 4 and 5). When we analyzed platelet-enriched protein, we not only confirmed the well-known roles of *ITGA2B* and *ITGB3*, but also and ascribe new roles many other platelet genes: *MYL9, CLU, PPKAR2B, TREML1*, and *CTTN* with respect to platelet function on aspirin and cardiovascular events. Additionally, recent genome wide association studies identified a *PEAR1* polymorphism associated with platelet *PEAR1* levels and platelet function on aspirin.(22) We excluded the probe set (228618_at) mapping to *PEAR1* because its variance (0.21) fell below our variance criteria (0.25, see Methods). However, in a *post hoc* analysis, *PEAR1* expression strongly correlated (r = 0.9) with ARS levels. Therefore, our approach identified previously known and novel platelet genes associated with platelet function in response to aspirin.

We observed an association between ARS and platelet function only after the administration of aspirin, suggesting that the latent effect of ARS genes on platelet function is unmasked in response to aspirin. Consistent with these findings, when we stratified the CATHGEN cohort by aspirin use, we observed that the association between the ARS and death/MI was higher in those using aspirin at the time of catheterization (OR = 1.4 vs. 1.1 in aspirin users vs. nonusers). We hypothesize that the molecular mechanisms represented by the ARS contributes minimally to native platelet function in the absence of aspirin. In contrast, when platelet COX-1, a protein not represented by the ARS, is suppressed by 325mg/day aspirin dosing(23), the effects of these platelet enriched genes is revealed such that the resulting level of

platelet function is then determined by the ARS. Alternatively, aspirin exposure may alter the genomic and protein content of circulating platelets. The precise mechanism by which platelet function on aspirin is related to the expression of the ARS genes and proteins on aspirin is the subject of ongoing work.

Limitations

Several limitations deserve consideration. Neither platelet function nor mean platelet volume (MPV) were measured in CATHGEN. Therefore, we cannot know whether heightened ARS levels altered platelet function or volumes in addition to an increased risk of death/MI. To our knowledge, large cohorts with platelet function, banked RNA, longitudinal follow up, and a sufficient number of events are not available. Further, in our discovery and validation cohorts, the association of the ARS with PFS was independent of platelet count and MPV, suggesting the ARS provides an independent parameter of platelet function that underlies cardiovascular events. Second, although there was no association between the ARS and modifiable risk factors (e.g. diabetes, hyperlipidemia, or hypertension), because we did not assess the degree to which these risk factors were controlled we do not know if addressing these risk factors could modulate ARS levels. Finally, the comparison of the ARS gene set with that of platelets, megakaryocytes, and platelet proteomics analyses demonstrate that the top ARS genes correlative of platelet function on aspirin were of platelet origin. However, some ARS genes (e.g. *TTC7B* and *FSTL1*) are also expressed in non-platelet cell types, suggesting that mechanism(s) represented by ARS genes may involve more than just platelets.

Conclusion

In summary, we used aspirin as a probe in conjunction with RNA profiling and identified novel biomarkers that identify individuals at highest risk for death/MI independent of clinical risk factors.

References

1. Wingrove JA, Daniels SE, Sehnert AJ, et al. Correlation of Peripheral-Blood Gene Expression With the Extent of Coronary Artery Stenosis. Circ Cardiovasc Genet 2008;1:31-38.
2. Helgadottir A, Thorleifsson G, Manolescu A, et al. A Common Variant on Chromosome 9p21 Affects the Risk of Myocardial Infarction. Science 2007;316:1491-1493.
3. Samani NJ, Erdmann J, Hall AS, et al. Genomewide association analysis of coronary artery disease. N Engl J Med 2007;357:443-53.
4. Reilly MP, Li M, He J, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. Lancet 2011;377:383-92.
5. Frelinger AL, III, Li Y, Linden MD, et al. Association of Cyclooxygenase-1-Dependent and -Independent Platelet Function Assays With Adverse Clinical Outcomes in Aspirin-Treated Patients Presenting for Cardiac Catheterization. Circulation 2009;120:2586-2596.
6. Rosenberg S, Elashoff MR, Beineke P, et al. Multicenter Validation of the Diagnostic Accuracy of a Blood-Based Gene Expression Test for Assessing Obstructive Coronary Artery Disease in Nondiabetic Patients. Annals of Internal Medicine 2010;153:425-434.
7. Pham MX, Teuteberg JJ, Kfoury AG, et al. Gene-Expression Profiling for Rejection Surveillance after Cardiac Transplantation. New England Journal of Medicine 2010;362:1890-1900.
8. Voora D, Ortel TL, Lucas JE, Chi JT, Becker RC, Ginsburg GS. Time-dependent changes in non-COX-1-dependent platelet function with daily aspirin therapy. Journal of Thrombosis and Thrombolysis 2012;33:246-57.

9. Fallahi P, Katz R, Toma I, et al. Aspirin insensitive thrombophilia: Transcript profiling of blood identifies platelet abnormalities and HLA restriction. Gene 2013;Epub ahead of print.

10. Wang L, Hauser ER, Shah SH, et al. Peakwide mapping on chromosome 3q13 identifies the kalirin gene as a novel candidate gene for coronary artery disease. American journal of human genetics 2007;80:650-63.

11. Rosati RA, McNeer JF, Starmer CF, Mittler BS, Morris JJ, Jr., Wallace AG. A New Information System for Medical Practice. Arch Intern Med 1975;135:1017-1024.

12. Shah SH, Granger CB, Hauser ER, et al. Reclassification of cardiovascular risk using integrated clinical and molecular biosignatures: Design of and rationale for the Measurement to Understand the Reclassification of Disease of Cabarrus and Kannapolis (MURDOCK) Horizon 1 Cardiovascular Disease Study. American Heart Journal 2010;160:371-379 e2.

13. Wang Q, Carvalho C, Lucas JE, West M. BFRM: Bayesian factor regression modeling. Bulletin of the International Society of Bayesian Analysis 2007;14:4-5.

14. Carvalho CM, Chang J, Lucas JE, Nevins JR, Wang Q, West M. High-Dimensional Sparse Factor Modeling: Applications in Gene Expression Genomics. Journal of the American Statistical Association 2008;103:1438-1456.

15. Mark DB, Nelson CL, Califf RM, et al. Continuing evolution of therapy for coronary artery disease. Initial results from the era of coronary angioplasty. Circulation 1994;89:2015-25.

16. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45.

17. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Statistics in Medicine 2008;27:157-72; discussion 207-12.

18. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. Statistics in Medicine 2011;30:11-21.

19. Thompson CB, Eaton KA, Princiotta SM, Rushin CA, Valeri CR. Size dependent platelet subpopulations: relationship of platelet volume to ultrastructure, enzymatic activity, and function. British Journal of Haematology 1982;50:509-19.

20. D'Agostino RB, Sr., Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA : the journal of the American Medical Association 2001;286:180-7.

21. Mathias R, Kim Y, Sung H, et al. A combined genome-wide linkage and association approach to find susceptibility loci for platelet function phenotypes in European American and African American families with coronary artery disease. BMC Medical Genomics 2010;3:22.

22. Faraday N, Yanek LR, Yang XP, et al. Identification of a specific intronic PEAR1 gene variant associated with greater platelet aggregability and protein expression. Blood 2011;118:3367-3375.

23. Gurbel PA, Bliden KP, DiChiara J, et al. Evaluation of dose-related effects of aspirin on platelet function: results from the Aspirin-Induced Platelet Effect (ASPECT) study. Circulation 2007;115:3156-64.

Figure legends

Figure 1: The aggregate expression of a set of coexpressed, whole blood genes correlates with platelet function in response to aspirin. Two independent cohorts of healthy volunteers were exposed to 325mg/day aspirin, followed by whole blood microarray profiling. Platelet function was assessed by the platelet function score (PFS(8)). The aggregate expression of a set of coexpressed genes (aspirin response signature [ARS], x-axis), is plotted against the PFS (y-axis) after aspirin exposure. Pearson correlation coefficients and p-values are reported.

Figure 2. Aspirin response signature (ARS) is associated with platelet function in patients at risk for cardiovascular disease. Patients treated with 81mg/day aspirin were assessed with the VerifyNow Aspirin device.(9) Three categories of individuals were profiled by microarray based on their aspirin response units (ARU): Aspirin resistant (AR, ARU > 550); High normal (HN, 500 < ARU < 550); Aspirin sensitive (AS, ARU < 550). ARS values are for each group are plotted and compared using two-sample t-tests. P-values are one-sided.

Figure 3: PCR-based assays verify the microarray-based gene expression values for aspirin response signature genes. Real-time PCR assays were designed to verify selected transcripts represented by the aspirin response signature (ARS) in the HV2 cohort. The deltaCq for each assay was correlated with the RMA normalized, probe set expression for the corresponding ARS gene using Pearson correlation (see Supplementary Data, Table 2). For the four genes with the highest PCR vs. microarray-based correlation (*ITGA2B*, *MYL9*, *TREML1*, and *MPL)*, we plot the relative quantity ($2^{-\text{deltaCq}}$, x-axis, log-scale) vs. the corresponding probe set expression (y-axis), correlation coefficient, and p-value.

Figure 4: A set of coexpressed peripheral blood genes does not correlate with native, pre-aspirin platelet function. The aggregate expression of coexpressed genes, is plotted against the platelet function before the administration of aspirin in the discovery cohort (HV1, A, n = 45) and validation cohort (HV2, B, n = 50) healthy volunteers. Pearson correlation coefficients and p-values are reported. ARS = aspirin response signature; PFS = platelet function score.

Figure 5. Peripheral blood gene expression adds additional prognostic information for death or myocardial infarction. Patients in the case:control and observational cohorts were combined and analyzed with respect to death/myocardial infarction (MI) outcomes. The receiver operating characteristics curves were plotted for predictive models containing cardiovascular risk factors, platelet count, presence of coronary artery disease, cohort (collectively, CV) and gene expression, or both were compared. ARS = aspirin response signature. The probe set, 216956_s_at represents *ITGA2B* gene expression.

Table 1. Genes represented by the ARS and their correlation with platelet function with aspirin*				
Affymetrix Probe ID	**Gene Symbol**	**Gene Description**	**Combined PFS beta coefficient ***	**Combined P-value**
Factor 14	n/a	n/a	-0.76088	0.0017
Individual Factor 14 transcripts				
208782_at	*FSTL1*	follistatin-like 1	-1.6579	0.0003
201059_at	*CTTN*	cortactin	-1.2817	0.0015
201906_s_at	*CTDSPL*	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like	-1.3795	0.0025
1555659_a_at	*TREML1*	triggering receptor expressed on myeloid cells-like 1	-1.0767	0.0034
212667_at	*SPARC*	secreted protein, acidic, cysteine-rich (osteonectin)	-1.214	0.0048
216956_s_at	*ITGA2B*	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41)	-1.0689	0.0048

230942_at	*CMTM5*	CKLF-like MARVEL transmembrane domain containing 5	-1.1641	0.0061
57588_at	*SLC24A3*	solute carrier family 24 (sodium/potassium/calcium exchanger), member 3	-1.3053	0.0063
207550_at	*MPL*	myeloproliferative leukemia virus oncogene	-0.931	0.0066
219090_at	*SLC24A3*	solute carrier family 24 (sodium/potassium/calcium exchanger), member 3	-1.1123	0.0080
208791_at	*CLU*	clusterin	-0.9584	0.0085
206494_s_at	*ITGA2B*	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41)	-0.7279	0.0087
227189_at	*CPNE5*	copine V	-1.2062	0.0088
220496_at	*CLEC1B*	C-type lectin domain family 1, member B	-1.2077	0.0090
206493_at	*ITGA2B*	integrin, alpha 2b (platelet glycoprotein IIb of IIb/IIIa complex, antigen CD41)	-0.8966	0.0094
206049_at	*SELP*	selectin P (granule membrane protein 140kDa, antigen CD62)	-1.1642	0.0104

203819_s_at	*IGF2BP3*	insulin-like growth factor 2 mRNA binding protein 3	-1.2947	0.0123
225354_s_at	*SH3BGRL2*	SH3 domain binding glutamic acid-rich protein like 2	-1.0895	0.0146
207808_s_at	*PROS1*	protein S (alpha)	-1.1049	0.0174
207206_s_at	*ALOX12*	arachidonate 12-lipoxygenase	-0.8756	0.0207
212813_at	*JAM3*	junctional adhesion molecule 3	-1.0454	0.0215
1560262_at	*LRRC32*	leucine rich repeat containing 32	-0.9376	0.0226
204628_s_at	*ITGB3*	integrin, beta 3 (platelet glycoprotein IIIa, antigen CD61)	-0.968	0.0242
214146_s_at	*PPBP*	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)	-0.713	0.0243
211026_s_at	*MGLL*	monoglyceride lipase	-1.0027	0.0249
208792_s_at	*CLU*	clusterin	-0.8122	0.0266
201108_s_at	*THBS1*	thrombospondin 1	-0.9169	0.0276
201058_s_at	*MYL9*	myosin, light chain 9, regulatory	-0.5909	0.0287

206390_x_at	*PF4*	platelet factor 4 (chemokine (C-X-C motif) ligand 4)	-0.9017	0.0296
206655_s_at	*GP1BB*	glycoprotein Ib (platelet), beta polypeptide	-0.8934	0.0317
209651_at	*TGFB1I1*	transforming growth factor beta 1 induced transcript 1	-0.7618	0.0326
207414_s_at	*PCSK6*	proprotein convertase subtilisin/kexin type 6	-0.8566	0.0351
200665_s_at	*SPARC*	secreted protein, acidic, cysteine-rich (osteonectin)	-0.8261	0.0410
212077_at	*CALD1*	caldesmon 1	-0.5688	0.0505
203817_at	*GUCY1B3*	guanylate cyclase 1, soluble, beta 3	-0.8511	0.0546
227088_at	*PDE5A*	phosphodiesterase 5A, cGMP-specific	-0.918	0.0571
226152_at	*TTC7B*	tetratricopeptide repeat domain 7B	-0.7986	0.0594
206167_s_at	*ARHGAP6*	Rho GTPase activating protein 6	-0.8437	0.0677
37966_at	*PARVB*	parvin, beta	-0.7708	0.0717
208601_s_at	*TUBB1*	tubulin, beta 1	-0.5959	0.0736
204115_at	*GNG11*	guanine nucleotide binding protein (G protein), gamma 11	-0.5622	0.1229

241133_at	*PRSS1*	protease, serine, 1 (trypsin 1)	-0.4814	0.1243
203680_at	*PRKAR2B*	protein kinase, cAMP-dependent, regulatory, type II, beta	-0.5049	0.1365
205442_at	*MFAP3L*	microfibrillar-associated protein 3-like	-0.4724	0.1385
212151_at	*PBX1*	pre-B-cell leukemia transcription factor 1	-0.6059	0.1729
212573_at	*ENDOD1*	endonuclease domain containing 1	-0.7276	0.1735
230690_at	*TUBB1*	tubulin, beta 1	-0.578	0.1864
230645_at	*FRMD3*	FERM domain containing 3	-0.6391	0.2102
225974_at	*TMEM64*	transmembrane protein 64	0.38321	0.2227
1553842_at	*BEND2*	chromosome X open reading frame 20	-0.5657	0.2258
228708_at	*RAB27B*	RAB27B, member RAS oncogene family	-0.4836	0.2512
227180_at	*ELOVL7*	ELOVL family member 7, elongation of long chain fatty acids (yeast)	-0.3943	0.2823
212148_at	*PBX1*	pre-B-cell leukemia transcription factor 1	-0.3139	0.2970
203414_at	*MMD*	monocyte to macrophage differentiation-associated	-0.4287	0.3236
1552773_at	*CLEC4D*	C-type lectin domain family 4, member D	0.37545	0.3543

222717_at	*SDPR*	serum deprivation response (phosphatidylserine binding protein)	-0.3009	0.3830
224823_at	*MYLK*	myosin, light chain kinase	-0.2911	0.4644
214974_x_at	*CXCL5*	chemokine (C-X-C motif) ligand 5	-0.1621	0.5011
229778_at	*C12ORF39*	chromosome 12 open reading frame 39	-0.2032	0.5020
235331_x_at	*PCGF5*	polycomb group ring finger 5	0.22781	0.5470
212651_at	*RHOBTB1*	Rho-related BTB domain containing 1	-0.2395	0.5755
206110_at	*HIST1H3H*	histone cluster 1, H3h	-0.2021	0.5827
215779_s_at	*HIST1H2BG*	histone cluster 1, H2bg	-0.2623	0.5896
207815_at	*PF4V1*	platelet factor 4 variant 1	-0.0927	0.6139
226188_at	*LGALSL*	lectin, galactoside-binding-like-	0.23728	0.6142
221556_at	*CDC14B*	CDC14 cell division cycle 14 homolog B (S. cerevisiae)	-0.2127	0.6530
207156_at	*HIST1H2AG*	histone cluster 1, H2ag	-0.1534	0.6882
210387_at	*HIST1H2BG*	histone cluster 1, H2bg	-0.1494	0.6906
225166_at	*ARHGAP18*	Rho GTPase activating protein 18	0.15541	0.7353

206272_at	*RAB4A*	RAB4A, member RAS oncogene family	0.09962	0.7967
210986_s_at	*TPM1*	tropomyosin 1 (alpha)	0.08749	0.8404
227451_s_at	*C6ORF79*	chromosome 6 open reading frame 79	-0.0134	0.9791

* = the beta coefficient for the expression of either the aggregate expression of the ARS or each probe set represented by the ARS using the combined HV1 and HV2 datasets from a regression model containing gene expression and cohort (HV1 vs. HV2) with corresponding p-value; PFS = platelet function score.

Table 2. Aspirin response signature proteins identified in platelet protein and their correlations with PFS on aspirin

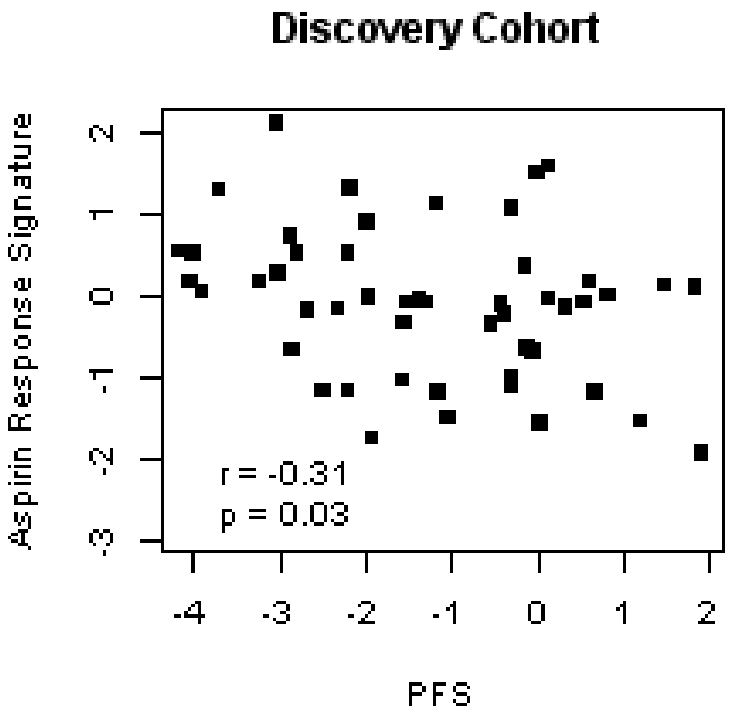
Protein Name	Correlation with PFS	p-value
TBB1	-0.32	0.02
GP1BB	-0.29	0.03
ITA2B	-0.29	0.03
ITB3	-0.28	0.04
MYL9	-0.27	0.05
RB27B	-0.26	0.06
LEGL	-0.24	0.08
TSP1	-0.24	0.08
CALD1	-0.22	0.11
SRC8	-0.21	0.12
SH3L2	-0.20	0.16
CXCL7	-0.20	0.15
SDPR	-0.18	0.19
PLF4	-0.18	0.20
SPRC	-0.14	0.31
PDE5A	-0.09	0.49
CLUS	0.06	0.67

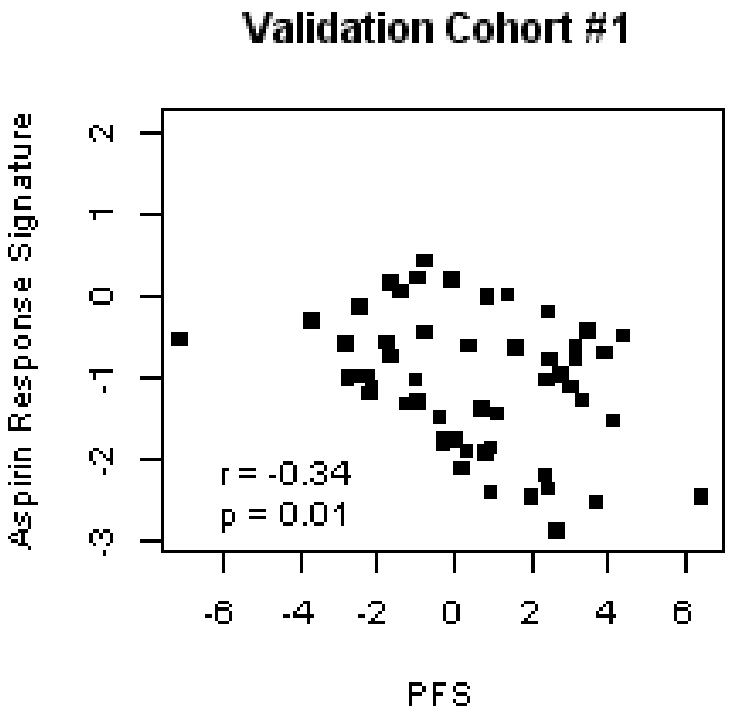
Table 3. Measures of discrimination with and without inclusion of gene expression profiles

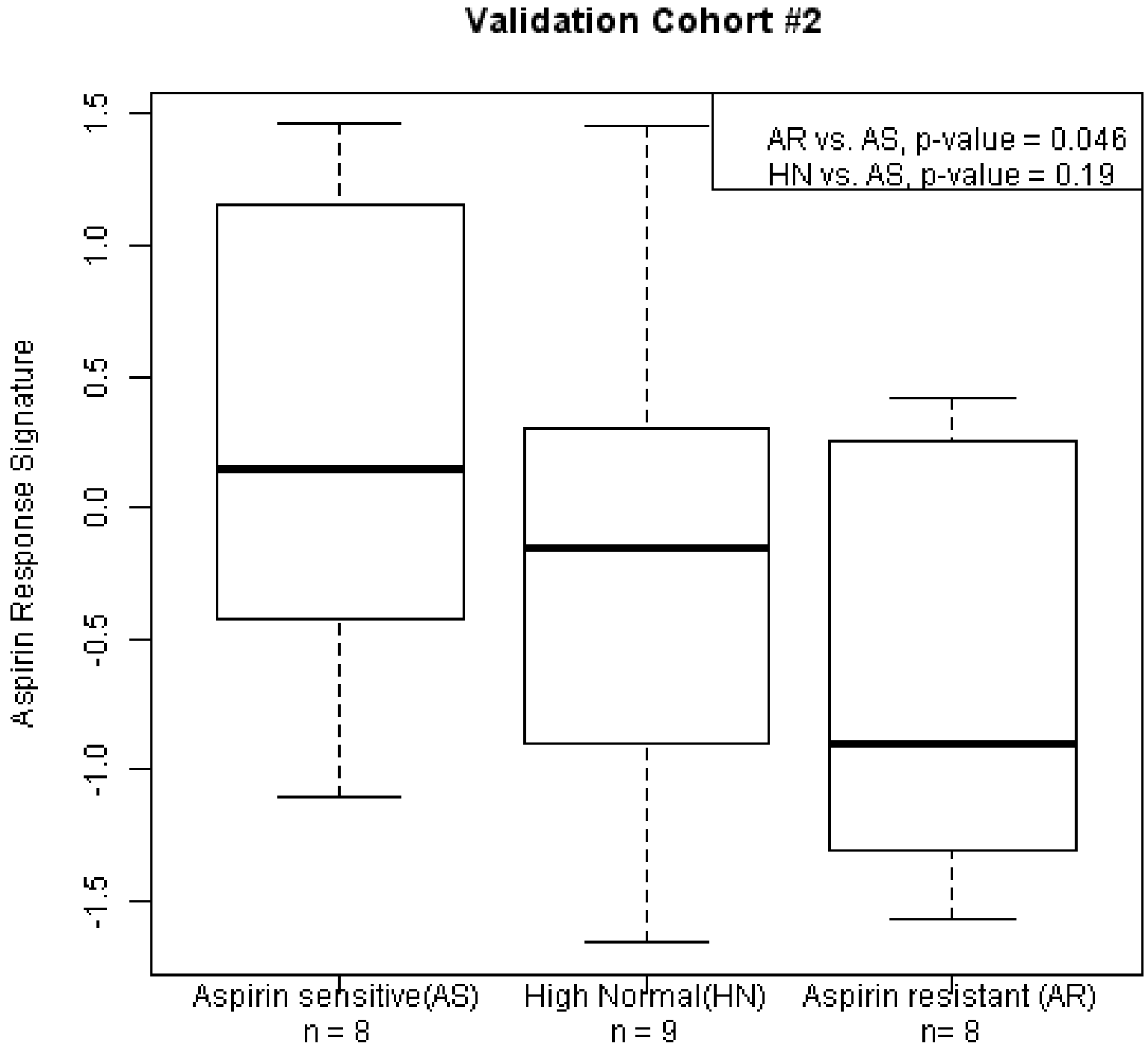
Measure	**Traditional Risk Factors**	**Traditional Risk Factors + ARS**	**Traditional Risk Factors + 216956_s_at** (*ITGA2B*)**
Area under ROC curve 95% confidence interval [CI] p-value*	0.72 [0.68-0.76] n/a	0.73 [0.69-0.77] 0.3	0.74 [0.70-0.78] 0.04
Net reclassification index (<10%, 10-20%, > 20%) CI p-value	-	0.06 [0.02 – 0.10] 0.005	0.12 [0.07 - 0.17] < 1e-05
Net reclassification index (category-free) CI p-value	-	0.31 [0.15 – 0.47] 2e-04	0.37 [0.21 – 0.54] 8.7e-06
Integrated discrimination improvement CI p-value	-	0.01 [0.002 – 0.02] 0.006	0.03 [0.02 – 0.05] 2e-05

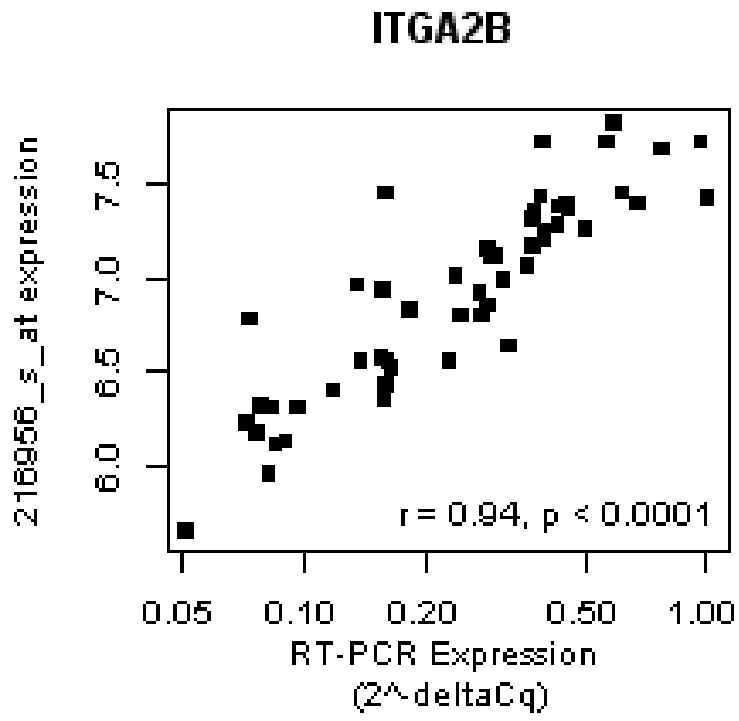
* all p-values are for comparisons with 'Traditional Risk Factors' model which includes: age, sex, African-American race, smoking, diabetes, hypertension, hyperlipidemia, cohort, and the presence of

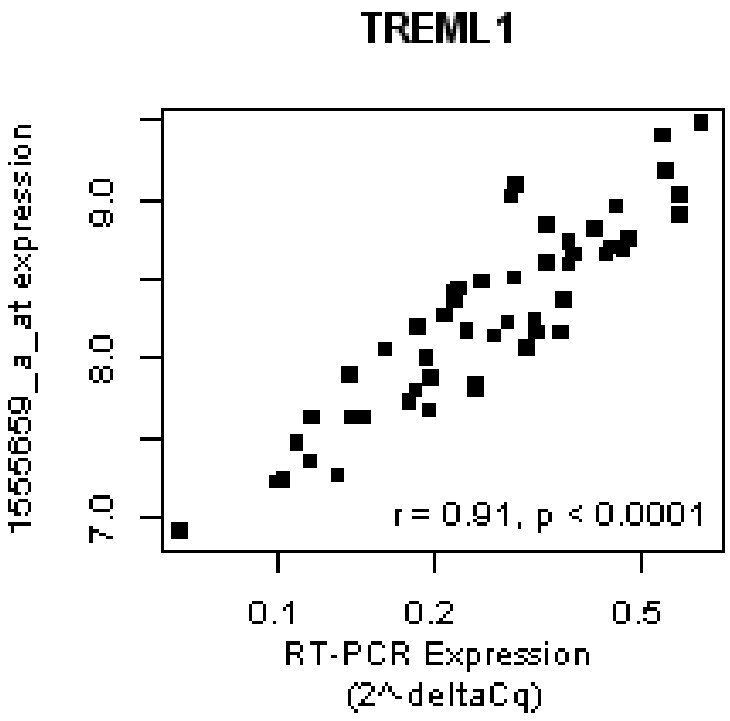
coronary artery disease; **The 216956_s_at probe set represents *ITGA2B* gene expression on the Affymetrix microarray; ARS = aspirin response signature; ROC = receiver operating characteristic

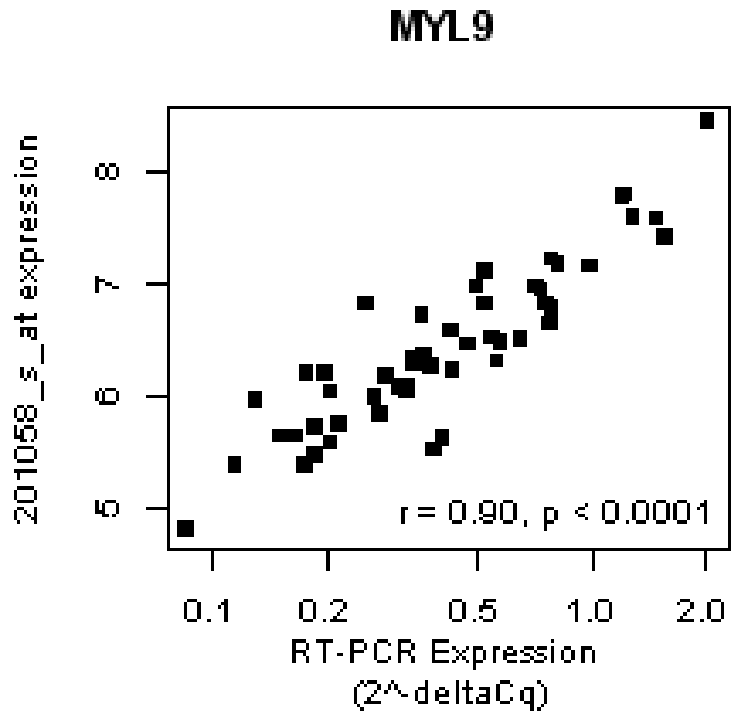

Discovery Cohort
Aspirin Response Signature
PFS
r = -0.31
p = 0.03
-4
-3
-2
-1
0
1
2
-3
-2
-1
0
1
2

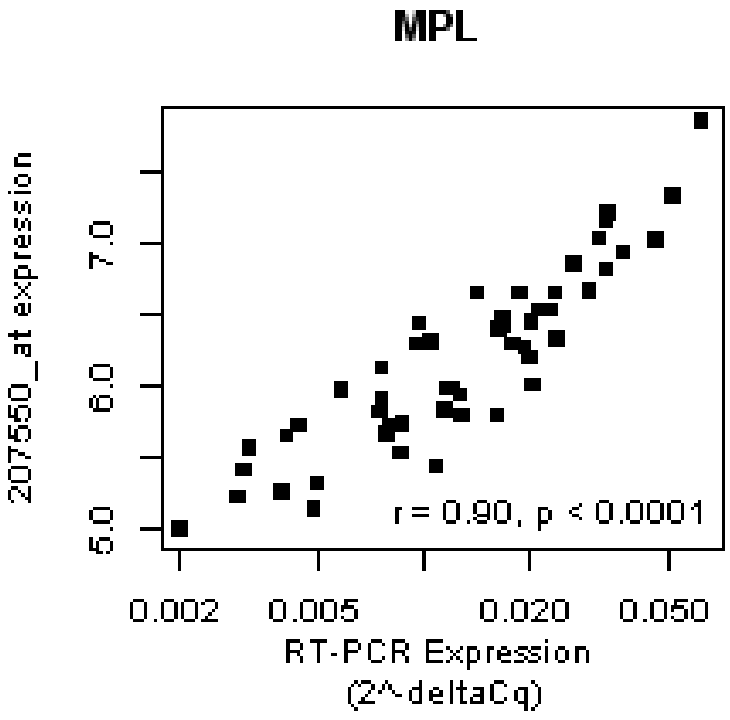

Validation Cohort #1
Aspirin Response Signature
PFS
r = -0.34
p = 0.01
-6
-4
-2
0
2
4
6
-3
-2
-1
0
1
2

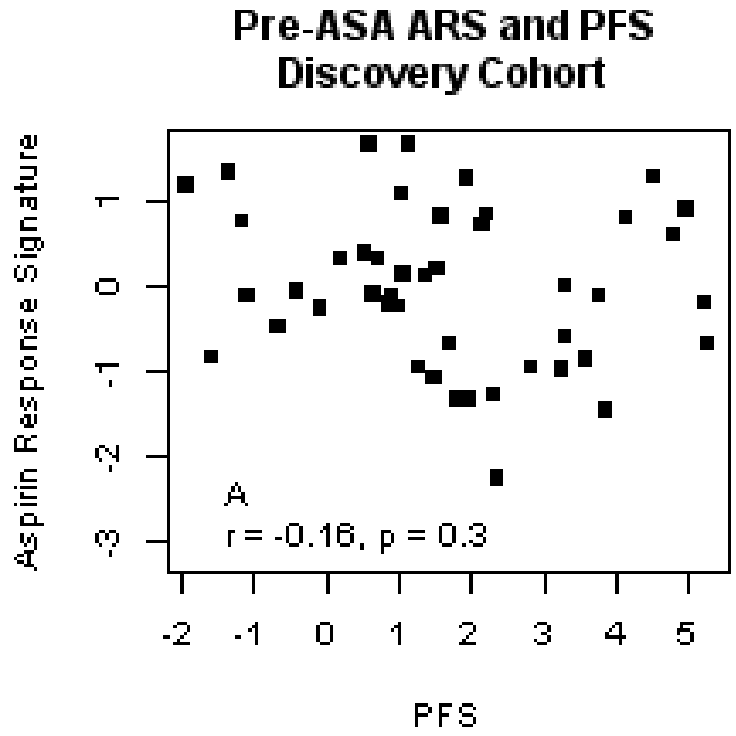

Validation Cohort #2
AR vs. AS, p-value = 0.046
HN vs. AS, p-value = 0.19
Aspirin Response Signature
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
Aspirin sensitive(AS)
n = 8
High Normal(HN)
n = 9
Aspirin resistant (AR)
n= 8

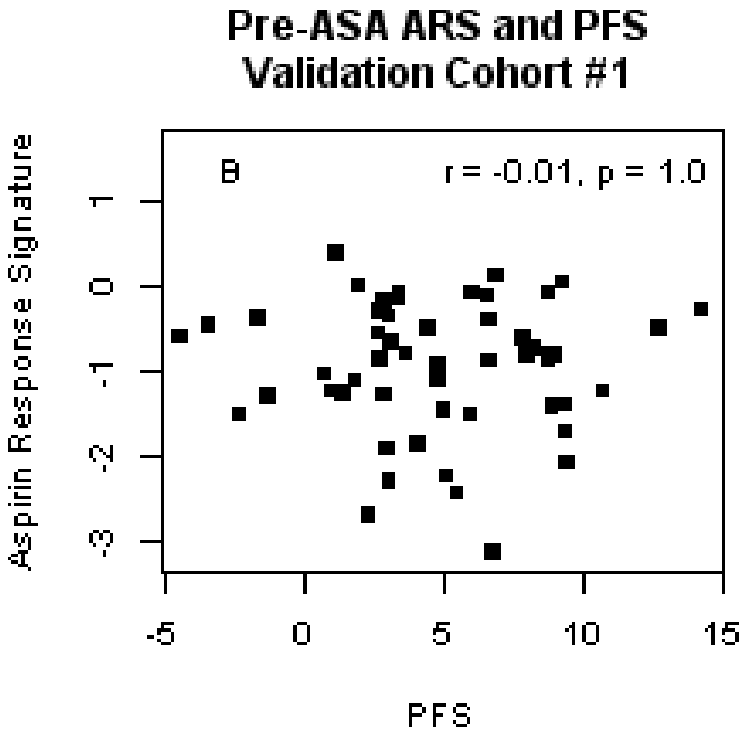

ITGA2B
216956_s_at expression
6.0
6.5
7.0
7.5
r = 0.94, p < 0.0001
0.05
0.10
0.20
0.50
1.00
RT-PCR Expression
(2^-deltaCq)

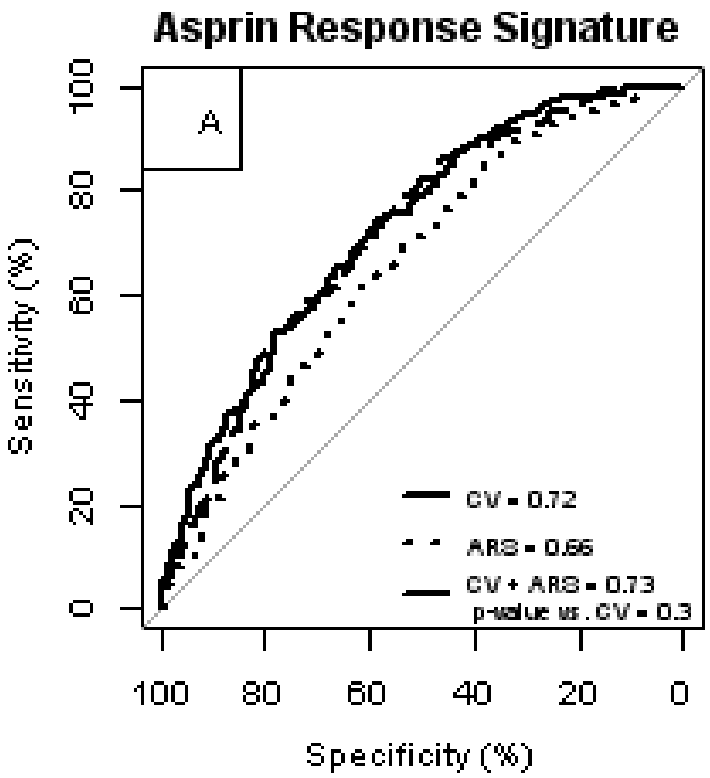

TREML1
1555659_a_at expression
7.0
8.0
9.0
r = 0.91, p < 0.0001
0.1
0.2
0.5
RT-PCR Expression
(2^-deltaCq)

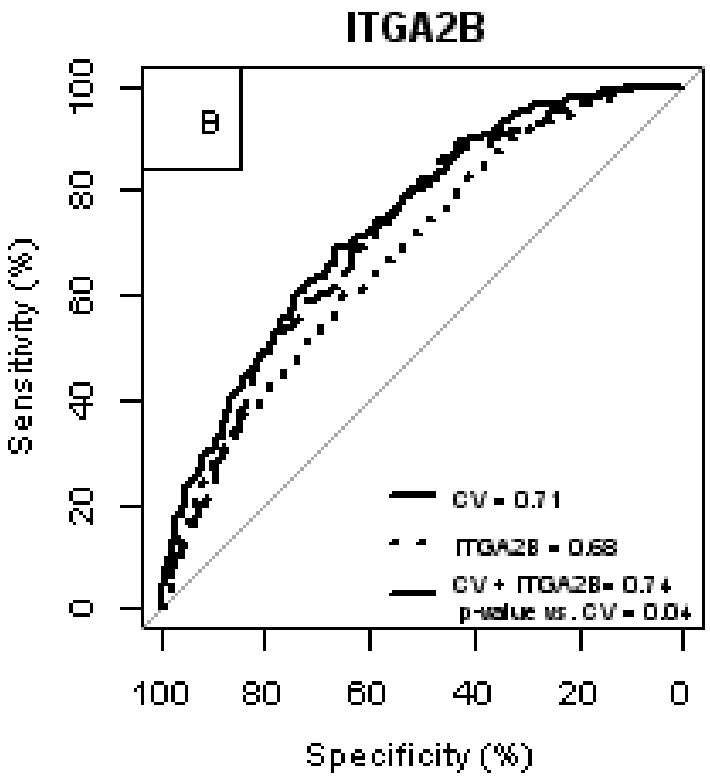

MYL9
201058_s_at expression
5
6
7
8
r = 0.90, p < 0.0001
0.1
0.2
0.5
1.0
2.0
RT-PCR Expression
(2^-deltaCq)


MPL
207550_at expression
5.0
6.0
7.0
r = 0.90, p < 0.0001
0.002
0.005
0.020
0.050
RT-PCR Expression
(2^-deltaCq)


Pre-ASA ARS and PFS
Discovery Cohort
Aspirin Response Signature
PFS
A
r = -0.16, p = 0.3
-3
-2
-1
0
1
-2
-1
0
1
2
3
4
5


Pre-ASA ARS and PFS
Validation Cohort #1
Aspirin Response Signature
PFS
B
r = -0.01, p = 1.0
-3
-2
-1
0
1
-5
0
5
10
15


Asprin Response Signature
A
Sensitivity (%)
Specificity (%)
100
80
60
40
20
0
CV = 0.72
ARS = 0.66
CV + ARS = 0.73
p-value vs. CV = 0.3


ITGA2B
B
Sensitivity (%)
Specificity (%)
100
80
60
40
20
0
CV = 0.71
ITGA2B = 0.68
CV + ITGA2B = 0.74
p-value vs. CV = 0.04

Supplementary Data

Table 1. Factors identified using Bayesian factor regression modeling of peripheral blood gene expression data and their characteristics.

Factor Number	Number of genes within Factor	GO terms	KEGG terms
1	521	protein modification, cellular physiological process, protein transport, establishment of protein localization, protein localization organismal physiological process ubiquitin cycle, G-protein coupled receptor protein signaling pathway neurophysiological process, cell surface receptor linked signal transduction, protein metabolism intracellular protein transport, response to external stimulus, cellular protein metabolism, intracellular transport, cellular macromolecule metabolism, macromolecule metabolism, response to stimulus	NS
2	142	cellular physiological process	NS
3	16	NS	NS
4	58	protein biosynthesis, macromolecule biosynthesis, cellular biosynthesis biosynthesis, cellular protein metabolism, protein metabolism macromolecule metabolism, cellular macromolecule metabolism	oxidative phosphorylation
5	29	response to biotic stimulus, immune response, defense response, deoxyribonucleotide metabolism	NS
6	91	response to biotic stimulus, immune response, defense response, response to stimulus, organismal physiological process, progesterone metabolism	NS
7	120	NS	NS
8	291	immune response, response to external biotic stimulus, response to pest, pathogen or parasite, response to biotic stimulus, defense response, inflammatory response, response to wounding	NS
9	153	NS	NS
10	68	NS	NS
11	11	NS	NS
12	248	cellular physiological process	NS
13	53	gas transport, oxygen transport	NS

14	62	blood coagulation, coagulation hemostasis, regulation of body fluids, platelet activation	NS
15	99	NS	NS
16	32	response to biotic stimulus	Nicotinate and nicotinamide metabolism
17	115	NS	NS
18	49	NS	NS
19	52	NS	NS
20	73	NS	NS
GO = Gene Ontology; KEGG = Kyoto Encyclopedia of Genes and Genomes; NS = no significantly associated terms/pathways (see Methods for significance level)			

Table 2. Validation of microarray based gene expression for Factor 14 probe sets with RT-PCR in HV2 cohort whole blood RNA.

Gene name	Taqman RTPCR assay name	Affymetrix microarray probe set name	Correlation coefficient	Correlation p-value
ITGA2B	Hs01116228_m1	206494_s_at	-0.94421	9.05E-26
TREML1	Hs00698316_m1	1555659_a_at	-0.90545	3.04E-20
ITGA2B	Hs01116228_m1	206493_at	-0.90253	6.28E-20
MPL	Hs00180489_m1	207550_at	-0.89971	1.24E-19
MYL9	Hs00697086_m1	201058_s_at	-0.89826	1.74E-19
PCSK6	Hs00159844_m1	207414_s_at	-0.893	5.76E-19
ITGA2B	Hs01116228_m1	216956_s_at	-0.88851	1.53E-18
CLU	Hs00971656_m1	208791_at	-0.88747	1.90E-18
CLU	Hs00971656_m1	208792_s_at	-0.88153	6.40E-18
SPARC	Hs00277762_m1	200665_s_at	-0.87198	3.95E-17
ALOX12	Hs00167524_m1	207206_s_at	-0.86138	2.53E-16
CMTM5	Hs00370784_m1	230942_at	-0.85545	6.71E-16
SH3BGRL2	Hs00230283_m1	225354_s_at	-0.85496	7.26E-16
CTDSPL	Hs00505109_m1	201906_s_at	-0.84553	3.12E-15
PPBP	Hs00234077_m1	214146_s_at	-0.84171	5.48E-15
PF4	Hs00236998_m1	206390_x_at	-0.83419	1.59E-14
PBX1	Hs00295499_s1	212151_at	-0.82951	3.01E-14
CTTN	Hs01124225_m1	201059_at	-0.82216	7.89E-14
SPARC	Hs00277762_m1	212667_at	-0.80984	3.60E-13
ITGB3	Hs01001469_m1	204628_s_at	-0.79402	2.17E-12
TUBB1	Hs00258236_m1	230690_at	-0.79401	2.17E-12
GNG11	Hs00914578_m1	204115_at	-0.78661	4.78E-12
TGFB1I1	Hs00210887_m1	209651_at	-0.76354	4.63E-11
PBX1	Hs00295499_s1	212148_at	-0.76222	5.23E-11
GP1BB	Hs00236857_m1	206655_s_at	-0.74866	1.75E-10
TUBB1	Hs00258236_m1	208601_s_at	-0.74411	2.59E-10
ELOVL7	Hs00405151_m1	227180_at	-0.72899	8.91E-10
THBS1	Hs00962914_m1	201108_s_at	-0.70496	5.42E-09
SELP	Hs00356351_m1	206049_at	-0.69728	9.30E-09
SLC24A3	Hs00221141_m1	57588_at	-0.65927	1.07E-07
CALD1	Hs00921982_m1	212077_at	-0.64826	2.04E-07
SLC24A3	Hs00221141_m1	219090_at	-0.63347	4.65E-07
MYLK	Hs00364926_m1	224823_at	-0.6135	1.33E-06
ARHGAP6	Hs00241801_m1	206167_s_at	-0.60772	1.77E-06
PRKAR2B	Hs00176966_m1	203680_at	-0.57731	7.45E-06
PDE5A	Hs00903251_m1	227088_at	-0.55683	1.81E-05
FSTL1	Hs00907496_m1	208782_at	-0.53538	4.32E-05
TTC7B	Hs00406077_m1	226152_at	-0.52431	6.62E-05

GUCY1B3	Hs00168336_m1	203817_at	-0.52258	7.07E-05
MMD	Hs00948031_m1	203414_at	-0.39744	0.003528
FRMD3	Hs00604157_m1	230645_at	-0.21862	0.119455
SDPR	Hs00190538_m1	222717_at	-0.18796	0.182072
CPNE5	Hs00326218_m1	227189_at	-0.07961	0.574802
CCDC90A	Hs00254417_m1	227451_s_at	-0.06118	0.666593
CDC14B	Hs00269351_m1	221556_at	-0.04119	0.771874
*coefficient and p-value correspond to Pearson correlation test between RMA normalized microarray expression for a given probe set and delta Ct (where higher values represent lower transcript abundance) for the corresponding RT-PCR assay.				

Table 3. Concordance of correlations with PFS between microarray and RT-PCR based gene expression for Factor 14 genes in HV2 cohort

Gene Name	Affymetrix probe set ID	RT-PCR based Correlation coefficient*	Microarray based Correlation coefficient*
CTDSPL	201906_s_at	0.34	-0.38
FSTL1	208782_at	0.18	-0.38
ITGA2B	216956_s_at	0.22	-0.37
TREML1	1555659_a_at	0.35	-0.35
SPARC	212667_at	0.23	-0.31
ITGA2B	206494_s_at	0.22	-0.30
ITGA2B	206493_at	0.22	-0.30
MPL	207550_at	0.30	-0.29
CTTN	201059_at	0.24	-0.29
CMTM5	230942_at	0.24	-0.29
SELP	206049_at	0.19	-0.28
CLU	208791_at	0.23	-0.27
GP1BB	206655_s_at	0.27	-0.27
ITGB3	204628_s_at	0.30	-0.26
TGFB1I1	209651_at	0.20	-0.25
ALOX12	207206_s_at	0.19	-0.24
PBX1	212151_at	0.26	-0.24
CPNE5	227189_at	0.33	-0.24
MYL9	201058_s_at	0.31	-0.24
CLU	208792_s_at	0.23	-0.22
PPBP	214146_s_at	0.24	-0.22
SH3BGRL2	225354_s_at	0.22	-0.20
PF4	206390_x_at	0.30	-0.20
THBS1	201108_s_at	0.13	-0.20
SLC24A3	57588_at	0.06	-0.20
PBX1	212148_at	0.26	-0.19
SPARC	200665_s_at	0.23	-0.18
GNG11	204115_at	0.24	-0.16
TUBB1	208601_s_at	0.16	-0.16
PCSK6	207414_s_at	0.28	-0.15
SLC24A3	219090_at	0.06	-0.15
PDE5A	227088_at	0.20	-0.14
GUCY1B3	203817_at	0.17	-0.12
CALD1	212077_at	0.13	-0.11
PRKAR2B	203680_at	0.23	-0.10
TTC7B	226152_at	0.00	-0.10
ARHGAP6	206167_s_at	0.20	-0.10
TUBB1	230690_at	0.16	-0.09

ELOVL7	227180_at	0.20	-0.08
MYLK	224823_at	0.22	-0.05
MMD	203414_at	0.21	-0.05
SDPR	222717_at	0.21	-0.05
FRMD3	230645_at	-0.12	0.00
CDC14B	221556_at	0.13	0.18
CCDC90A	227451_s_at	-0.14	0.28

*Correlation coefficients represent Pearson correlations between gene expression based on microarray or RTPCR for Factor 14 transcripts and post-aspirin platelet function score (PFS) in HV2 cohort.

Table 4. Comparison of platelet and megakaryocyte genes with Factor 14

Gene set description*	Number of features	Number of overlapping genes with Factor 14***	GSEA NES	GSEA P-value	GSEA Q-value**	Reference
Factor 14 genes (reference)	62	62	2.23	<0.0001	<0.0001	N/A
Platelet genes (A)	248	25	2.18	<0.0001	<0.0001	[1]
Platelets specific genes (B)	36	12	2.12	<0.0001	0.001	[2]
Megakaryocytes genes (C)	261	20	2.06	<0.0001	0.001	[3]
Platelet genes (D)	196	14	1.97	<0.0001	0.008	[4]
Platelet proteins (E)	99	31	1.84	<0.0001	0.01	[5]
Platelet genes (F)	36	6	1.73	0.004	0.02	[6]

GSEA = Gene Set Enrichment Analysis (GSEA); NES = Normalized enrichment score; *The gene lists were as follows: A) the abundant platelet transcripts identified by RNA-sequencing[1]; B) platelet-specific transcripts identified in platelets of sickle cell disease patients or controls[2]; C) megakaryocyte-specific genes[3]; D) abundant platelet genes identified in platelets of patients with systemic lupus or controls[4]; or E) platelet proteins from healthy donors[5]; F) the top 50 genes identified in purified platelets from healthy volunteers[6]; **refers to false discovery rate q-value;***p ≤ 0.001 for all overlaps.

Table 5. Comparison of Factor 14 genes with non-platelet derived peripheral blood cell-type genes*.

Gene set description*	Number of features	Number of overlapping genes with Factor 14*	GSEA NES	GSEA P-value**	GSEA Q-value***
$CD4^+$ Th lymphocytes	36	0	-0.7	0.7	0.7
$CD8^+$ Tc lymphocytes	4	0	0.7	0.7	0.8
$CD14^+$ Monocytes	164	0	0.6	0.8	0.9
$CD19^+$ B lymphocytes	53		0.5	1.0	0.9
$CD56^+$ NK cells	605	0	1.4	0.2	0.2
$CD66^+$ granulocytes	257	0	1.7	0.03	0.03
Erythroblasts	38	0	1.3	0.2	0.2

*Cell-type specific gene lists were obtained from Watkins et al[3]; Abbreviations and footnotes are per Table 2.

Table 6. Baseline characteristics in CATHGEN cohorts and their association with Factor 14 levels			
Variable	**Case:Control Cohort (n = 397)**	**Observational cohort (n = 190)**	**P-value for association with Factor 14***
Age (years, median, [IQR])	65 [57 -73]	56 [47 -67]	0.15
Hypertension (%)	70.3%	70.5%	0.23
CAD	26%	78%	0.47
Diabetes (%)	34.3%	23.2%	0.14
RACE (%) **White** **African-American** **Other**	 74.3% 21.7% 4.0%	 66.8% 26.8% 6.3%	 Reference 2.13e-06 0.53
Female (%)	32.2%	41.1%	7.37e-06
Smoking history (%)	47.9%	53.7%	0.06
Hyperlipidemia (%)	61.0%	61.6%	0.37
*The combined Case:Control and Observational cohorts were used to test for associations with Factor 14 levels, while controlling for cohort using linear regression.			

Table 7. Factor 14 individual probe sets are associated with death or myocardial infarction in two independent datasets.					
Probe set ID	Gene name	Odds Ratio*	Odds Ratio p-value	Hazard Ratio*	Hazard Ratio p-value
Factor 14	n/a	1.21	0.028036541	1.47	0.001454
Factor 14 individual probe set associations					
201058_s_at	*MYL9*	1.30	0.00400748	1.75	7.65E-06
206494_s_at	*ITGA2B*	1.37	0.00351754	1.83	1.88E-05
216956_s_at	*ITGA2B*	1.52	0.000531982	1.82	3.09E-05
230942_at	*CMTM5*	1.31	0.066298986	2.52	3.50E-05
1555659_a_at	*TREML1*	1.36	0.010689841	1.89	5.63E-05
212573_at	*ENDOD1*	1.94	4.95E-05	2.10	0.000188
212148_at	*PBX1*	1.46	2.45E-06	1.47	0.0002
226188_at	*LGALSL*	1.38	0.071992059	2.62	0.000206
212151_at	*PBX1*	1.65	1.48E-05	1.73	0.000281
206493_at	*ITGA2B*	1.46	0.002887947	1.83	0.000351
210986_s_at	*TPM1*	1.87	1.45E-05	1.83	0.000422
57588_at	*SLC24A3*	1.15	0.440525089	2.19	0.000472
209651_at	*TGFB1I1*	1.32	0.015158949	1.63	0.000616
203819_s_at	*IGF2BP3*	1.45	0.033315085	1.86	0.000619
206049_at	*SELP*	1.46	0.021063226	2.01	0.000636
1552773_at	*CLEC4D*	1.48	0.00175683	1.74	0.000708
226152_at	*TTC7B*	1.43	0.029162426	2.06	0.000931
204628_s_at	*ITGB3*	1.41	0.059389555	1.81	0.00112
208791_at	*CLU*	1.56	0.002006075	1.84	0.001166
206655_s_at	*GP1BB*	1.32	0.043694557	1.69	0.001235
210387_at	*HIST1H2BG*	1.74	0.000124728	1.98	0.001269
201108_s_at	*THBS1*	1.39	0.02328534	1.87	0.001633
208792_s_at	*CLU*	1.51	0.003791609	1.80	0.001869
219090_at	*SLC24A3*	1.16	0.373804241	2.01	0.001907

225974_at	*TMEM64*	0.59	0.008235562	0.41	0.002877
207206_s_at	*ALOX12*	1.21	0.138154505	1.74	0.003564
229778_at	*C12orf39*	1.39	0.004259803	1.58	0.005073
215779_s_at	*HIST1H2BG*	1.68	0.002837678	1.81	0.005208
37966_at	*PARVB*	1.04	0.771921449	1.81	0.005396
225354_s_at	*SH3BGRL2*	1.34	0.049135136	1.79	0.00551
212667_at	*SPARC*	1.25	0.167170138	1.78	0.005604
207414_s_at	*PCSK6*	1.23	0.150039721	1.72	0.006252
235331_x_at	*PCGF5*	1.93	5.27E-06	1.42	0.006562
200665_s_at	*SPARC*	1.44	0.008859476	1.72	0.006808
207808_s_at	*PROS1*	1.45	0.018336331	1.58	0.006933
212077_at	*CALD1*	1.11	0.338890694	1.52	0.006972
201906_s_at	*CTDSPL*	1.30	0.11440241	1.90	0.007299
241133_at	*TRBV27*	1.28	0.019888406	1.52	0.011429
206390_x_at	*PF4*	1.13	0.398156787	1.85	0.0117
211026_s_at	*MGLL*	1.16	0.327461963	1.85	0.012801
230645_at	*FRMD3*	1.85	0.000131807	1.79	0.014483
206110_at	*HIST1H3H*	1.72	3.26E-05	1.57	0.018138
208601_s_at	*TUBB1*	1.16	0.201369942	1.50	0.01831
203680_at	*PRKAR2B*	1.43	0.005278287	1.54	0.025644
227180_at	*ELOVL7*	1.25	0.070697319	1.49	0.032551
214146_s_at	*PPBP*	1.55	0.000423437	1.83	0.037707
227189_at	*CPNE5*	0.71	0.079532372	1.63	0.03795
208782_at	*FSTL1*	1.28	0.086692126	1.53	0.039651
1560262_at	*LRRC32*	1.24	0.155422462	1.55	0.050918
201059_at	*CTTN*	1.16	0.299273772	1.53	0.057153
220496_at	*CLEC1B*	1.50	0.002955397	1.43	0.079647
212651_at	*RHOBTB1*	1.32	0.08218736	1.43	0.096564
204115_at	*GNG11*	1.41	0.004779687	1.45	0.118268
206167_s_at	*ARHGAP6*	1.14	0.422952376	1.36	0.151355

227088_at	*PDE5A*	1.35	0.046937627	1.37	0.151501
203817_at	*GUCY1B3*	1.28	0.098280545	1.35	0.160362
207550_at	*MPL*	1.05	0.717720328	1.28	0.210728
1553842_at	*BEND2*	1.42	0.079659139	1.33	0.233
212813_at	*JAM3*	0.97	0.872083306	1.34	0.240894
230690_at	*TUBB1*	1.09	0.589122948	1.31	0.265404
207156_at	*HIST1H2AG*	1.54	0.004128653	1.27	0.303256
225166_at	*ARHGAP18*	1.53	0.034773026	1.30	0.338505
224823_at	*MYLK*	1.24	0.099105304	1.21	0.357528
227451_s_at	*CCDC90A*	1.07	0.751398281	0.76	0.381367
228708_at	*RAB27B*	1.21	0.257379973	0.84	0.546707
207815_at	*PF4V1*	1.07	0.397575492	1.07	0.581265
221556_at	*CDC14B*	0.78	0.197214947	0.86	0.60266
222717_at	*SDPR*	1.08	0.545883997	1.10	0.679631
206272_at	*RAB4A*	1.54	0.020878373	0.91	0.803266
214974_x_at	*CXCL5*	0.88	0.201323204	0.97	0.867518
205442_at	*MFAP3L*	0.90	0.409405359	1.01	0.968416
203414_at	*MMD*	1.19	0.292053865	1.01	0.985414

*Individual Factor 14 probe sets were associated with death or myocardial infarction in the case control and observational cohorts, yielding odds ratios and hazards ratios, respectively. For each probe set analyses were adjusted for log(platelet count), race, and sex.

Table 8. Net Reclassification Table for combined CATHGEN cohorts				
Factor 14				
	10%	10-20%	>20%	% Reclassified
10%	10	4	0	29
10-20%	4	55	6	15
>20%	0	14	474	3
Net Reclassification Index				4.6% (p = 0 007)
ITGA2B				
	10%	10-20%	>20%	% Reclassified
10%	8	5	1	43
10-20%	13	42	10	35
>20%	1	24	463	5
Net Reclassification Index				7.5% (p = 0.002)