Supplementary Materials.

1) Detailed Methods and Materials
2) Supplementary Table 1. Validation of the of the 29 probe sets from the Tang et al, 2006 study [1]. Cross-validated Probabilities.
3) Supplementary Table 2. Classification Accuracy (% correct classification) of 3h and 24h Ischemic Stroke (IS) Predictors. Cross-validated Probabilities.
4) Supplementary Table 3. Annotations for the combined 3h and 24h IS predictors.
5) Supplementary Figure 1. Diagram of the analysis work flow for the identification of IS predictors.
6) Supplementary Figure 2. PAM prediction accuracy of IS and healthy using the 29 probe set predictors of IS from Tang et al, 2006. Cross-validated Probabilities
7) Supplementary Figure 3. PAM 3h IS vs. Healthy. test set + test set confusion matrix.
8) Supplementary Figure 4. PAM 3h IS vs. MI. test set + test set confusion matrix.
9) Supplementary Figure 5. PAM 3h IS vs. SAVVY. test set + test set confusion matrix.
10) Supplementary Figure 6. PAM 24h IS vs. Healthy. test set + test set confusion matrix.
11) Supplementary Figure 7. PAM 24h IS vs. MI. test set + test set confusion matrix.
12) Supplementary Figure 8. PAM 24h IS vs. SAVVY test set + test set confusion matrix.
13) Supplementary Figure 9. Combined 3h, 24h and 3+24h IS Predictors. Cross validated probabilities for 3h IS predictors (A), 24h IS predictors (B) and 3h+24h IS predictors.

Supplementary Material – Materials and Methods

Study Participants

Ischemic Stroke (IS) Patients

Participants with acute IS (n=68) were recruited from the CLEAR trial, a multicenter, randomized double blind safety study of recombinant tissue-plasminogen activator (r-tPA) and eptifibatide as previously described [8] (NCT00250991 at Clinical-Trials.gov). Blood samples were collected at ≤ 3 hours (3h IS), 5 hours (5hr IS) and 24 hours (24 IS) following ischemic stroke onset. r-tPA, with or without eptifibatide, was administered following the 3h blood draw. IS was diagnosed by a stroke neurologist with access to all clinical and diagnostic tests including neurovascular imaging data.

Control Groups

Vascular Risk Factor Subjects (SAVVY)

Subjects with at least one cardiovascular risk factor (hypertension, diabetes mellitus, hyperlipidemia, or tobacco smoking) were recruited from the SAVVY (Sex, Age and Variation in Vascular functionalitY) study (n=52). These subjects are referred to as vascular risk factor SAVVY Controls in the current study. Exclusion criteria were past history of cardiovascular disease (including stroke, coronary artery disease, peripheral artery disease or deep vein thrombosis), BMI > 46kg/m^2, history of cancer, chronic infection, autoimmune disease or blood dyscrasias.

Patients with Myocardial Infarction (MI)

Subjects with MI (n=16) were recruited from the University of California Davis Medical Center. The average time since the event was 58.0h (range 19.3-176.5). All were treated acutely with anti-platelet drugs and an anticoagulant prior to the blood draw. Angioplasty (n=8)

or CABG (n=1) were performed in some of the patients prior to the blood draw. No MI patient received r-tPA.

Healthy Controls

Healthy controls were recruited from the University of Cincinnati (n=15), UC Davis (n=3) and Stanford (n=20). These subjects had never been hospitalized, were on no medications, and had no known major medical, surgical or psychiatric diseases.

Baseline demographic data were compared between the previous [1] and current study as well as between current IS and control subjects using Student's 2-tail t-test for continuous variables (age) and a χ^2 or Fisher Exact tests for categorical variables (gender, race).

Probe-level Data Analysis

Raw expression values of each probe from the Affymetrix U133 Plus 2.0 expression arrays were collapsed into probe set level data using Robust Multichip Averaging (RMA) normalization [9], as well as by modified internal-gene normalization (manuscript in preparation) to a subset of stably expressed internal genes [10]. This involved Median Polishing summarization step, division of each individual gene expression value by the geometric mean of the reference genes, and $\log_2$-transformation. For the analysis in Objective 1, both RMA and Internal control gene normalized values were used. For all the analysis of Objective 2, the derivation of the discriminatory genes was performed using the internal control gene normalized values. The same values were used in developing the Classifiers.

Batch Correction

Due to the unbalanced nature of the batches, bias is introduced when batch is used as a factor in an ANCOVA model. However, it is still desirable to account for the existing technical variation. This was accomplished by selecting genes that were common to the ANCOVA output sets with and without batch as a factor. While this technique introduced strict criteria for the

selection of discriminating genes, it was intended to improve the chance of validation of the results upon subsequent studies and to achieve greater generalization, which can be translated into IS predictive clinical test.

Identification of Discriminatory Genes

Analysis of each comparison (IS per time-point (3h and 24h) vs Healthy, MI and SAVVY, respectively) was performed individually. The samples were randomly split, stratified by Group, in order to perform a split-sample analysis, where the Prediction Algorithms are trained on half of the samples (Training Set), and the performance of the Classifiers is tested on the second half of the samples (Test Set). The Analysis Workflow Chart is shown in Supplementary Materials Figure 1. The feature selection for the derivation of the discriminatory genes between Healthy and IS at 3h and IS at 24h, respectively, involved finding common probe sets from four different ANCOVA analysis, referred to here as Models 1-4. All factors used in the analysis were common to all models (Group, Age, Gender) with the exception of Batch, which was only factored in Model 1 and 3. Models 1 and 2 were applied to a randomly selected one-half of the samples stratified by Group and time-point (for the IS samples) named here 1st random half, whereas Models 3 and 4 were applied to the complete data sets. Overlap of models with and without batch was performed due to the unbalanced nature of batches in an attempt to select more reliable probe sets. Overlap of complete-set and split-set models was performed to achieve greater generalization compared to the split set model which can be translated into IS predictive clinical test.

Gene lists satisfying the following criteria were developed: FDR-corrected p-value (Group) $\leq$0.05 and fold-change $\leq$-1.5 or $\geq$1.5, as well as being not-significant for the rest of the factors (uncorrected p (Age)>0.5 and uncorrected p (Gender)>0.05 and, for the models including Batch, uncorrected p (Batch) >0.05). The goal is to find genes whose expressions are not affected by significant technical (batch), gender, or age effects.

Exception to Flow Chart Analysis for IS at 24h vs Healthy was at Model 1, where the uncorrected p (Group) <0.01 was used to generate a larger gene list. Analysis of SAVVY vs IS at 3h and IS at 24h, respectively, included only Models 2 and 4, since Batch could not be factored in, due to the complete confounding of the batches. Analysis of MI vs IS at 3h and IS at 24h, respectively, included only Models 3 and 4, since the sample size of the MI patients was very small (n=17). In this case a 10-fold cross-validation procedure was used to determine the performance of the Classification Algorithms. If the number of the probe sets at the feature selection step was large, we proceeded with excluding probe sets not annotated, annotated as chromosomal segments, annotated as hypothetical proteins, probe sets which per Affimetrix annotation may potentially detect more than one unique gene (* _x_at, *_a_at, *_s_at), and exclusion of duplicates.

Predictions/Classification

Different prediction algorithms were used. Prediction Analysis of Microarrays (PAM) uses the K-nearest neighbor as a classification engine (default k=10) as well as nearest shrunken centroid as a feature-selection method [11]. The differentially expressed genes that passed the criteria outlined above were input into PAM and the minimum numbers of genes with the optimal classification accuracy were selected. In addition, multiple other classification methods were evaluated in the analysis of the combined 3h IS predictors, 24h IS predictors and 3h plus 24h IS predictors in order to find an optimal model and to produce an unbiased estimate of prediction accuracy (analysis performed in Partek Genomics Suite, Partek Inc., St. Louis, MI, USA). We used a combination of the ANCOVA models and nearest-shrunken centroids for our feature reduction step. In addition to PAM, the classification models used in this study were K-Nearest Neighbor (K-NN) with k = 1, 3, 5, 7, and 9 number of neighbors with Euclidian Distance similarity measure; Nearest-Centroid (NC) with equal and proportional prior probabilities;

Quadratic Discriminant Analysis (QDA) with equal and proportional prior probabilities, Linear Discriminant Analysis (LDA) with equal and proportional prior probabilities, and Support Vector Machine, constituting a 121-model space. For overview of these methods, see [13, 14]. 2-level nested cross-validation (CV) was performed to generate a less biased estimate of classification success (reported as accuracy (normalized) estimate). In this approach, an "outer" cross-validation is performed in order to produce an unbiased estimate of prediction error (by holding out samples as an independent test set). To select the optimal model to be applied to the held out test sample, additional "inner" cross-validation is performed on the training data (which is the data not held out as test data by the "outer" cross-validation). Full leave-one-out cross validation (CV) was used in cases where the complete set was used to train and CV the prediction accuracy.

For Table 4 in the Results section, the following parameters were used: [*]Accuracy (normalized) estimate of 121-Model Space=91.2% (80.3/88). Best Model: SVM (shrink=yes, cost=101, nu=0.5, tol=0.001, kern rbf deg = 3, radial basis function (gamma) = 0.01, coef=0.0). Kappa =0.83. [†]Accuracy (normalized) estimate of 121-Model Space=87.9% (76.4/87). Best Model: SVM (shrink=yes, cost=101, nu=0.5, tol=0.001, kern rbf deg = 3, radial basis function (gamma) = 0.0001, coef=0.0). Kappa=0.83. [‡]Accuracy (normalized) estimate of 121-Model Space=91.2% (110/121). Best Model: SVM (shrink=yes, cost=701, nu=0.5, tol=0.001, kern rbf deg = 3, radial basis function (gamma) = 0.00001, coef=0.0). [‖]Correct classification at 3h=76%, at 24h=97%. [#]Correct classification at 3h=94%, at 24h=97%.

Gene Enrichment Analysis of Discriminatory Genes to Identify Biological Themes in the Combined 3h and 24h IS Predictors

Ingenuity Pathway Analysis (IPA 8.0, Ingenuity® Systems) was used for identifying over-represented biological functions in the combined 97 probe set list of 3h and 24h predictors. A Fisher's exact test ($p<0.1$) was used to determine whether there was over representation of the 97 probe sets/genes in any given biological function. Gene ontology of the stroke predictors was extracted from Affymetrix NetAffix website (https://www.affymetrix.com/user/login.jsp?toURL=-/analysis/netaffx/index.affx).

SUPPLEMENTARY MATERIALS

Supplementary Figure 1. Diagram of the analysis work flow for the identification of IS predictors.

Supplementary Figure 2. PAM prediction accuracy of IS and healthy using the 29 probe set predictors of IS from Tang et al, 2006. The internal gene normalized expression values of all IS (n=70, 199 samples) and healthy (n=38) for the 29 IS predictors from Tang et al, 2006 were used as input in PAM. K-NN (number of neighbors n=10) threshold =0 (including all 29 predictors, and a 10-fold cross-validation was used to estimate prediction accuracy. X-axis represents sample number and the Y-axis represents cross-validated probability of diagnosis. A sample is considered misclassified if the predicted class does not match the known class with a probability greater than 0.5.

Supplementary Figure 3. PAM 3h vs. Healthy test set + test set confusion matrix

Supplementary Figure 4. PAM 3h vs. MI CV + CV confusion matrix

Supplementary Figure 5. PAM 3h vs. SAVVY test set + test set confusion matrix

Supplementary Figure 6. PAM 24h vs. healthy test set + test set confusion matrix

Supplementary Figure 7. PAM 24h vs. MI CV + CV confusion matrix

Supplementary Figure 8. PAM 24h vs. SAVVY test set + test set confusion matrix

Supplementary Figure 9. PAM on Combined 3h, 24h and 3+24h IS predictors. CV Probabilities.

Figure 9A. 3h IS predictors. Combined 60-probe set predictors from combined analysis on 3h IS vs all controls (healthy, MI and SAVVY) were input in PAM.

Figure 9B. 24h IS predictors. Combined 46-probe set predictors from combined analysis on 24h IS vs all controls (healthy, MI and SAVVY) were input in PAM.

Figure 9C. Combined 3h and 24h IS predictors. Combined 97-probe set predictors from combined analysis on 3h IS and 24h IS vs all controls (healthy, MI and SAVVY) were input in PAM.

SUPPLEMENTARY TABLES

Supplementary Table 1. Validation of the of the 29 probe sets from the Tang et al, 2006 study [1]. Cross-validated Probabilities. Trained and cross- validated on current study samples (IS: n =70, 199 samples) and Healthy (n=38, 38 samples).

Normalization Method	Class Prediction	Study	3h	5h	24h	All Time Points
RMA	IS, Sensitivity, %	Tang *et al*, 2006	66.7	86.7	100	84.4
		Current Study	86.6	98.5	89.4	91.5
	Healthy, Specificity, %	Tang *et al*, 2006	N/A	N/A	N/A	100
		Current Study	N/A	N/A	N/A	84.2
Internal Genes	IS, Sensitivity, %	Tang *et al*, 2006	73.3	93.3	100	88.9
		Current Study	86.6	98.5	95.5	93.5
	Healthy, Specificity, %	Tang *et al*, 2006	N/A	N/A	N/A	100
		Current Study	N/A	N/A	N/A	89.5

Sensitivity = % correct classification of IS samples

Specificity = % correct classification of healthy samples

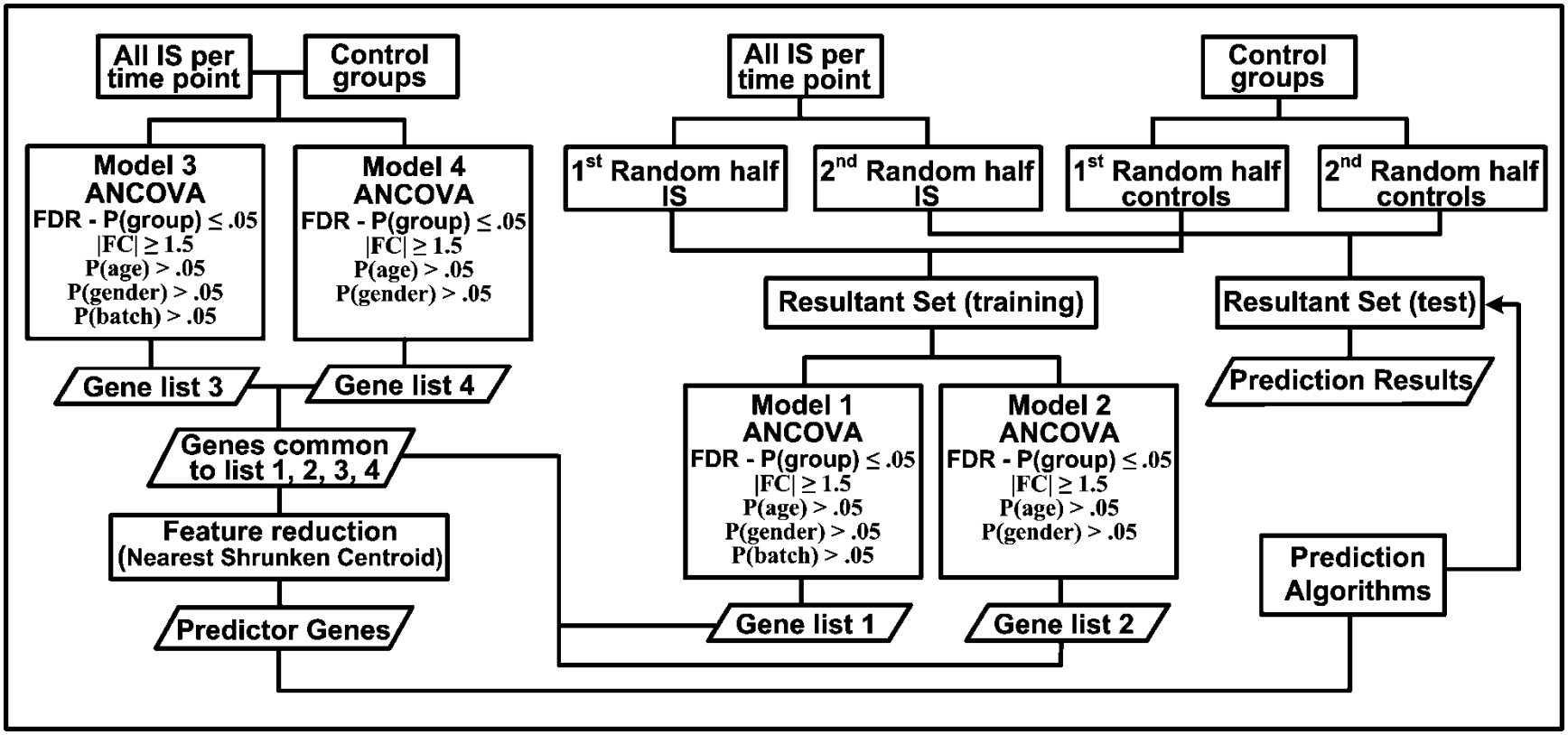
Supplementary Table 2. Classification Accuracy (% correct classification) of 3h and 24h Ischemic Stroke (IS) Predictors. Sample sizes used for Cross-Validation were n=67 at 3h IS, n=66 at 24h IS, n=52 for SAVVY, n=17 for MI. Sample sizes used for split-sample prediction performance estimate on the test set were n=33 at 3h IS, n=33 at 24h IS, n=26 for SAVVY, n=8 for MI. The 60–probe set 3h IS predictors represented the sum of the 3h IS comparison to the three control groups: Healthy (17 probe sets), SAVVY controls (22 probe sets) and MI (31 probe sets). The 46–probe set 24h IS predictors represented the sum of the 24h IS comparison to the three control groups: Healthy (20 probe sets), SAVVY controls (9 probe sets) and MI (17 probe sets). The 3h and 24h IS Combined predictors represent the sum of the 3h IS predictors (60 probe sets) and 24h IS predictors (n=46) of which 9 were common, thus yielding 97 probe sets.

	60 probe sets 3h IS vs Controls (Healthy, MI, SAVVY)		46 probe sets 24h IS vs Controls (Healthy, MI, SAVVY)		97 probe sets 3h and 24h IS Combined vs Controls (Healthy, MI, SAVVY)	
Group	**PAM**	**SVM***	**PAM**	**SVM†**	**PAM**	**SVM‡**
IS	90	91	88	91	90‖	96
SAVVY	94	98	98	98	94	98
MI	71	88	65	82	71	82
Healthy	82	84	79	84	79	76

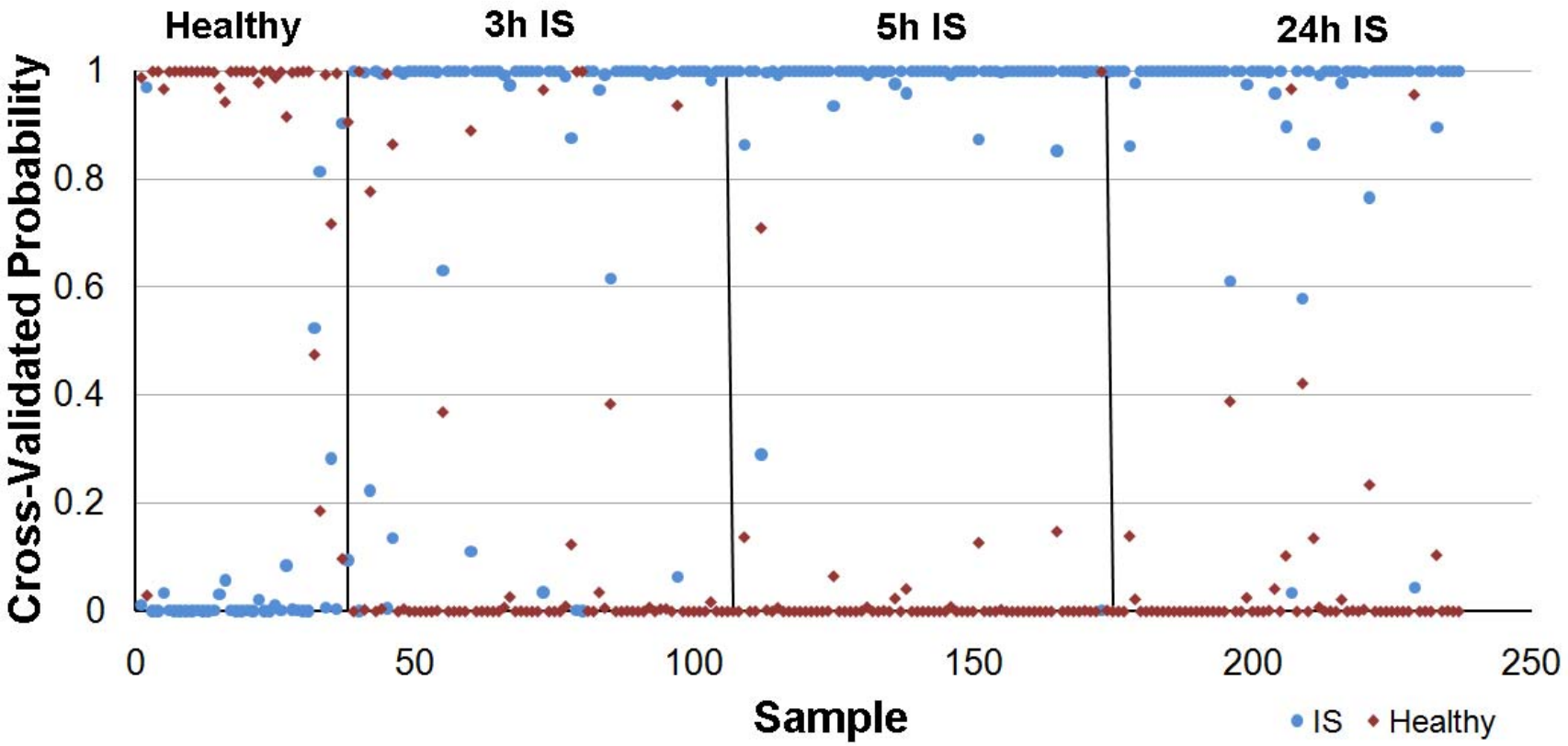
*Accuracy (normalized) estimate of 121-Model Space=86.4% (150/174). Best Model: SVM (shrink=yes, cost=201, nu=0.5, tol=0.001, kern rbf deg = 3, radial basis function (gamma) = 0.001, coef=0.0). †Accuracy (normalized) estimate of 121-Model Space=89.2% (154/173). Best Model: SVM (shrink=yes, cost=201, nu=0.5, tol=0.001, kern rbf deg = 3, radial basis function (gamma) = 0.0001, coef=0.0). ‡Accuracy (normalized) estimate of 121-Model Space=88.2% (212/240). Best Model: SVM (shrink=yes, cost=101, nu=0.5, tol=0.001, kern rbf deg = 3, radial basis function (gamma) = 0.01, coef=0.0). ‖Correct classification at 3h=87%, at 24h=96%

Supplementary Table 3. Annotations for the combined 3h and 24h IS predictors – in a separate Excel file.

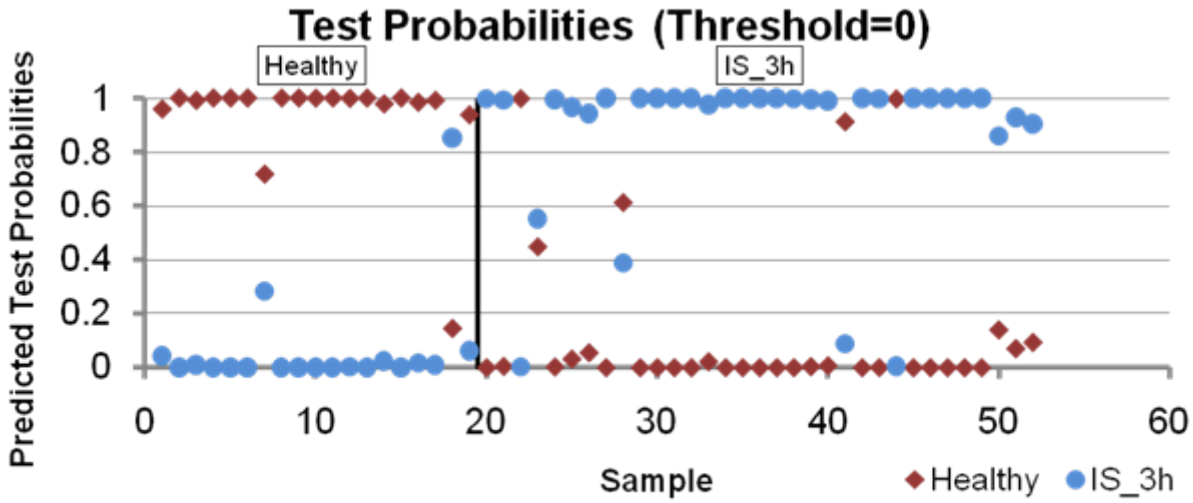
Supplementary Materials Figure 1
Analysis Workflow



Supplementary Materials Figure 2
29-probe set from Tang et al, 2006 study on our IS and Healthy Subjects. Cross-Validated Probability.



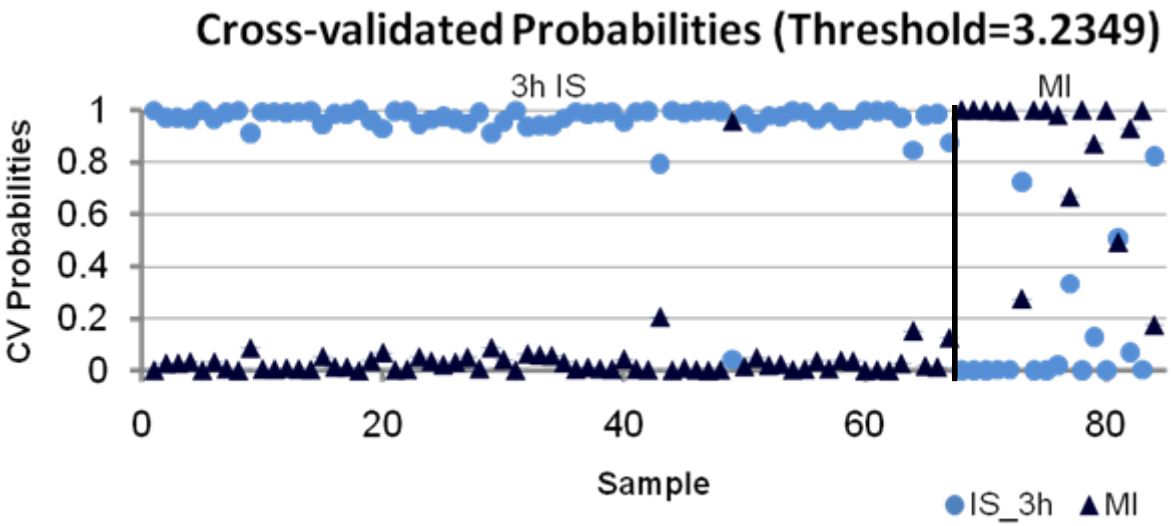
Supplementary Materials Figure 3
Ischemic Stroke at 3h versus Healthy



Test set Prediction Confusion Matrix (Threshold=0)

True\Predicted	Healthy	IS_3h	Correct Classification, %
Healthy	18	1	94.73
IS_3h	4	29	87.9

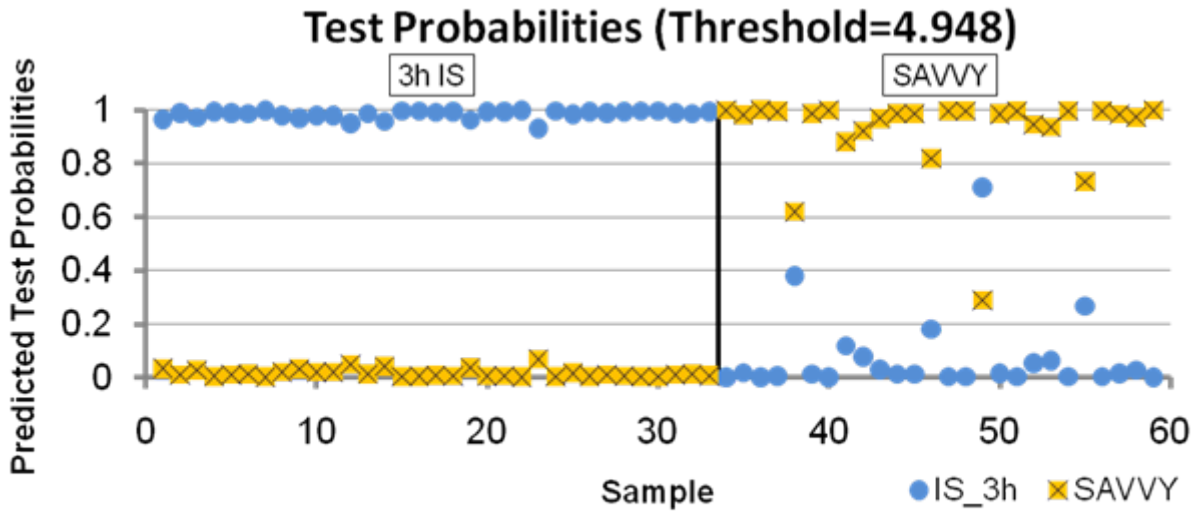
Supplementary Materials Figure 4
Ischemic Stroke at 3h versus Myocardial Infarction



CV Confusion Matrix (Threshold=3.23495)

True\Predicted	IS_3h	MI	Correct Classification, %
IS_3h	66	1	98.5
MI	3	14	82.4

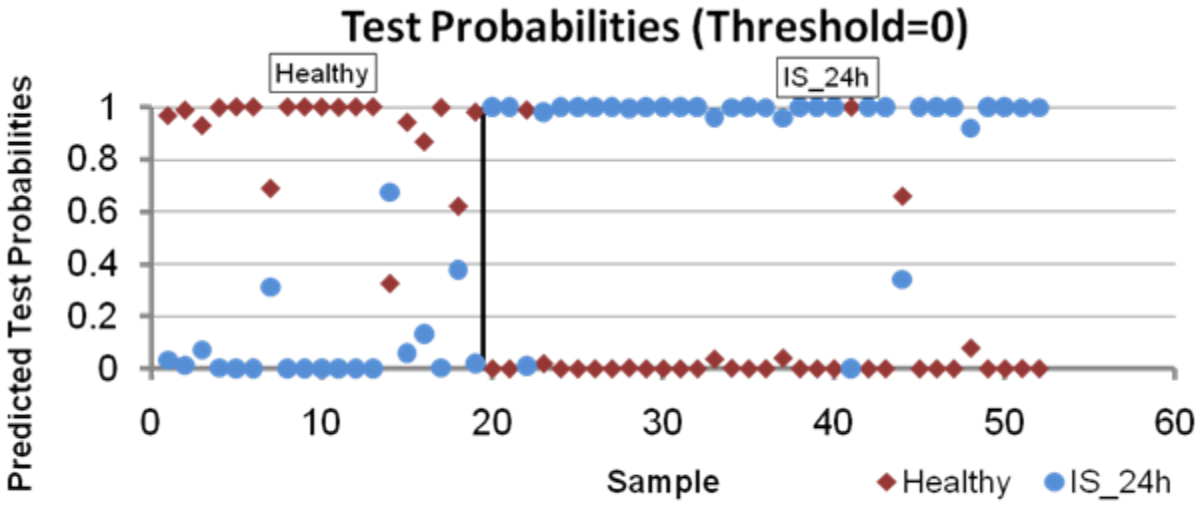
Supplementary Materials Figure 5
Ischemic Stroke at 3h versus SAVVY



Test set Prediction Confusion Matrix (Threshold=4.948)

True\Predicted	IS_3h	SAVVY	Correct Classification, %
IS_3h	33	0	100
SAVVY	1	25	96.2

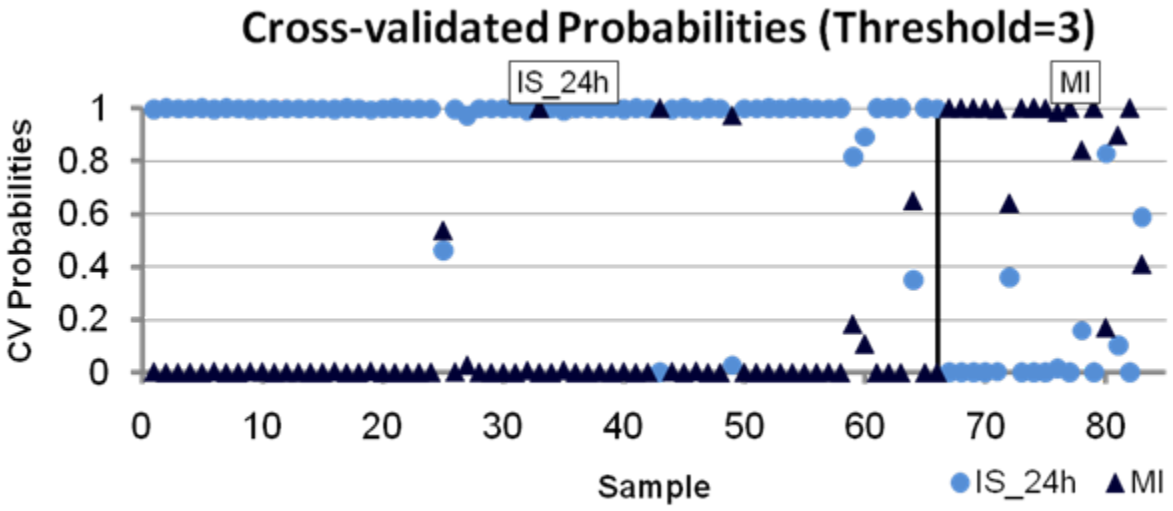
Supplementary Materials Figure 6
Ischemic Stroke at 24h versus Healthy



Test set Prediction Confusion Matrix (Threshold=0)

True\Predicted	Healthy	IS_24h	Correct Classification, %
Healthy	18	1	94.7
IS_24h	3	30	90.9

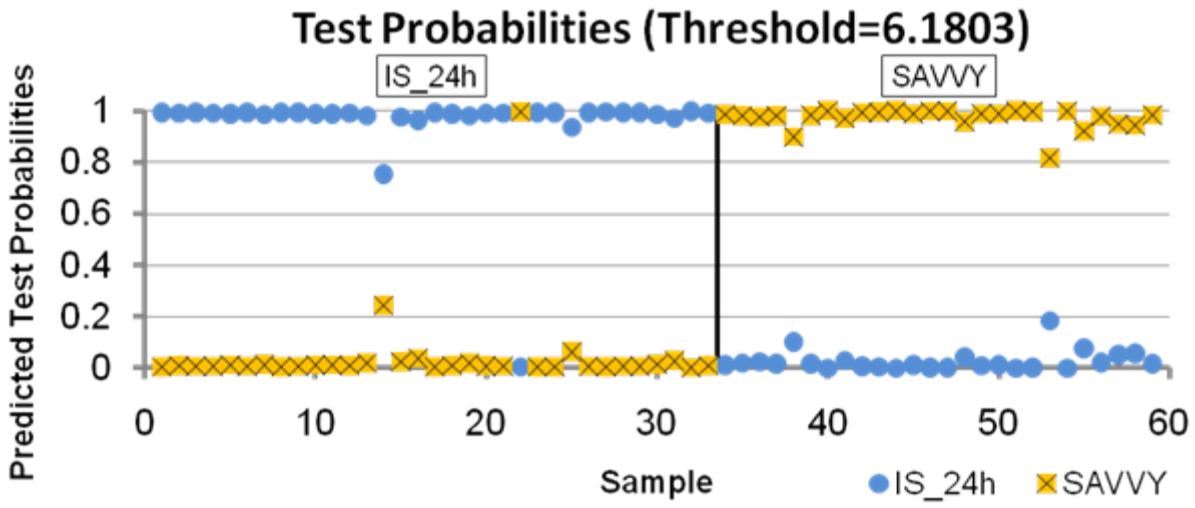
Supplementary Materials Figure 7
Ischemic Stroke at 24h versus Myocardial Infarction



CV Confusion Matrix (Threshold=2.92544)

True\Predicted	IS_24h	MI	Correct Classification, %
IS_24h	62	4	93.9
MI	2	15	88.2

Supplementary Materials Figure 8
Ischemic Stroke at 24h versus SAVVY



Test set Prediction Confusion Matrix (Threshold=6.1803)

True\Predicted	IS_24h	SAVVY	Correct Classification, %
IS_24h	32	1	97
SAVVY	0	26	100

Supplementary Materials Figure 9. Cross Validated Probabilities of
A. 3h IS predictors.
B. 24 IS predictors.
C. Combined 3h and 24h IS predictors