

Prediction of Ischemic Events on the Basis of Transcriptomic and Genomic Profiling in Patients Undergoing Carotid Endarterectomy

Lasse Folkersen,¹ Jonas Persson,^{1,2} Johan Ekstrand,³ Hanna E Agardh,¹ Göran K Hansson,¹ Anders Gabrielsen,¹ Ulf Hedin,³ and Gabrielle Paulsson-Berne¹

¹Center for Molecular Medicine, Department of Medicine, Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden; ²Department of Clinical Sciences, Karolinska Institutet, Danderyd University Hospital, Stockholm, Sweden; and ³Department of Molecular Medicine and Surgery at Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden

Classic risk factors, including age, smoking, serum cholesterol, diabetes and blood pressure, constitute the basis of present risk prediction models but fail to identify all individuals at risk. The objective of this study was to investigate if genomic and transcriptional patterns improve prediction of ischemic events in patients with established carotid artery disease. Genotype and gene expression profiles were obtained from carotid plaque tissue ($n = 126$) and peripheral blood mononuclear cells ($n = 97$) of patients undergoing carotid endarterectomy. Patients were followed for an average of 44 months, and 25 ischemic events occurred (18 ischemic strokes and 7 myocardial infarctions). Blinded leave-one-out cross-validation on Cox regression coefficients was used to assign gene expression-based risk scores to each patient. When compared with classic risk factors, addition of carotid plaque gene expression-based risk score improved the prediction of future ischemic events from an area under the curve (AUC) of 0.66 to an AUC of 0.79. The inclusion of gene expression risk score from peripheral blood mononuclear cells or from 25 established myocardial infarction risk single nucleotide polymorphisms only exhibited marginal effects on the prediction of ischemic events. Prediction of ischemic events is improved by inclusion of gene expression profiling from carotid endarterectomy tissue compared with prediction on the basis of classic risk markers alone in patients with atherosclerosis. The method may be developed to identify subjects at very high risk of ischemic events.

Online address: <http://www.molmed.org>
doi: 10.2119/molmed.2011.00479

INTRODUCTION

Ischemic events contribute substantially to morbidity in patients with atherosclerotic disease despite guideline-based treatment, and it is of interest to identify cardiovascular patients with an excess risk. If novel biomarkers of excess risk of ischemic events in secondary prevention are identified, it would be an important first step in applying individualized preventive measures. Classic risk factors, such as age, gender, smoking, diabetes, hypercholesterolemia and hyper-

tension are well documented and easily identifiable (1), but there might be additional predictive power in other biomarkers. This additional predictive power may be speculated to be coupled to genetic factors and phenotypic gene expression profiles.

Several genome-wide association studies (GWASs) have associated single nucleotide polymorphisms (SNPs) with early-onset myocardial infarction and other phenotypes of cardiovascular disease (2–4). Therefore, in GWAS co-

orts, there is good evidence to support the role of common genetic variation with increased risk of early myocardial infarction. In addition, there have been some inquiry into the efficacy of risk-associated SNPs in predicting outcomes in diseased populations (5), but the role of genotype-based prediction of adverse events in secondary cardiovascular prevention remains poorly explored. It is tempting to speculate that SNPs associated with early myocardial infarction may provide additional information on the risk of ischemic events in patients.

Most gene expression studies in patients with cardiovascular disease have attempted to distinguish current patient phenotypes on the basis of gene expression profile data. For example, investigations have compared circulating blood cell RNA profiles among normotensive and hypertensive patients, ischemic and nonischemic dilated cardiomyopathy pa-

Address correspondence to Lasse Folkersen, Center for Molecular Medicine, CMM L8:03, Karolinska University Hospital Solna, 171 76 Stockholm, Sweden. Phone: +46-851776406; Fax: +46-8313147; E-mail: lasse.folkersen@ki.se.

Submitted December 8, 2011; Accepted for publication February 23, 2012; Epub (www.molmed.org) ahead of print February 24, 2012.

tients, healthy controls and patients with thoracic aortic aneurysm, as well as patients differing in their extent of coronary artery disease (6–10). Other studies have examined diseased tissue biopsied from the site of disease, with a similar aim of identifying disease-specific gene expression (11–15). Although many such disease-specific differences in gene expression profiles have been noted to be present at the time of sampling, it is unknown whether expression patterns predict future prognosis. In the context of a single gene, it was recently demonstrated that expression of fatty acid-binding protein 4 (*FABP4*) in the carotid atherosclerotic plaque (16,17) provides prognostic information. Furthermore, a study of genome-wide gene expression patterns in endomyocardial biopsies from patients with heart failure, comparing individuals who had a 5-year event-free survival with those who had a major cardiovascular event within 2 years, identified 46 genes for which expression levels predicted disease outcome (18). Supporting this notion, studies in cancer patients have demonstrated that expression profiling can be used to predict prognosis. These expression profile methods have identified two distinct forms of diffuse large B-cell lymphoma that were associated with significantly different overall survival rates (19), and several prognostic gene expression signatures have been patented, some of which are approaching clinical application (20). Consequently, we speculate that genome-wide gene expression profiling provide additional information on risk of ischemic events.

In the current study, we hypothesize that combining risk-SNP genotype information and genome-wide gene expression profiling with classic risk markers will provide novel information on the risk of ischemic events in atherosclerotic carotid disease.

MATERIALS AND METHODS

Sample Collection

Materials from the Biobank of Karolinska Endarterectomies (BiKE) were obtained

Table 1. Demographic characteristics of patients.

	Plaque	PBMCs
n	126	97
Age (years)	70.6 ± 8.9	70.8 ± 9.38
Female gender	22.2% (28)	25.8% (25)
Body mass index (kg/m ²)	26.4 ± 3.8	26.5 ± 4.15
LDL (mmol/L)	2.52 ± 0.94	2.46 ± 0.97
Current or previous smoker	49.2% (62)	49.5% (48)
Type 2 diabetes mellitus	25.4% (32)	26.8% (26)
Symptomatic carotid artery disease	67.5% (85)	69.1% (67)
Follow-up time (d)	1,333 ± 728	1,159 ± 631
Ischemic event during follow-up	19.8% (25)	21.6% (21)

Numerical variables are given as mean ± standard deviation. Categorical variables are given as percentage (number of patients). All variables except follow-up time and ischemic event percentage are indicated at baseline.

after all participants provided informed written consent, per the Declaration of Helsinki and the Karolinska Institute ethics committee (file number 02-147 and 2009/295-31/2). The study included 98 peripheral blood mononuclear cell (PBMC) samples and 126 atherosclerotic plaque tissue samples from patients who were undergoing carotid endarterectomy at Karolinska University Hospital (Stockholm, Sweden). The patients were selected for operation according to the criteria in the North American Symptomatic Carotid Endarterectomy Trial (NASCET) study (21). The plaque and PBMC data sets had 97 overlapping samples; accordingly, the latter approximated a subset of the former. An overview of patient characteristics can be seen in Table 1. Expression data have also been described (22,23).

Follow-up and Definition of Ischemic Events

Patients were followed for 1,333 ± 728 d (average ± standard deviation). Sample collection started in October 2002 and ended in March 2009. Last follow-up date was 31 December 2010. Ischemic events were defined as having any of the following conditions reported to the Swedish Hospital Discharge Register or the Swedish Cause of Death Register: acute myocardial infarction (International Classification of Diseases, 10th Revision [ICD-10], codes I21, I22, I23), other

forms of ischemic heart disease (I24.8, I24.9) and cerebral infarction (I63). In addition, reports of angina pectoris (I20) or chronic ischemic heart disease (I25) were considered events when reported to the Swedish Cause of Death Register. Of the ischemic events, 7 were myocardial infarctions and 18 were ischemic strokes.

The retrieval of myocardial infarction and stroke incidence data from the Swedish Hospital Discharge Register and the Swedish Cause of Death Register is a reliable, validated alternative to the use of revised hospital discharge and death certificates (24,25).

Gene Expression Microarrays

Total RNA was isolated using the RNeasy Mini Kit (Qiagen) and treatment with the RNase-free DNase set (Qiagen) according to the manufacturer's instructions. RNA quality was analyzed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and RNA concentration was measured on a NanoDrop (Thermo Scientific). Low-concentration and low-quality samples were omitted.

RNA samples were hybridized to Affymetrix HG-U133 plus 2.0 oligonucleotide arrays and scanned at the Karolinska Institute Bioinformatics and Expression Analysis core facility. The resulting CEL files were preprocessed using robust multiarray average (RMA) (26) normalization, as implemented in the

Affymetrix Power Tools 1.10.2 package apt-probeset-summarize. As part of the RMA normalization, all expression measurements were \log_2 -transformed. Low-expression probe sets for which average expression levels were less than the genome-wide median value were omitted from the analysis. The data set is available from Gene Expression Omnibus (accession number GSE21545).

Genotyping

DNA samples from patients were genotyped using Illumina Human 610W-Quad BeadArray at the SNP technology platform at Uppsala University. The GenomeStudio™ software from Illumina was used for genotype calling and quality control, and the MACH 1.0 algorithm (Center for Statistical Genetics, University of Michigan, Ann Arbor, MI) was used for imputation on the basis of references from the 1,000 genomes project. The average call rate per SNP was 99.84%. Replicate genotyping of 12 samples demonstrated an overall concordance of 99.99%. A total of 29 SNPs previously reported to predict early-onset myocardial infarction were chosen to create a genotype score (2–4). Of the 29 SNPs, 4 were omitted because of unsatisfactory imputation quality (Rsq score >0.3 using the MACH algorithm), Hardy-Weinberg disequilibrium ($P < 0.01$) or frequency mismatch (>0.2 compared with reported frequencies). The reported allele frequencies of risk-SNPs were compared with the measured allele frequencies to ensure the correct choice of risk allele. Further details available are in Supplementary Table 1. Genotype risk score was calculated as the sum of risk alleles in each patient: SNPs that were homozygous for the risk allele were counted as 2, and SNPs that were heterozygous were counted as 1.

Gene Set-Based Expression Prediction and Cross-Validation

A complete leave-one-out cross-validation scheme was used as previously recommended (27): for each iteration, all samples except one were used to select probe sets for which expression

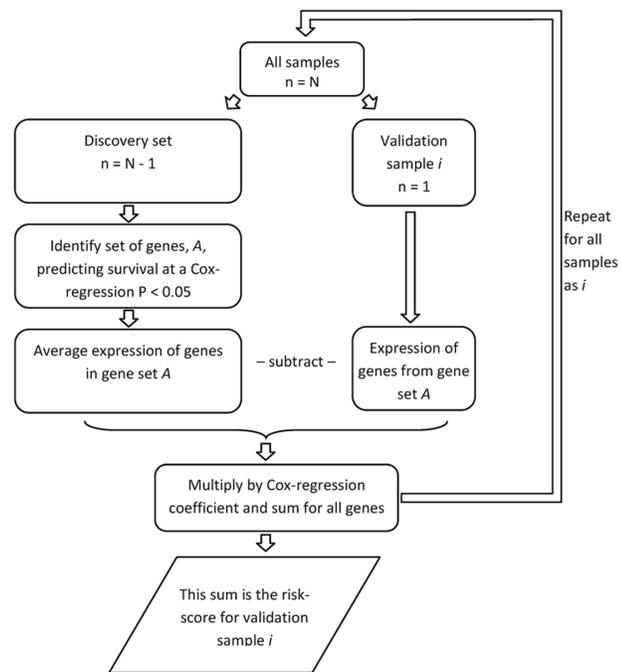


Figure 1. Flow diagram of cross-validation scheme. R-code for this scheme can be found as Supplementary Table 4.

levels were predictive of ischemic events. For each probe set, this step was performed by Cox regression with the `coxph` function from the survival Bioconductor package using default settings. All probe sets that had a Cox regression significance of $P < 0.05$ were selected to predict the risk score of the omitted sample.

Stated more formally, the risk score of each omitted sample (i) is:

$$\text{risk}_i = \sum_{j \in A} \text{coefficient}_j \times (\text{expression}_{ij} - \overline{\text{expression}}_j),$$

where j is probe set, A is the group of probe sets with a Cox regression significance of $P < 0.05$, coefficient_j is the Cox regression coefficient for a probe set, expression_{ij} is the probe set expression in the omitted sample and $\overline{\text{expression}}_j$ is the mean expression of the probe set in the remaining samples. The iteration was repeated until all samples had been assigned a risk score, on the basis of gene selection and gene expression weights, selected independently of the expression profile in the sample. This scheme is shown in Fig-

ure 1. If the predictive probe sets are incorrectly tested in the same data, vastly overestimated prediction values of area under the curve (AUC) >0.9 are observed.

Cox Regression and Calculation of Area under the Curve

Classic clinical risk markers, cross-validated gene expression risk score and risk allele counts of genotypes were used as variables in the overall prediction analysis, as indicated in the results. Similar to the selection of gene expression parameters, regression was performed by `coxph` function from the survival Bioconductor package using default settings. The proportional hazards assumption for a Cox regression model fit was tested using the `cox.zph` function. The assumption holds globally as well as for individual variables in both data sets ($P > 0.05$). Receiver operating characteristic (ROC) curves were plotted on the basis of the Cox regression calculations, and AUCs were calculated using the `risksetROC` R package.

All supplementary materials are available online at www.molmed.org.

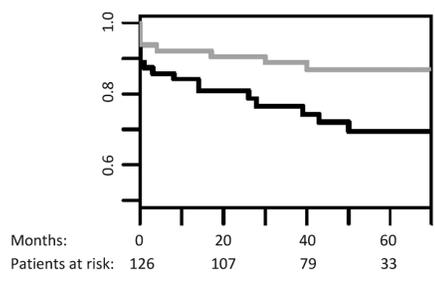


Figure 2. Kaplan-Meier plot of ischemic events stratified by gene expression. The plot shows the fraction of event-free survival as a function of months after operation. Each line represents half of the 126 plaque samples: the gray line denotes the group with expression risk scores below the median, and the black line denotes the group with expression risk scores above the median.

RESULTS

Gene Expression-Based Prediction

Global gene expression profiles of carotid plaque tissue (n = 126) and PBMC samples (n = 97) were obtained from patients undergoing carotid endarterectomy. No individual genes were significantly predictive of future ischemic events at a false discovery rate of 5%. We therefore tested whether risk prediction of ischemic events was improved by the use of gene expression profiles from multiple genes (so-called gene expression signatures).

Lack of suitable validation cohorts prevented us from using a classic discovery and validation cohort setup. We therefore adopted a leave-one-out cross-validation scheme as previously recommended (27) and as used in previous development of cancer diagnostics where gene expression signatures provided more robust predictors of survival (19,20). In this method, a risk score was calculated for each sample by using information from the remaining samples (Figure 1). For plaque expression signatures, this step resulted in risk scores that differed significantly between patients with events and without events (Student *t* test, $P = 1.21 \times 10^{-4}$). For PBMC sam-

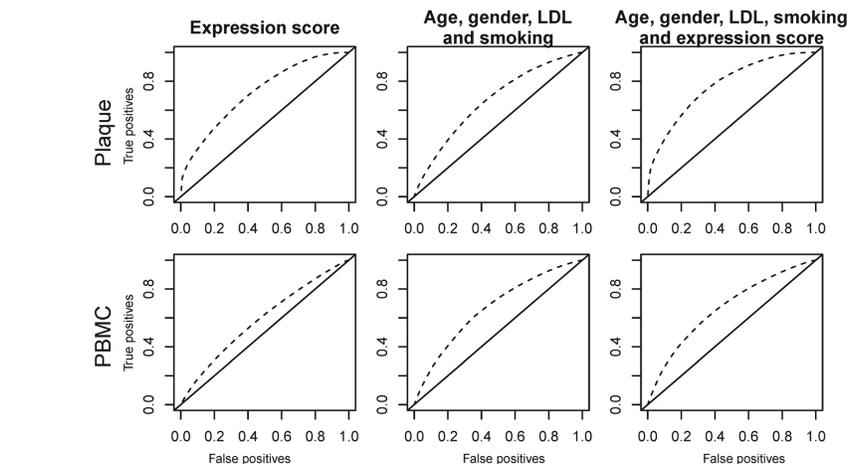


Figure 3. Receiver operating characteristic curves for different risk factors at 300 d. Top plots show data from the plaque data set. Bottom plots show data from the PBMC data set. Plots are calculated only from gene expression profiles (left), only from the established risk factors: age, gender, LDL and smoking (middle), or from both gene expression profile and the four established risk factors (right). The straight diagonal line illustrates prediction by pure chance (AUC 0.5), and the curved line describes the prediction possible with the given predictive variables. A theoretical curve along the top left corner would indicate perfect prediction (AUC 1).

ples, the differences were not significant ($P = 0.145$). A Kaplan-Meier plot of ischemic events as a function of time stratified by the risk scores from plaque samples clearly shows the capacity of this approach (Figure 2).

Risk estimates were then incorporated into prediction models. Figure 3 demonstrates the predictive power of three models: expression signature score only, classic risk factor profiling and the combination of classic risk factor profiling and expression signature score. The area under the ROC curve (AUC) calculated for both data sets and with several models is depicted in Table 2. As shown, plaque expression signatures predicted outcomes better than PBMC expression signatures. Using gene expression risk signatures in addition to classic risk factors improved prediction of ischemic events in all cases. Figures 2 and 3 are based on 300 d of follow-up, but similar effects could be shown at 100 and 30 d.

We further investigated if the independence between the gene expression risk signature and the classic risk factor model was affected by choice of classic risk variables included. Throughout the

text, the classic risk variables were defined as age, gender, smoking and serum low-density lipoprotein (LDL) concentration. However, in the more than 70 other registered clinical variables of medical history, medication, ultrasound measurements and serum concentrations measured for all patients, we observed that diabetes, serum creatinine and calculated glomerular filtration rate had additional borderline univariate predictive properties in this data set. We therefore repeated the analysis including all of these seven clinical variables, with or without the gene expression signature. The change in prognostic strength added by the gene expression signature was negligible: although the classic risk factor AUC increased slightly from 0.66 to 0.69, the combined risk factor AUC increased similarly from 0.79 to 0.83. This result indicates that the gene expression risk signature is independent of established classic risk predictors.

The cross-validated risk prediction score demonstrates the general utility of gene expression signature-based prediction. Because of the cross-validation scheme, however, the actual list of probe

sets that is used for profiling depends on the sample that is omitted. On average, 192 ± 38 probe sets were used to profile each leave-one-out iteration. Further, 18.3% of the probe sets were present in >90% of the iterations, and 66.2% of the probe sets were present in <10% of the iterations. The 146 probe sets that were present in >90% of the iterations are shown in Supplementary Table 2.

Gene Set-Based Prediction with Custom Annotation Files due to Probe Set Redundancy

A total of 32 genes in the list were measured by multiple probe sets in our lists or had no annotation in the latest Bioconductor microarray annotation (Bioconductor 2.8) (28). We therefore examined the use of an updated annotation through custom chip definition files (CDF) (29). This updated annotation contains 18,185 probe sets, one-third of the original Affymetrix annotation, and virtually none of these are duplicate measurements of one gene.

Repeating all calculations on this updated annotation yielded essentially the same predictive score for the plaque profiles (maximum 0.001 change from the numbers in Table 2). However, the number of probe sets used for the prediction in each of the iterations decreased to 62 ± 13 . Although the percentage of probe sets that were found in >90% of the iterations remained the same, the number decreased to 45, of which 17 were also identified by standard CDF analysis. This list of more cleanly annotated updated genes is available as Supplementary Table 3. From these results, we conclude that the expression profile risk prediction is not an artifact of problematic probe set annotation.

Description of Predictive Gene Set

The expression profile of the identified 146 probe sets (common to >90% of the iterations) are the best predictors of ischemic events in the BiKE data set. The gene products of these 146 genes localized to the cytoplasm (13.2%), the nucleus (11.4%), the plasma membrane

Table 2. AUC at 300 d since operation.

	Model	AUC
Plaque data set (n = 126)	Genotype score only	0.55
	Expression score only	0.72
	Classic risk markers	0.66
	Combination	0.79
Peripheral blood mononuclear cells data set (n = 97)	Genotype score only	0.55
	Expression score only	0.59
	Classic risk markers	0.67
	Combination	0.68

Genotype score is on the basis of the risk allele count in 25 previously published myocardial infarction-associated risk variants. The expression score is on the basis of the mRNA expression levels of a set of predictive genes. Classic risk score is based on age, gender, serum LDL and smoking status. Combination is the inclusion of all parameters in a multivariate model.

(14.9%) and the mitochondria (3.51%), according to the cellular compartment library of Gene Ontology (30). Other libraries in Gene Ontology, such as biological process and molecular function, did not demonstrate any substantial trends; the most highly represented process was signal transduction, into which 4.39% of the genes were annotated.

Notably, inspection of the list revealed only a few genes that would be expected to be biomarkers on the basis of the known pathophysiology of atherosclerosis. Such examples include *PDGFD*, a member of the platelet-derived growth factor family that was recently identified as being associated with myocardial infarction (3); interleukin-17C (*IL17C*), which encodes a cytokine distinct to the T-helper 17 subtype of CD4⁺ T cells; C-X-C motif receptor 6 (*CXCR6*), a chemokine receptor; and IL-7 receptor (*IL7R*), the receptor for the hematopoietic growth factor IL-7.

Genotype-Based Prediction

The 25 risk-SNPs previously reported to be predictive of myocardial infarction in healthy individuals (2–4) were used to generate genotype risk scores. This score was comprised of a simple count of risk alleles in each patient and had an average value of 28.5 ± 3.35 . Depending on the data set and time frame, the predictive value was between 0.54 and 0.55. Its contribution to the combined prediction

of ischemic events was minor; in both data sets, the combined AUC values from Table 2 decreased by 0.02 when genotype score was removed.

DISCUSSION

We have demonstrated, as proof of concept, that genome-wide gene expression signatures from carotid endarterectomy samples improve risk prediction of ischemic events. This might have future implications in clinical care by helping identify high-risk subjects where an individualized tailored approach may be implemented to more effectively prevent ischemic events.

Prediction using gene expression signatures improves the accuracy and sensitivity of classic risk markers. Choosing detection thresholds that will detect 77% of patients who will have an ischemic event will have a false-positive result for 35% of the remaining patients, who will in fact be event-free. If gene expression profiling is not performed, however, only 58% of the patients who will have ischemic events will be detected at the same false-positive rate. For every six patients tested, one extra correct prediction will be made. Whether this improvement justifies the use of expression profiling in the clinical setting is a matter of cost versus benefit, but the costs will likely decline as the price of microarrays decreases in response to competition with next-generation sequencing or through

the use of medium-throughput PCR based measurements

When developing gene expression signatures as biomarker tools, it is important to address several points. First, variations in predictivity by gene expression profiling have several causes that result from the study design. Prospective studies provide more realistic estimates than case-control studies but are unrealistic as initial research tools. Similarly, the use of intracohort cross-validation or independent cohort validation is a matter of resources, in that the latter essentially requires complete collection and analysis of an additional biobank. Our carotid plaque expression data set is to our knowledge unique, and an independent validation cohort is therefore not possible. Under this circumstance, intracohort cross-validation provides the most efficient use of data.

Second, the choice of algorithm is also an important decision. Several algorithms exist to make predictions from high-dimensional genomic data (31), any of which will slightly alter the results. However, when using a cross-validation setup, the selection of algorithms can undermine the fundamental model of cross-validation (32); thus, we initially chose an approach that, compared with other methods, can be considered simplistic but that did not require any tuning of the parameters or optimization of the study setup. The data set that we examined is now available in Gene Expression Omnibus for use with various algorithms and for independent validation with other data sets.

Third, the clinical benefit depends to a high degree on the predictivity of expression signatures. In cancer profiling, the topic of predictivity has been controversial—receiving comments ranging from praise as the histopathology of the 21st century to critical comparisons with pure chance (20,33). In a large prospective validation study, the MammaPrint® gene signature score for time to distant breast cancer metastasis at 5 years had an AUC value of 0.681 (34).

In this investigation, as in cancer research, we found it necessary to use several genes in a gene signature, rather than focus on individual genes. It is likely that gene signatures are more robust, because they use information from several different pathways. In fact, the realization that any defined type of tumor is a highly heterogeneous mix of adverse expression profiles is cited as one of the chief contributions to prediction profiling in cancer (20). On the basis of our results from this investigation, the same may hold true for cardiovascular disease.

We also observed that gene expression signatures from plaque tissue predict ischemic events better than gene expression signatures from circulating blood cells. Therefore, there is no support for the part of our initial hypothesis that asks if circulating blood cells carry predictive properties. Possibly, the signs of future disease are too weak in circulating blood cells, compared with carotid artery plaque tissue. Regardless, the choice of tissue must be carefully considered in future studies of gene expression on the basis of risk profiling. In this context, it is of interest to note that the obtained plaque gene expression profile is not from the plaque that subsequently causes an ischemic event. Rather, it is likely to be indicative of the general state of plaques in a patient, an idea previously discussed in relation to morphology of endarterectomized plaques (35).

Finally, we observed that risk profiling using plaque expression performed better than risk profiling using risk-SNP genotypes. In contrast to GWASs, our results are from patients with established atherosclerosis, and a direct comparison to GWAS findings for early myocardial infarction in previously healthy individuals is therefore not obvious. This result has nevertheless been observed previously in a study of 846 coronary artery bypass graft patients. Here it was found that the inclusion of a single SNP in the 9p21 locus improved the mortality prediction using classic risk parameters from an AUC of 0.777–0.782 (5). On the

basis of our data, however, we conclude that, in patients with established atherosclerosis, the gene expression signature of a disease should be more helpful toward prediction than measurements of genotype traits.

Our predictions are based on the gene expression profiles in 126 patients who experienced 25 new or recurrent ischemic events. All gene selections and risk score calculations were performed in a blinded, cross-validation iteration and therefore constitute a single risk score variable in the prediction calculations in Figure 3. However, a limitation of the study is that six variables were fit by Cox regression with 25 events—four classic variables, one gene expression-based variable and one genotype-based variable (although the latter did not affect the results). Regardless, compared with several combinations of classic risk profiling markers, prediction is improved using the single variable of gene expression risk signature.

CONCLUSION

This study provides proof of concept that gene expression signatures from the atherosclerotic plaque provide information that can aid prediction of ischemic events. This prediction is superior to prediction on the basis of genotypes or expression profiling based on circulating blood cells. The benefit of preferentially treating patients who are at higher risk has already been acknowledged with regard to the established use of risk score algorithms that are on the basis of routinely collected patient characteristics. Improving these risk score algorithms could enhance the quality of clinical care. Further work will be devoted to develop clinically useful tests on the basis of gene expression data.

ACKNOWLEDGMENTS

This work was supported by grants from the Swedish Research Council; Combine Sweden; DASTI (Danish Agency for Science, Technology and Innovation); the regional agreement on medical training and clinical research

(ALF); Swedish Heart-Lung Foundation, Foundation for Strategic Research; Vin-nova Foundation; and the European Commission (AtheroRemo project).

DISCLOSURE

The authors wish to declare that a US patent application is pending regarding the application of this method (13/396,005).

REFERENCES

- Furie KL, et al. (2011) Guidelines for the prevention of stroke in patients with stroke or transient ischemic attack: a guideline for healthcare professionals from the American Heart Association / American Stroke Association. *Stroke*. 42:227–76.
- Kathiresan S, et al. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* 41:334–41.
- Coronary Artery Disease (C4D) Genetics Consortium. (2011) A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* 43:339–44.
- Schunkert H, et al. (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* 43:333–8.
- Muehlschlegel JD, et al. (2010) Chromosome 9p21 variant predicts mortality after coronary artery bypass graft surgery. *Circulation*. 122:S60–5.
- Chon H, et al. (2004) Broadly altered gene expression in blood leukocytes in essential hypertension is absent during treatment. *Hypertension*. 43:947–51.
- Cappuzzello C, et al. (2009) Gene expression profiles in peripheral blood mononuclear cells of chronic heart failure patients. *Physiol. Genomics*. 38:233–40.
- Wang Y, et al. (2007) Gene expression signature in peripheral blood detects thoracic aortic aneurysm. *PLoS One*. 2:e1050.
- Wingrove JA, et al. (2008) Correlation of peripheral-blood gene expression with the extent of coronary artery stenosis. *Circ. Cardiovasc. Genet.* 1:31–8.
- Aziz H, Zaas A, Ginsburg GS. (2007) Peripheral blood gene expression profiling for cardiovascular disease assessment. *Genomic Med.* 1:105–12.
- Faber BC, et al. (2001) Identification of genes potentially involved in rupture of human atherosclerotic plaques. *Circ. Res.* 89:547–54.
- Vemuganti R, Dempsey RJ. (2005) Carotid atherosclerotic plaques from symptomatic stroke patients share the molecular fingerprints to develop in a neoplastic fashion: a microarray analysis study. *Neuroscience*. 131:359–74.
- Papaspiridonos M, et al. (2006) Novel candidate genes in unstable areas of human atherosclerotic plaques. *Arterioscler. Thromb. Vasc. Biol.* 26:1837–44.
- Ijas P, et al. (2007) Microarray analysis reveals overexpression of CD163 and HO-1 in symptomatic carotid plaques. *Arterioscler. Thromb. Vasc. Biol.* 27:154–60.
- Saksi J, et al. (2011) Gene expression differences between stroke-associated and asymptomatic carotid plaques. *J. Mol. Med. (Berl)*. 89:1015–26.
- Agardh HE, et al. (2011) Expression of fatty acid-binding protein 4/aP2 is correlated with plaque instability in carotid atherosclerosis. *J. Intern. Med.* 269:200–10.
- Peeters W, et al. (2011) Adipocyte fatty acid binding protein in atherosclerotic plaques is associated with local vulnerability and is predictive for the occurrence of adverse cardiovascular events. *Eur. Heart J.* 32:1758–68.
- Heidecker B, et al. (2008) Transcriptomic biomarkers for individual risk assessment in new-onset heart failure. *Circulation*. 118:238–46.
- Alizadeh AA, Eisen MB, Davis RE, Staudt LM. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 403:503–11.
- Weigelt B, Baehner FL, Reis-Filho JS. (2010) The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. *J. Pathol.* 220:263–80.
- Ferguson GG, et al. (1999) The North American Symptomatic Carotid Endarterectomy Trial: urgent results in 1415 patients. *Stroke*. 30:1751–8.
- Razuvaev A, et al. (2011) Correlations between clinical variables and gene-expression profiles in carotid plaque instability. *Eur. J. Vasc. Endovasc. Surg.* 42:722–30.
- Gabrielsen A, et al. (2010) Thromboxane synthase expression and thromboxane A2 production in the atherosclerotic lesion. *J. Mol. Med. (Berl)*. 88:795–806.
- Tunstall-Pedoe H, et al. (1994) Myocardial infarction and coronary deaths in the World Health Organization MONICA Project: Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation*. 90:583–612.
- Merlo J, et al. (2000) Comparison of different procedures to identify probable cases of myocardial infarction and stroke in two Swedish prospective cohort studies using local and national routine registers. *Eur. J. Epidemiol.* 16:235–43.
- Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 4:249–64.
- Simon R, Radmacher MD, Dobbin K, McShane LM. (2003) Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J. Natl. Cancer Inst.* 95:14–8.
- Gentleman RC, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5:R80.
- Dai M, et al. (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* 33:e175.
- Ashburner M, et al. (2000) Gene ontology: tool for the unification of biology: the Gene Ontology Consortium. *Nat. Genet.* 25:25–9.
- Bovelstad HM, Nygard S, Borgan O. (2009) Survival prediction from clinico-genomic models: a comparative study. *BMC Bioinformatics*. 10:413.
- Jelizarow M, Guillemot V, Tenenhaus A, Strimmer K, Boulesteix AL. (2010) Over-optimism in bioinformatics: an illustration. *Bioinformatics*. 26:1990–8.
- He YD, Friend SH (2001) Microarrays—the 21st century divining rod? *Nat. Med.* 7:658–9.
- Buyse M, et al. (2006) Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J. Natl. Cancer Inst.* 98:1183–92.
- Hellings WE, et al. (2010) Composition of carotid atherosclerotic plaque is associated with cardiovascular outcome: a prognostic study. *Circulation*. 121:1941–50.