Online supplement

Novel blood-based transcriptional biomarker panels predict the late phase asthmatic response

Amrit Singh[1,2,3], Casey P. Shannon[2], Young Woong Kim[1,2], Chen Xi Yang[1,2], Robert Balshaw[4], Gabriela V. Cohen Freue[5], Gail M. Gauvreau[6], J. Mark FitzGerald[7,8], Louis-Philippe Boulet[9], Paul M. O'Byrne[6], Scott J. Tebbutt[1,2,8]

Supplemental Methods

Blood collection and processing

Blood samples were collected immediately prior to (pre) and 2-3 hours after (post) allergen inhalation challenge using standard operating protocols at each participating centre. At each time point blood was collected into EDTA tubes (3mL of blood) and PAXgene (PreAnalytiX – Qiagen/BD, Valencia, CA, USA) Blood RNA tubes. Complete blood cell counts and differentials were obtained using a hematology analyzer [Cell Dyn 3700 System (Abbott Diagnostics, IL, USA)]. The EDTA blood tubes were then used to prepare buffy coat, plasma and concentrated erythrocyte fractions by centrifugation at 500 x g for 10 minutes at room temperature. 2.5mL of blood was collected in PAXgene Blood RNA tubes which contain an additive that lyses red blood cells and stabilizes intracellular RNA preventing both degradation and changes in gene expression profiling due to sample handling. These samples were kept at -80°C prior to shipment to the Tebbutt laboratory in Vancouver, Canada.

After overnight thawing of PAXgene tubes, total RNA was purified using RNA extraction kits such as PAXgene Blood RNA Kit (PreAnalytiX-Qiagen, Germany) and PAXgene Blood miRNA Kit (PreAnalytiX-Qiagen, Germany) using 5mL of the PAXgene solution (~2.5mL of blood + 6.4 mL of stabilizing reagent). RNA concentrations were determined using NanoDrop 8000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA). RNA quality was assessed using the Agilent 2100 Bioanalyzer following the RNA 6000 Nano Kit protocol (Agilent Technologies, Santa Clara, CA, USA).

RNA-Sequencing

Total RNA was purified and sent to Génome Québec (Centre d'Innovation Genome Quebec et Université McGill) for whole transcriptome sequencing (8 samples per lane). Quality control

(Nano Chip Bioanalyzer and Nanodrop) for each sample was again performed by Genome Quebec. External RNA Control Consortium (ERCC) spike-in controls (92 sequences) were added to all samples. rRNA/globin-depleted stranded cDNA libraries were sequenced using an Illumina HiSeq 2000 as 100 bp paired end reads.

RNA-Seq data analysis

All RNA-Sequencing files passed quality control metrics based on FastQC standards. Initially the first 12 bases of all 100 bp reads were trimmed and all left reads (R1) files were concatenated into one file and all right (R2) end reads were concatenated into another file using all samples (Note: for the purposes of the *de novo* assembly additional RNA-Seq data from asthmatic subjects from repeat challenges was included). The left and right concatenated files were used by the Trinity software(1) (version r20131110) to construct a *de novo* assembly of the blood transcriptome (Reverse forward RF-stranded library type was specified). The abundance estimates of Trinity contigs were estimated using RSEM (2) (RNA-Seq by Expectation Maximization, version 1.2.11) using the Bowtie aligner.

A similar approach was used for genome-guided assemblies. However, after advice from Illumina on how to increase alignment rates, Seqtk (version 1.0) was used to trim the first 5 bases and the last 25 bases of all 100 bp paired end reads (this was not repeated for the Trinity *de novo* assembly). The University of California, Santa Cruz (UCSC) gene transfer format (GTF) file [which contained the genomic positions for Coding sequences (CDS)] and the list of known UCSC gene-isoforms (http://genome.ucsc.edu/cgi-bin/hgTables?hgsid=419083647_CPaaYnAZklf85tfa5qclmzQszSKh), were used to extract transcript sequences from the 2013 human reference genome (GRCh38 build) (http://hgdownload.cse.ucsc.edu/goldenPath/hg38/chromosomes/) in order to build a reference

transcriptome for RSEM. The abundance estimates of UCSC gene and gene-isoforms were estimated using RSEM (version 1.2.19) using the Bowtie2 (version 2.2.4) aligner. RSEM estimates the abundance of gene-isoforms and then sums up the isoforms of a given gene to determine the abundance of genes. 89,357 UCSC gene-isoforms corresponded to 42,465 genes whereas 258,403 Trinity "isoforms" corresponded to 212,373 Trinity "genes". The Trinity dataset at the "isoform" sequence level was used for downstream analyses such that, if needed, probe-specific sequences could be designed.

STAR (Spliced Transcripts Alignment to a Reference, version 2.5.0a), a fast RNA-Seq alignment software (3) was also used to align the paired end reads to the human genome using annotations from the GENCODE project (4) (GENCODE release v21, http://www.gencodegenes.org/releases/reference_releases.html), which include protein coding genes, transcribed variants, long non-coding RNAs, and pseudogenes. Feature counts (5) in the Subread/Rsubread package (version 1.5.0) was used to estimate the abundance of 60,155 Ensembl transcripts as well the 92 ERCC control sequences.

UCSC and Ensembl identifiers were annotated using the biomaRt R library (version 2.26.0). Trinity contigs were annotated using Trinotate: Transcriptome Functional Annotation and Analysis (https://trinotate.github.io/), which uses BLASTX to query nucleotide sequences (translated in protein sequences) against protein databases such as SwissProt and UniPort. For selected Trinity contigs that were transferred to the NanoString platform, the 100bp probe sequences were also queried against the human reference genome (GRCh38.p2) using BLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

The count data for all mRNA datasets were normalized to log2-counts per million ($\log_2$ cpm) (6):

$$X_{norm} = \log_2\left(\frac{\left(X_{counts} + 0.5\right)^T}{\left(lib.size + 1\right) * 10^6}\right)$$

All alignment and mapping statistics for all four RNA-Seq datasets, UCSC genes, UCSC gene-isoforms, Ensembl and Trinity are provided below. The 92 ERCC spike-ins were used to establish a lower limit of detection.

RNA-Sequencing alignment results.

	UCSC genes	UCSC gene-isoforms	Ensembl transcripts	Trinity contigs
Trimming	seqtk – remove first 5 and last 25 bases	seqtk – remove first 5 and last 25 bases	seqtk – remove first 5 and last 25 bases	seqtk – remove first 12 bases sickle
Aligner/Counter	Bowite2/RSEM	Bowite2/RSEM	STAR/subread	Bowtie/RSEM
Number of transcripts	42,465	89,357	60,155	258,403
Total number of read pairs (Millions) Mean±SD (Min, Max)	25.2±2.6 (18.0, 31.6)		21.3±2.5 (16.3, 26.5)	24.6±2.5 (17.7, 30.8)
Alignment rate (%)	35.6±5.1 (14.8, 48.4)*		83.9±4.3 (66.4, 89.9)** 38.2±4.9 (23.2, 49.8)*	44.9±5.35* (14.80, 55.46)

*Alignment to the transcriptome
**Alignment to the genome

Pre-filtering of datasets:

Four expression datasets were generated from the RNA sequencing data: UCSC genes dataset (42,465 genes), UCSC gene-isoforms dataset (89,357 gene-isoforms), Ensembl dataset (60,155 transcripts) and a Trinity dataset (258,403 contigs). Below 14 attomoles/μL or 3 $\log_2$ counts per million mapped reads (cpm), the linearity between the ERCC spike-in concentrations and RNA-

Seq counts, was diminished (Figure S1). Pre-filtering of transcripts below 3 $\log_2$ cpm resulted in 6,598 UCSC genes, 6,078 UCSC gene-isoforms, 7,518 Ensembl transcripts and 5,227 Trinity contigs (additional details can be found in the online supplement).

nCounter Elements TagSets

100 base pairs (bp) of oligonucleotide probes were designed for all biomarker candidates using the nDesign portal (http://www.NanoString.com/) and with the help of the bioinformatics team at NanoString Technologies. For unknown targets, the sequences were selected based on coverage plots of RNA-Seq data and provided to NanoString Technologies. NanoString Technologies provided the nCounter Elements TagSets which contain fluorescently labeled reporter tags (unique linear combinations of 6 spots of color specific for each specific RNA transcript target) and a biotinylated universal capture tag (8). During hybridization the reporter and capture probes bind to target-specific probes, probe A (50 bases), and probe B (50 bases), both of which bind to the single stranded RNA transcript target sequence. After several wash steps, this Tag Complex is immobilized to the glass cartridge and scanned allowing digital counting and quantification of RNA molecules (8, 9).

Probe A and Probe B pairs for all biomarker candidates as well as both Probe A pools and Probe B pools were ordered from Integrated DNA Technologies, Inc. (IDT, Iowa, USA) (pools consists of all probes A or B for all biomarker candidates). Prior to commencing assay, 30X Working Probe Pools were created for a set number of assays resulting in a 8.3x fold dilution of the master Probe A and B stocks. The final concentration of each Probe A in the 30X Working Probe A and B Pool was 0.6nM and 3nM respectively. For a single run of 12 samples, 130µL of the hybridization buffer was added to 65µl of the TagSet followed by 65µL of the

Extension TagSet. Then 13µl of the 30X Working Probe A Pool was added. After briefly inverting and spinning down of the TagSet tube 13µl of the 30X Working Probe B Pool was added. Lastly, 39µl of RNAase-free water was added bringing the total volume of the master mix for 13 assays (dead volume of one extra assay) to 325µl (25µl per assay). Hybridization reactions were performed in sets of 12 using a strip of 12 tubes. 25µl of the master mix was added to each tube in addition to 5µl (100ng) of the RNA sample. After 16 hours hybridization at 67°C using the Bio-Rad PCR system, the strip of tubes was placed in the nCounter Prep Station for fully automated purification and immobilization of Tag Complexes to the flow cell. After 3 hours of high sensitivity preparation, the flow cell was scanned using the nCounter Digital Analyzer under the maximum field of view (max FOV) setting (555 FOVs).

NanoString data quality control and normalization

All NanoString data were assessed for various quality control (QC) metrics (10) such as the number of fields of view, binding density, linearity of positive control spike-ins and the lower limit of detection (see below for a complete list of QC parameters):

1. Imaging QC: % FOV (field of view) must be greater than 75 FOV.
2. Binding Density QC: Binding density must be between 0.05 and 2.25.
3. Positive Control Linearity QC: The R2 must be greater than 0.9 between the counts and concentrations of the 6 positive controls.
4. Positive Control Limit of Detection QC: The second lowest positive control spike in (0.5fM) must have counts greater than the Mean±2SD of the negative controls for each sample.

NanoString uses six positive control ERCC spike-in controls corresponding to six different concentrations in the 30µL hybridization; 128fM, 32fM, 8fM, 2fM, 0.5fM and 0.125fM in order assess the technical performance of the assays. Data normalization was performed independently in the discovery, validation and control cohorts. The geometric mean of all six

positive controls for each sample was calculated. Then the mean of all sample geometric means was divided by each sample geometric mean in order to determine a normalization factor for each sample. The expression data for each sample was multiple by the corresponding sample normalization factor. This resulted in all geometric means of all positive controls for each sample to be equivalent thus normalizing for assay to assay variability. Technical variability was assessed by running replicates (aliquots) of the same RNA sample across cartridges and across time (same day and different days).

Statistical Analyses

Classification algorithms

Elastic net(11) (Enet) and random forest (12) (Rf), two commonly used classification methods (13–16) for high dimensional datasets (p>>n, ie. the number of variables (*e.g.* genes/transcripts) far exceeds the number of samples), were used to identify biomarker panels of the late phase asthmatic response for each RNA-Seq data. Enet is a penalized regularization method that performs both shrinkage of regression coefficients (increases predictive performance) and variable selection (selects a limited number of variables from high dimensional datasets) resulting in a parsimonious model. This is achieved by finding β (regression coefficients) that minimizes

$\left\|y - X\beta\right\|^2 + \lambda P(\alpha)$, where the elastic net penalty is $P(\alpha) = \frac{1-\alpha}{2}\left\|\beta\right\|_2^2 + \alpha\left\|\beta\right\|_1$.

The tuning parameter λ shrinks the coefficient estimates towards zero and is estimated using cross-validation. The selection parameter α $(\alpha \in [0,1])$ controls the level of sparsity in the model, where $\alpha = 0$ corresponds to the model containing all p variables and $\alpha = 1$ corresponds to a model with many zero coefficients and a few non-zero coefficients (selected variables).

Random forest is an ensemble method which combines the predictions of many tree classifiers. For the present analysis 500 tree classifiers were built from 500 bootstrap (with replacement) datasets. Each tree was built by splitting randomly selected m variables (m= $\sqrt{\text{total number of variables}}$) in each split with no pruning. However, variables are ranked based on the Gini importance score which reflects how often a given variable is selected for a split and its ability to split samples into their respective classes. In order to keep the number of selected variables by Enet and Rf the same, for a given Enet panel size k, k top ranked variables based on the importance score were selected for Rf. Enet and Rf were implemented using the glmnet (version 2.0-1) and randomForest (version 4.6-12) R libraries.

Classification performance

Deep cross-validation (k x M-fold) was used to assess the classification performance of the biomarker panels. Briefly, the data was split into M folds, 1/M served as the test set and (M-1)/M served as the training set. For each fold, differential expression analysis was performed using the training set and the top ranked 500 transcripts (with the smallest p-values) were used to build the classification model. Each classification model was then used to predict the probabilities of the subjects in the held out test set. This process was repeated until all folds had been used as a test set (M times). The result was a list of probabilities for all subjects that were used to calculate the area under the receiver operating curve (AUROC), a measure of classification performance. The M-fold cross-validation is repeated k times. When M = N, also called a leave-one-out cross-validation, each subject is treated as a test set. The output of the deep cross-validation (k x M-fold) is k number of AUROCs, which are then used to calculated the Mean ± standard deviation of the AUROC per biomarker panel. For both the discovery and validation phase an AUROC

cut-off of 0.70 was used as per recommended guidelines for clinical biomarker implementation (17).

Gene-set enrichment analysis

Tissue enrichment analysis was performed using the sear R library (version 0.1, https://github.com/cashoes/sear), which uses a hypergeometric test to test for over representation (enrichment) of various cell-specific genes in the input gene list.

Pathway enrichment was performed using Enrichr(18), a web based application that allows the user to upload a list of gene symbols. The gene list is then compared to curated lists of canonical pathways/molecular processes from various databases [Kyoto Encyclopedia of Genes and Genomes (KEGG), Biocarta, and WikiPathways and tests for enrichment using a Fisher's exact test.

Supplemental Results

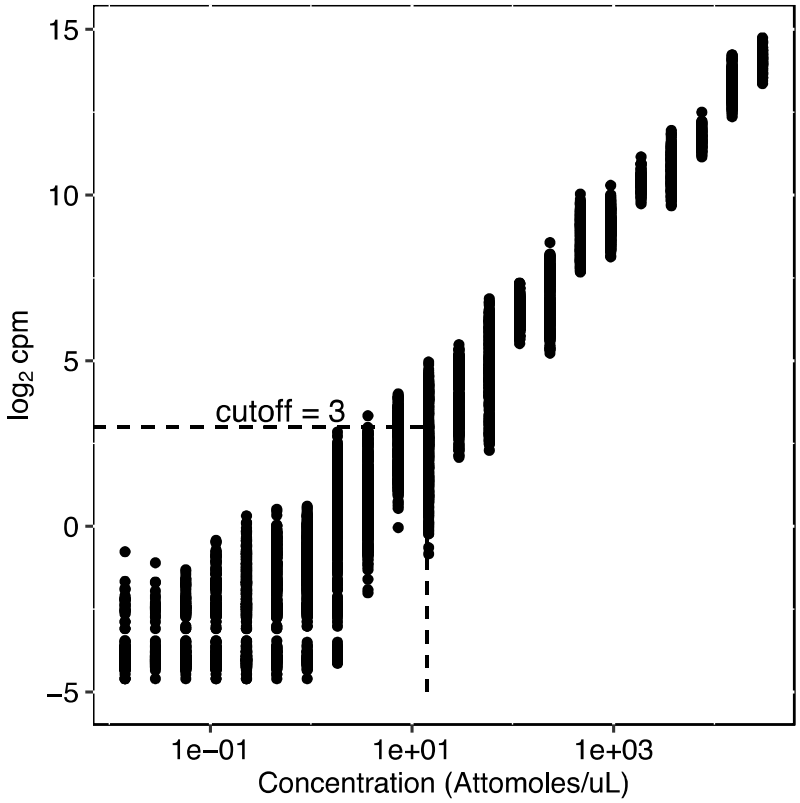

Figure S1: Establishing a limit of detection in RNA-Seq data. A cut-off of 3 log2 cpm (counts per million) was used to remove all genes across all samples below this threshold.

NanoString Elements probe design

Six house-keeping probes as well 75 probes corresponding to 117 biomarker candidates (due to gene redundancy across datasets) were designed. Transcripts with UCSC gene and Ensembl identifiers were annotated to their corresponding gene symbol. Based on the gene name, a probe was selected from the nDesign portal (www.nanostring.com) with the assumption that for each gene, the 100 base NanoString probe would have an expression similar to that of the much longer sequence of nucleotides from the RNA-Seq data. The table below describes the characteristics for each biomarker probe in the different mRNA biomarker panels.

Table S1: NanoString probe sequences

House-keepers

	Transcript	Accession	Position	Target sequence	**Tm CP**	**Tm RP**
1	ARPC4	NM_005718.4	971-1070	GACTTTGTGATCCACTTCATGGAGGAGATTGACAAGGAGATCAGTGAGAT GAAGCTGTCAGTCAATGCCCGTGCCCGCATTGTGGCTGAAGAGTTCCTTA	83	84
2	TMBIM6	NM_003217.2	1221-1320	GGCCCTTCCTTCCTCATTGTTGTTTGGTATGCGCACAGTTCCTGTGGGACT GGGCCGTGAGTTTTCCATTGGAAAGAAGTTCAGTGGTCCCATTGTTAAC	80	78
3	RHOA	NM_001664.2	1231-1330	GGTACTCTGGTGAGTCACCACTTCAGGGCTTTACTCCGTAACAGATTTTG TTGGCATAGCTCTGGGGTGGGCAGTTTTTTGAAAATGGGCTCAACCAGAA	79	84
4	MED13	NM_005121.1	1386-1485	GCTGGACAACAAGGACAGGCACCATCTTTAGGTCAGCAACAACAAATACT TCCTAAGCACAAGACCAATGAGAAGCAAGAAAAGAGTGAAGAGCCACAGA	80	79
5	TOR1AIP2	NM_145034.4	206-305	AGGCGTGAGCCACCACGCCTGGCCTCACTCAAACAGTAAACGAGTCACTGT ATATATGCCTTATTTCCTGGGTCCATGGGAAGCATGAGTCTGTTGGGAC	84	82
6	WAC	NM_100486.2	756-855	CCTCTGGACTGAACCCCACATCTGCACCTCCAACATCTGCTTCAGCGGTCC CTGTTTCTCCTGTTCCACAGTCGCCAATACCTCCCTTACTTCAGGACCC	82	80

Tm CP: melting temperature of the capture probe

Tm RP: melting temperature of the reporter probe

UCSC genes biomarker panel

	Transcript	Accession	Position	Target sequence	**Tm CP**	**Tm RP**
1	ABHD5	NM_016006.4	956-1055	TAAAATGCACCCTGACATTCCAGTTTCAGTGATCTTTGGCGCCCGATCCT GCATAGATGGCAATTCTGGCACCAGCATCCAGTCCTTACGACCACATTCA	80	84
2	AHCTF1	NM_015446.3	406-505	TAATCATGGAGGAGCCAGTGCAAGCACTCAGCATTTACATCCAAGTCTGC GATGGCTTTTTGGAGTGGCAGCTGTGGTCACTGATGTTGGACAGATCCTT	82	83
3	ATP8A1	NM_006095.2	871-970	GAGGATTTCTGGCAGAATTGAGTGTGAAAGTCCAAACAGACATCTCTACG ATTTTGTTGGAAACATAAGGCTTGATGGACATGGCACCGTTCCACTGGGA	81	82
4	B3GNT5	NM_032047.4	131-230	GGAAGGAAAGCCGACCTCCGATTTGGACATTTAAAGAGCTGGGCTTGAAC TTCGTGAGTTTCGCTCTAAACTGCCCTTGAAATGAAGCTGGACTTGGAGG	79	76
5	C9orf78	NM_016520.2	911-1010	AGAAGTTCAAGAAAATGAATAGGCGGTACTGAGTTGTGCAGAGTGGGAT GTAAATATCGCCTTCCTCTCCCTATATCCCTCCCATGAAAAATGGCTTCCT	84	79
6	CARM1	NM_199141.1	1166-1265	TCCTGATGGCCAAGTCTGTCAAGTACACGGTGAACTTCTTAGAAGCCAAA GAAGGAGATTTGCACAGGATAGAAATCCCATTCAAATTCCACATGCTGCA	84	81
7	CASP8	NM_001228.4	302-401	AGATGGACTTCAGCAGAAATCTTTATGATATTGGGGAACAACTGGACAGT GAAGATCTGGCCTCCCTCAAGTTCCTGAGCCTGGACTACATTCCGCAAAG	82	85
8	CD300LB	NM_174892.2	1531-1630	ACGAAAACCATCGCAGGAAATGGCACCCTCCCTTTTCGGTGATGTTGAAA	82	80

				TCATGTTACTAATGAAAACTGTCCTAGGGAAGTGGTTCTGTCTCCTCACA		
9	CD4	NM_000616.4	976-1075	TGGCAGGCGGAGAGGGCTTCCTCCTCCAAGTCTTGGATCACCTTTGACCT GAAGAACAAGGAAGTGTCTGTAAAACGGGTTACCCAGGACCCTAAGCTCC	86	86
10	CD8A	NM_001768.5	1321-1420	GCTCAGGGCTCTTTCCTCCACACCATTCAGGTCTTTCTTTCCGAGGCCCCT GTCTCAGGGTGAGGTGCTTGAGTCTCCAACGGCAAGGGAACAAGTACTT	83	83
11	CLEC4E	NM_014358.2	571-670	GAGTTTTTTATTGGACTGTCAGACCAGGTTGTCGAGGGTCAGTGGCAAT GGGTGGACGGCACACCTTTGACAAAGTCTCTGAGCTTCTGGGATGTAGGGG	78	85
12	CNTNAP3	NM_033655.3	1433-1532	TCGAACGTGGAACAGAGCAGGACATTTGCTTTTCGGCGAACTTCGACGTG GTTCAGGGAGTTTCGTCCTCTTTCTTAAGGATGGCAAGCTCAAACTGAGT	81	79
13	DAP	NM_004394.1	1791-1890	CTGAGGGAGCATGGCACAGCCTCACACTTGAAAGACGGTGTTTGGTTTCC CATCTAATCAACTTAAGGGAAGCCGGCATGTACCCTTCAAGGCCCTGTCA	82	79
14	F13A1	NM_000129.3	3197-3296	TTCAGGTCCCCTTTCAGAGATATAATAAGCCCAACAAGTTGAAGAAGCTG GCGGATCTAGTGACCAGATATATAGAAGGACTGCAGCCACTGATTCTCTC	80	85
15	FAM8A1	NM_016255.2	1792-1891	CCTGGGGAAATTGTCTTTGGTGTTTAGAGGAGGGAATGAGAACACAAATT GGATAATCCACTGTCTCCCATCCCAGGAGGTGGTGAGTTGGCTACAAGAG	80	84
16	FOS	NM_005252.2	1476-1575	ACTCAAGTCCTTACCTCTTCCGGAGATGTAGCAAAACGCATGGAGTGTGT ATTGTTCCCAGTGACACTTCAGAGAGCTGGTAGTTAGTAGCATGTTGAGC	79	81
17	FUT7	NM_004479.3	1711-1810	ACTGGCATGAATGAGAGCCGATACCAACGCTTCTTTGCCTGGCGTGACAG GCTCCGCGTGCGACTGTTCACCGACTGGCGGGAACGTTTCTGTGCCATCT	83	84
18	GNLY	NM_006433.2	306-405	CAGGAGCTGGGCCGTGACTACAGGACCTGTCTGACGATAGTCCAAAAACT GAAGAAGATGGTGGATAAGCCCACCCAGAGAAGTGTTTCCAATGCTGCGA	82	81
19	HBA2	NM_000517.4	437-536	CCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAAT ACCGTTAAGCTGGAGCCTCGGTAGCCGTTCCTCCTGCCCGCTGGGCCTC	89	90
20	HLA-A	NM_002116.5	1001-1100	GGAAGAGCTCAGATAGAAAAGGAGGGAGTTACACTCAGGCTGCAAGCAG TGACAGTGCCCAGGGCTCTGATGTGTCCCTCACAGCTTGTAAAGTGTGAGA	85	86
21	KRT23	NM_015515.3	1736-1835	CGGGAAGAATCAAAGTCGAGCATGAAAGTGTCTGCAACTCCAAAGATCAA GGCCATAACCCAGGAGACCATCAACGGAAGATTAGTTCTTTGTCAAGTGA	83	82
22	LMBRD1	NM_018368.3	303-402	GGAGCTGGTGATCGGCTGGTGCATATTCGGCCTCTTACTACTGGCTATTTT GGCATTCTGCTGGATATATGTTCGTAAATACCAAAGTCGGCGGGAAAGT	84	81
23	NAPA	NM_003827.2	897-996	GTACAGCGCCAAAGACTACTTCTTCAAGGCGGCCCTCTGCCACTTCTGCAT CGACATGCTCAACGCCAAGCTGGCTGTCCAAAAGTATGAGGAGCTGTTC	83	84
24	NFKBIA	NM_020529.1	946-1045	GGATGAGGAGAGCTATGACACAGAGTCAGAGTTCACGGAGTTCACAGAGG ACGAGCTGCCCTATGATGACTGTGTGTTTGGAGGCCAGCGTCTGACGTTA	82	83
25	PABPC1	NM_002568.3	322-421	CCAGCGGCAGTGGATCGACCCCGTTCTGCGGCCGTTGAGTAGTTTTCAAT TCCGGTTGATTTTTGTCCCTCTGCGCTTGCTCCCCGCTCCCCTCCCCCCG	81	82
26	PPP3R1	NM_000945.3	2386-2485	GCCATCGCTGTTCCTTCAACTGAGTGCTGCACATCATGGGCTCTGTCTGT GAGAGAAAAATCCCGGTGCTTGGTGTCCTTGCATGACATGGAGTTTTGCA	83	81
27	RGS2	NM_002923.1	856-955	AACAGCTTCCCTCACTGTGTACAGAACGCAAGAAGGGAATAGGTGGTCTG AACGTGGTGTCTCACTCTGAAAAGCAGGAATGTAAGATGATGAAAGAGAC	82	76

28	SCARNA5	NR_003008.2	103-202	AGTCATGTGTATGGGATCATGGAGCAGCTGATAATTTGGGATTCTGTCAG TGTGTGTTTCTGAGAGTGATCGGCTCACAGCTGACGAGTATCCAACAAAA	80	82
29	SF3B1	NM_012433.2	1-100	GGAAGTTCTTGGGAGCGCCAGTTCCGTCTGTGTGTTCGAGTGGACAAAAT GGCGAAGATCGCCAAGACTCACGAAGATATTGAAGCACAGATTCGAGAAA	82	83
30	SH3BGRL3	NM_031286.2	391-490	AGATTGTCAACGGGGACCAGTACTGTGGGGACTATGAGCTCTTCGTGGAG GCTGTGGAACAAAACACGCTGCAGGAGTTCCTGAAGCTGGCTTGAGTCAA	84	82
31	SULT1A1	NM_177534.2	1394-1493	TGCGAATCAAACCTGACCAAGCGGCTCAAGAATAAAATATGAATTGAGGG CCCGGGACGGTAGGTCATGTCTGTAATCCCAGCAATTTGGAGGCTGAGGT	84	85
32	TGFBI	NM_000358.2	2031-2130	GTGGTCCATGTCATCACCAATGTTCTGCAGCCTCCAGCCAACAGACCTCAG GAAAGAGGGGATGAACTTGCAGACTCTGCGCTTGAGATCTTCAAACAAG	82	80
33	ZNF185	NM_007150.3	1124-1223	CTGAAGGCTTGGCTGCAGTAGACATCGGCTCCGAGAGAGGAAGCTCCAGT GCCACTTCAGTCTCTGCTGTCCCTGCTGATAGGAAGAGCAACAGCACAGC	85	84

Tm CP: melting temperature of the capture probe
Tm RP: melting temperature of the reporter probe

UCSC gene-isoforms biomarker panel

	Transcript	Accession	Position	Target sequence	**Tm CP**	**Tm RP**
1	ATP11A	NM_015205.2	456-555	ATGCCATGAACCAGTGTCCTGTTCATTTCATTCAGCACGGCAAGCTCGTT CGGAAACAAAGTCGAAAGCTGCGAGTTGGGGACATTGTCATGGTTAAGGA	79	83
2	ATP8A1	NM_006095.2	871-970	GAGGATTTCTGGCAGAATTGAGTGTGAAAGTCCAAACAGACATCTCTACG ATTTTGTTGGAAACATAAGGCTTGATGGACATGGCACCGTTCCACTGGGA	81	82
3	CD8A	NM_001768.5	1321-1420	GCTCAGGGCTCTTTCCTCCACACCATTCAGGTCTTTCTTTCCGAGGCCCCT GTCTCAGGGTGAGGTGCTTGAGTCTCCAACGGCAAGGGAACAAGTACTT	83	83
4	CLEC4E	NM_014358.2	571-670	GAGTTTTTTATTGGACTGTCAGACCAGGTTGTCGAGGGTCAGTGGCAAT GGGTGGACGGCACACCTTTGACAAAGTCTCTGAGCTTCTGGGATGTAGGGG	78	85
5	CNTNAP3	NM_033655.3	1433-1532	TCGAACGTGGAACAGAGCAGGACATTTGCTTTTCGGCGAACTTCGACGTG GTTCAGGGAGTTTCGTCCTCTTTCTTAAGGATGGCAAGCTCAAACTGAGT	81	79
6	F13A1	NM_000129.3	3197-3296	TTCAGGTCCCCTTTCAGAGATATAATAAGCCCAACAAGTTGAAGAAGCTG GCGGATCTAGTGACCAGATATATAGAAGGACTGCAGCCACTGATTCTCTC	80	85
7	FAM8A1	NM_016255.2	1792-1891	CCTGGGGAAATTGTCTTTGGTGTTTAGAGGAGGGAATGAGAACACAAATT GGATAATCCACTGTCTCCCATCCCAGGAGGTGGTGAGTTGGCTACAAGAG	80	84
8	FPR2	NM_001462.3	1201-1300	GATGGGGTCAGGGATATTTTGAGTTCTGTTCATCCTACCCTAATGCCAGT TCCAGCTTCATCTACCCTTGAGTCATATTGAGGCATTCAAGGATGCACAG	78	80
9	GNLY	NM_006433.2	306-405	CAGGAGCTGGGCCGTGACTACAGGACCTGTCTGACGATAGTCCAAAAACT GAAGAAGATGGTGGATAAGCCCACCCAGAGAAGTGTTTCCAATGCTGCGA	82	81
10	HBA2	NM_000517.4	437-536	CCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAAT ACCGTTAAGCTGGAGCCTCGGTAGCCGTTCCTCCTGCCCGCTGGGCCTC	89	90
11	IL1R2	NM_173343.1	114-	TGCTTCTGCCACGTGCTGCTGGGTCTCAGTCCTCCACTTCCCGTGTCCTCT	81	78

			213	GGAAGTTGTCAGGAGCAATGTTGCGCTTGTACGTGTTGGTAATGGGAGT		
12	MME	NM_000902.2	5060-5159	GGATTGTAGGTGCAAGCTGTCCAGAGAAAAGAGTCCTTGTTCCAGCCCTA TTCTGCCACTCCTGACAGGGTGACCTTGGGTATTTGCAATATTCCTTTGG	80	79
13	NAPA	NM_003827.2	897-996	GTACAGCGCCAAAGACTACTTCTTCAAGGCGGCCCTCTGCCACTTCTGCAT CGACATGCTCAACGCCAAGCTGGCTGTCCAAAAGTATGAGGAGCTGTTC	83	84
14	NFKBIA	NM_020529.1	946-1045	GGATGAGGAGAGCTATGACACAGAGTCAGAGTTCACGGAGTTCACAGAGG ACGAGCTGCCCTATGATGACTGTGTGTTTGGAGGCCAGCGTCTGACGTTA	82	83
15	PTPN18	NM_014369.3	3421-3520	CCTCTGTGTTGCTGGATAATGAGTCATCTATCTCTGGAGGAGAAGAAAGG CAGGTCCTCCACAGCCCTGATAAAATCTCCAAGTCTCCCAGTTTCGGGTC	84	84
16	QKI	NM_006775.2	839-938	TAATTTTGTTGGGAGAATCCTTGGACCTAGAGGACTTACAGCCAAACAAC TTGAAGCAGAAACCGGATGTAAAATCATGGTCCGAGGCAAAGGCTCAATG	82	83
17	RGS2	NM_002923.1	856-955	AACAGCTTCCCTCACTGTGTACAGAACGCAAGAAGGGAATAGGTGGTCTG AACGTGGTGTCTCACTCTGAAAAGCAGGAATGTAAGATGATGAAAGAGAC	82	76
18	SCARNA5	NR_003008.2	103-202	AGTCATGTGTATGGGATCATGGAGCAGCTGATAATTTGGGATTCTGTCAG TGTGTGTTTCTGAGAGTGATCGGCTCACAGCTGACGAGTATCCAACAAAA	80	82
19	SEMA4D	NM_001142287.1	1121-1220	AAGTGAACCCATCATCTCCCGAAATTCTTCCCACAGTCCTCTGAGGACAGA ATATGCAATCCCTTGGCTGAACGAGCCTAGTTTCGTGTTTGCTGACGTG	84	79
20	SF3B1	NM_012433.2	1-100	GGAAGTTCTTGGGAGCGCCAGTTCCGTCTGTGTGTTCGAGTGGACAAAAT GGCGAAGATCGCCAAGACTCACGAAGATATTGAAGCACAGATTCGAGAAA	82	83
21	SH3BGRL3	NM_031286.2	391-490	AGATTGTCAACGGGGACCAGTACTGTGGGGACTATGAGCTCTTCGTGGAG GCTGTGGAACAAAACACGCTGCAGGAGTTCCTGAAGCTGGCTTGAGTCAA	84	82
22	SULT1A1	NM_177534.2	1394-1493	TGCGAATCAAACCTGACCAAGCGGCTCAAGAATAAAATATGAATTGAGGG CCCGGGACGGTAGGTCATGTCTGTAATCCCAGCAATTTGGAGGCTGAGGT	84	85
23	TGFBI	NM_000358.2	2031-2130	GTGGTCCATGTCATCACCAATGTTCTGCAGCCTCCAGCCAACAGACCTCAG GAAAGAGGGGATGAACTTGCAGACTCTGCGCTTGAGATCTTCAAACAAG	82	80
24	CNTNAP3_isoform	uc004abk.1	4014-4113	ACATAGTTATTAAAATGGGAATAAGTAAGAAAATAGACCTGAGTCACCA CAGAGGAAGTAAATTACACATTGTCATCGGCATTGGAAGGAAAATATACTG	79	78
25	COPB1_isoform	uc001mli.2	193-292	AAGATGCGGAAGGGGAGCGACTAGGCCGCTTGCGTCTGGGCCTGGCAGAAG GGACCGGATTTTCTGGCATCCTTAAATCTTGTGTCAAGGATTGGTTATA	96	75
26	GNAS_isoform	uc002yae.3	425-524	GATCCGAACCCACAACTCCCTGAAGAACAGAATACTATGCTTTTTAGTCG GGATGTCTTTATGAAAGCAGTACTCCTAACTGACATGGTGCAATATGATT	80	79
27	MAP3K8_isoform	uc001ivj.2	25-124	GTCAGTTTCCCATGGGTCTTGAATGCAAATACAAATATCGTAAACTAAAT ATTTGTGTTTTCTTTCCTAGACTCTCCAGAAAGAGCAACAGTAATGGAGT	74	75
28	PELI1_isoform	uc002sct.4	178-277	CACCACAAAGCAGCCCCAACGCCTCTCCCTGCGTCCGCGGCTCCTCAGCGC TCGGCTCCGTGGTCAACTTCCCCTCGCTGGGCTCGGCTGGCGGGCGCGG	86	85
29	PLXNC1_isoform	uc010sut.2	211-310	TGGTTTTTGAGTTACCATGTCCCTCTGATGCAGCATCTCTGTCTCTTAG CGGCCGTGGGGGTGACCAGGCACAAATCGAAGGAGCTGAGTCGCAAACAG A	77	95
30	PTAR1_isoform	uc004ahi.3	742-	TTTCTACCTTCAGCATCACTTAAATGGTAGGTTTCCTCACAGCATGACCCA	78	86

			841	GTTGTCACCTGCAGACAGCCCTGGGGGGACTTTGAGTGACTTGCACCTT		
31	SF3B1_isoform	uc002uug.3	1253-1352	CTGCTATATGCCAGTCCTGTCTGCATTCTTAAGGGTGCAGTTCAACACAT CCTCTCTAGATTATGGTGAAAAAGTATTCCAAAGGAAGTCTTATCAGAGC	80	77
32	TIA1_isoform	uc002sgl.4	883-982	GACTTATTGCAGAAATAGATGAGAAGCAAATCAAGACTACTATTCAAAAA TCAAATACCAAACAGCTATCATATGATGAGGTTGTAAATCAGTCTAGTCC	77	77
33	VCAN_isoform	uc003kij.3	1356-1455	TTCCCTCCCCCTGATAGCAGATTTGATGCCTACTGCTTTAAACGTCGAATG AGTGATTTGAGTGTAATTGGTCATCCAATAGATTCAGAATCTAAAGAAG	77	74
34	VPS13A_isoform	uc004akp.4	9413-9512	CAGGTTGAGGGATGGGACTGGAAATCAAATGTTACAGGCATCAAAAAGT TTGATATGAAAAGTTAATGCATGACTTTGCAAGTGAAAGCCAACAGTAGAT	83	78
35	ZNF609_isoform	uc002ann.3	757-856	CCTGTTTCCACACCAGCAGTGCTGCCAATACACCTTTTGGTGCCAGTGGTC AACAATGACATCTCATCTCCTTGTGAGCAGATCATGGTTCGTACCCGAT	84	79

Tm CP: melting temperature of the capture probe
Tm RP: melting temperature of the reporter probe

Ensembl biomarker panel

	Transcript	Accession	Position	Target sequence	**Tm CP**	**Tm RP**
1	AHCTF1	NM_015446.3	406-505	TAATCATGGAGGAGCCAGTGCAAGCACTCAGCATTTACATCCAAGTCTGC GATGGCTTTTTGGAGTGGCAGCTGTGGTCACTGATGTTGGACAGATCCTT	82	83
2	ATP8A1	NM_006095.2	871-970	GAGGATTTCTGGCAGAATTGAGTGTGAAAGTCCAAACAGACATCTCTACG ATTTTGTTGGAAACATAAGGCTTGATGGACATGGCACCGTTCCACTGGGA	81	82
3	C9orf78	NM_016520.2	911-1010	AGAAGTTCAAGAAAATGAATAGGCGGTACTGAGTTGTGCAGAGTGGGAT GTAAATATCGCCTTCCTCTCCCTATATCCCTCCCATGAAAAATGGCTTCCT	84	79
4	CARM1	NM_199141.1	1166-1265	TCCTGATGGCCAAGTCTGTCAAGTACACGGTGAACTTCTTAGAAGCCAAA GAAGGAGATTTGCACAGGATAGAAATCCCATTCAAATTCCACATGCTGCA	84	81
5	CD300LB	NM_174892.2	1531-1630	ACGAAAACCATCGCAGGAAATGGCACCCTCCCTTTTCGGTGATGTTGAAA TCATGTTACTAATGAAAACTGTCCTAGGGAAGTGGTTCTGTCTCCTCACA	82	80
6	CD8A	NM_001768.5	1321-1420	GCTCAGGGCTCTTTCCTCCACACCATTCAGGTCTTTCTTTCCGAGGCCCCT GTCTCAGGGTGAGGTGCTTGAGTCTCCAACGGCAAGGGAACAAGTACTT	83	83
7	CHP1	XM_005254140.1	175-274	CCCTCCTTCCCTCCTGTCGCCGTCTCTTCTGGCGCCGCTGCTCCCGGAGGA GCTCCCGGCACGGCGATGGGTTCTCGGGCCTCCACGTTACTGCGGGACG	87	87
8	CISH	NM_145071.2	559-658	CACCAATGTACGCATTGAGTATGCCGACTCCAGCTTCCGTCTGGACTCCA ACTGCTTGTCCAGGCCACGCATCCTGGCCTTTCCGGATGTGGTCAGCCTT	84	86
9	CLEC4E	NM_014358.2	571-670	GAGTTTTTTATTGGACTGTCAGACCAGGTTGTCGAGGGTCAGTGGCAAT GGGTGGACGGCACACCTTTGACAAAGTCTCTGAGCTTCTGGGATGTAGGGG	78	85
10	CMC1	NM_182523.1	293-392	CTCTGGAGTTCTTATGGTAGTAAAATGCCGGAAAGAAAATTCTGCATTGA AAGAATGTCTAACTGCTTACTATAATGATCCAGCCTTTTATGAAGAATGC	77	76
11	CNTNAP3	NM_033655.3	1433-1532	TCGAACGTGGAACAGAGCAGGACATTTGCTTTTCGGCGAACTTCGACGTG GTTCAGGGAGTTTCGTCCTCTTTCTTAAGGATGGCAAGCTCAAACTGAGT	81	79

12	CTDSP2	NM_005730.3	1641-1740	GGGGAGAAGCTGAAAGACCAAGACTCTTCCCAAGTTAGCTTGTCTCCTCT CCTGTCACCCTAAGAGCCACTGAGTTGTGTAGGGATGAAGACTATTGAAG	81	82
13	CTSA	NM_001127695.1	1541-1640	TGCCACAATGGGACATGTGCAACTTTCTGGTAAACTTACAGTACCGCCGT CTCTACCGAAGCATGAACTCCCAGTATCTGAAGCTGCTTAGCTCACAGAA	79	81
14	DAP	NM_004394.1	1791-1890	CTGAGGGAGCATGGCACAGCCTCACACTTGAAAGACGGTGTTTGGTTTCC CATCTAATCAACTTAAGGGAAGCCGGCATGTACCCTTCAAGGCCCTGTCA	82	79
15	DESI1	NM_015704.2	2881-2980	CAAGACCCACTGATTTGCCAGTGTGCATGGAAATAATAGATTAGAGCAGA AACTAGCAGGGACTGTTGTATAATCGTGATCTACTAGCAGAATTGGGCCC	82	83
16	F13A1	NM_000129.3	3197-3296	TTCAGGTCCCCTTTCAGAGATATAATAAGCCCAACAAGTTGAAGAAGCTG GCGGATCTAGTGACCAGATATATAGAAGGACTGCAGCCACTGATTCTCTC	80	85
17	FAM8A1	NM_016255.2	1792-1891	CCTGGGGAAATTGTCTTTGGTGTTTAGAGGAGGGAATGAGAACACAAATT GGATAATCCACTGTCTCCCATCCCAGGAGGTGGTGAGTTGGCTACAAGAG	80	84
18	GBE1	NM_000158.3	1051-1150	CTTTGCAGCTTCCAGCCGTTATGGAACACCTGAAGAGCTACAAGAACTGG TAGACACAGCTCATTCCATGGGTATCATAGTCCTCTTAGATGTGGTACAC	81	81
19	GNLY	NM_006433.2	306-405	CAGGAGCTGGGCCGTGACTACAGGACCTGTCTGACGATAGTCCAAAAACT GAAGAAGATGGTGGATAAGCCCACCCAGAGAAGTGTTTCCAATGCTGCGA	82	81
20	HBA2	NM_000517.4	437-536	CCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAAT ACCGTTAAGCTGGAGCCTCGGTAGCCGTTCCTCCTGCCCGCTGGGCCTC	89	90
21	ITSN1	NM_003024.2	856-955	TAAATTACAAAAGGCACAGTCATTTGATGTGGCCAGTGTCCCACCAGTG GCAGAGTGGGCTGTTCCTCAGTCATCAAGACTGAAATACAGGCAATTATTC	82	81
22	KIAA1551	NM_018169.3	746-845	CTTCAGGAGTTACCCAAAACGTATGGTTGAACTCACCAATGAGGAATCCTG TGCATTCTCATATAGGGGCAACTGTATCTCATCAAACTGATTTTGGAGC	81	79
23	KRT23	NM_015515.3	1736-1835	CGGGAAGAATCAAAGTCGAGCATGAAAGTGTCTGCAACTCCAAAGATCAA GGCCATAACCCAGGAGACCATCAACGGAAGATTAGTTCTTTGTCAAGTGA	83	82
24	PABPC1	NM_002568.3	322-421	CCAGCGGCAGTGGATCGACCCCGTTCTGCGGCCGTTGAGTAGTTTTCAAT TCCGGTTGATTTTTGTCCCTCTGCGCTTGCTCCCCGCTCCCCTCCCCCCG	81	82
25	PPP3R1	NM_000945.3	2386-2485	GCCATCGCTGTTCCTTCAACTGAGTGCTGCACATCATGGGCTCTGTCTGT GAGAGAAAAATCCCGGTGCTTGGTGTCCTTGCATGACATGGAGTTTTGCA	83	81
26	RALGPS2	NM_152663.3	1486-1585	GAGGCCAAGCTGAAAGTTCTACTCTTTCTAGTGGAATATCAATAGGTAGC AGCGATGGTTCTGAACTAAGTGAAGAGACCTCATGGCCTGCTTTTGAAAG	81	83
27	RGS2	NM_002923.1	856-955	AACAGCTTCCCTCACTGTGTACAGAACGCAAGAAGGGAATAGGTGGTCTG AACGTGGTGTCTCACTCTGAAAAGCAGGAATGTAAGATGATGAAAGAGAC	82	76
28	SCARNA5	NR_003008.2	103-202	AGTCATGTGTATGGGATCATGGAGCAGCTGATAATTTGGGATTCTGTCAG TGTGTGTTTCTGAGAGTGATCGGCTCACAGCTGACGAGTATCCAACAAAA	80	82
29	SF3B1	NM_012433.2	1-100	GGAAGTTCTTGGGAGCGCCAGTTCCGTCTGTGTGTTCGAGTGGACAAAAT GGCGAAGATCGCCAAGACTCACGAAGATATTGAAGCACAGATTCGAGAAA	82	83
30	SH3BGRL3	NM_031286.2	391-490	AGATTGTCAACGGGGACCAGTACTGTGGGGACTATGAGCTCTTCGTGGAG GCTGTGGAACAAAACACGCTGCAGGAGTTCCTGAAGCTGGCTTGAGTCAA	84	82
31	SMCHD1	NM_015295.2	4676-	AACCACCTACACCAGCTGTTTCAAATGTTCGCTCAGTTGCCAGTAGGACCT	84	81

			4775	TGGTCAGAGATCTACATCTTAGTATCACGGATGACTACGACAACCATAC		
32	TGFBI	NM_000358.2	2031-2130	GTGGTCCATGTCATCACCAATGTTCTGCAGCCTCCAGCCAACAGACCTCAG GAAAGAGGGGATGAACTTGCAGACTCTGCGCTTGAGATCTTCAAACAAG	82	80
33	ZNF185	NM_007150.3	1124-1223	CTGAAGGCTTGGCTGCAGTAGACATCGGCTCCGAGAGAGGAAGCTCCAGT GCCACTTCAGTCTCTGCTGTCCCTGCTGATAGGAAGAGCAACAGCACAGC	85	84
34	ZNF281	NM_012482.3	2606-2705	AGCGTTTGGTTCTCAGTTTAAGTCGGGCAGCAGGGTGCCAATGACCTTTA TCACTAACTCTAATGGAGAAGTGGACCATAGAGTAAGGACTTCAGTGTCA	80	84
35	noNamelncRNA	ENST00000585152.1	1380-1479	GGGAAGGAATAAAGTATGAATTTTCAAGCTGGTTACCCATCTGGGCAACT GGAGCTCAGTTTTGCTGGGAAAGTCTGGAAGACAGTGTCAGATATAACTA	82	83

Tm CP: melting temperature of the capture probe
Tm RP: melting temperature of the reporter probe

Trinity biomarker panel

	Transcript	Accession	Position	Target sequence	**Tm CP**	**Tm RP**
1	CASP8	NM_001228.4	302-401	AGATGGACTTCAGCAGAAATCTTTATGATATTGGGGAACAACTG GACAGT GAAGATCTGGCCTCCCTCAAGTTCCTGAGCCTGGACTACATTCC GCAAAG	82	85
2	CECR1	NM_177405.2	1027-1126	TAGCCAGCCCTCTACAAGCTGTCTTCTTGCACACGCTGTCACTTC CTCTCA CTCGTTCTTGAATCAGCTCCATGTGCCCATGAAATCAATGGCCT CTGTA	83	81
3	FNIP1	NM_001008738.2	1265-1364	ACAATTTGTAATCTTTACACGATGCCACGAATTGGAGAACCTGT CTGGC TTACAATGATGTCGGGGACTCCAGAAAAGAACCACCTTTGCTAT CGTTTCA	79	80
4	FPR2	NM_001462.3	1201-1300	GATGGGGTCAGGGATATTTTGAGTTCTGTTCATCCTACCCTAAT GCCAGT TCCAGCTTCATCTACCCTTGAGTCATATTGAGGCATTCAAGGAT GCACAG	78	80
5	LYST	NM_000081.3	179-278	GGTCATGAGCACCGACAGTAACTCACTGGCACGTGAATTTCTGA CCGATG TCAACCGGCTTTGCAATGCAGTGGTCCAGAGGGTGGAGGCCAG GGAGGAA	83	83
6	QKI	NM_006775.2	839-938	TAATTTTGTTGGGAGAATCCTTGGACCTAGAGGACTTACAGCCA AACAAC	82	83

				TTGAAGCAGAAACCGGATGTAAAATCATGGTCCGAGGCAAAGGCTCAATG		
7	SETX	NM_015046.4	1141-1240	ATCAACAACGCAAGCTACAATAGAGAGATCCGACATATACGGAACAGCTC TGTAAGGACCAAGTTAGAACCGGAGTCCTATCTGGATGATATGGTGACTT	80	80
8	SF3B1	NM_012433.2	1-100	GGAAGTTCTTGGGAGCGCCAGTTCCGTCTGTGTGTTCGAGTGGACAAAAT GGCGAAGATCGCCAAGACTCACGAAGATATTGAAGCACAGATTCGAGAAA	82	83
9	unknown1_comp54405_c1_seq1	comp54405_c1_seq1.1	5-104	TTCTTATCGTGGTGGTAGTTCCACAATGTATACACACGTCTCAATCTACCA AACTACATGCTTCAAAAATGTACACTTTGTTTCATGCCCATTATACCTC	76	73
10	unknown2_comp55647_c0_seq2	comp55647_c0_seq2_2.1	1-100	AAGGGACTGGGTTTCATCACCAGTGACTGCCATGAGTCTAACTTAGTTAT CTGGTTGGATAGGGATGCATTTACTGAACTGGGGACCCAGGAGGAGTTGG	82	85
11	unknown2a_comp55647_c0_seq2	comp55647_c0_seq2_3.1	496-595	GCTGCACATCAAAATCACTGTTTCCTTTGACCAATACTCTAGACACACTCA AGCAGAGATTCTGGATTCTGATTTAATTGACACAGAGTAGGGCCTGGGC	79	84
12	unknown3_comp56590_c0_seq8	comp56590_c0_seq8.1	16-115	GCTGCCCCGAGCCCGCGAAGGGAGGGAAGTTCCAGAATCGAGAGAGGGAG GGAGTCAAGGTGGAACCCATAGAGTGAGCCTCCTGAAGACACAGAGCGGT	87	87
13	FPR1_intron_comp17070_c0_seq1	int_comp17070_c0_seq1.1	1-100	GGATGTTTGGTTTGAAGCTTTCAGGAGAAAGCAAAAGAGCCCTAACGACT TTATGATGGGTCATGGGGAAATGAGTGTAATACAGAAGCAGTCACCTTTC	82	81
14	IFRD1_intron_comp41141_c0_seq1	int_comp41141_c0_seq1.1	82-181	GATAGAGGGTAGGCTTCTTTAGCTCATATGTCTAAGCTTTCTATCTTTGA ATTACAGTTGAAGTTTAATGATCTAGTAAGCACCTGTAAAGCATACAGAA	73	75

Tm CP: melting temperature of the capture probe
Tm RP: melting temperature of the reporter probe

Quality control and reproducibility assessment

Probes for all transcripts were ordered from Integrated DNA Technologies (IDT, Iowa, USA) and assayed in 29/36 samples since RNA for some samples had been completely depleted. The test run of the custom Elements assays had extremely high binding densities for most samples. This was due to the Hemaglobin Subunit Alpha 2 (*HBA2*) transcript that was saturating the NanoString assay. Given the extreme saturation bias of *HBA2*, an inactive probe (19) was designed (same sequence as the *HBA2* but no bound reporter probe) in order to achieve a 95% attenuation (see sample calculation below):

Probe A is bound to the reporter probe therefore the inactive probe sequence for HBA2 was designed (see below) as the exact 50 bp sequence as probe A of HBA2.

5'-TTT GGA GGT CAG CAC GGT GCT CAC AGA AGC CAG GAA CTT GTC CAG GGA GG-3'

Properties
Tm (50nM NaCL)*: 72.6°C
GC Content: 58.0%
Molecular Weight: 15,541.1
nmoles/OD260:2.1
μg/OD260:31.9
Ext.Coefficient: 487,700L/(mole·cm)
Secondary Structure calculations: Lowest folding free energy (kcal/mole): -3.51 at 25C

The concentration of probe A for each tag in the 30X Working Probe A Pool is 0.6 nM. 11.514nM of the inactive probe is added to the 0.6nM of the active probe A bringing the total concentration to 12.146nM in the 30X Working Probe A Pool (see below for calculation).

$$\frac{\left[HBA2_{inactive}\right]}{\left[HBA2_{inactive}\right]+\left[HBA2_{active}\right]} = 0.95$$

$$\frac{\left[HBA2_{inactive}\right]}{\left[HBA2_{inactive}\right]+0.606nM} = 0.95$$

$$\left[HBA2_{inactive}\right] = \frac{0.95 * 0.606}{1-0.95}$$

$$\left[HBA2_{inactive}\right] = 11.514nM$$

The concentration of the inactive probe A sequence of HBA2 received from IDT was 19µM. A 100x dilution resulted in a concentration of 190nM. The inactive probe was added when creating the 30X Working Probe Pool A (see table below):

Number of Assays	Aliquot from Master Probe Stock (µL)	Volume of inactive Probe A	TE-Tween (µL)	Final Volume (µL)
12	4	2.00	27.00	33
24	4	2.00	27.00	33
36	5	2.55	34.45	42
48	7	3.51	47.49	58
60	8	4.06	54.94	67
72	10	5.03	67.97	83

Spiking in the *HBA2* inactive probe resulted in a significant reduction in the binding densities of the discovery and validation cohort samples.

References

1. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 2013;8:1494–1512.
2. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;12:1.
3. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21.
4. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, Barnes I, Bignell A, Boychenko V, Hunt T, Kay M, Mukherjee G, Rajan J, Despacio-Reyes G, Saunders G, Steward C, Harte R, Lin M, Howald C, Tanzer A, Derrien T, Chrast J, Walters N, Balasubramanian S, Pei B, *et al.* GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* 2012;22:1760–1774.
5. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014;30:923–930.
6. Law CW, Chen Y, Shi W, Smyth GK. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 2014;15:R29.
7. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;3:1–12.

8. Singh A, Yamamoto M, Ruan J, Choi JY, Gauvreau GM, Olek S, Hoffmueller U, Carlsten C, FitzGerald JM, Boulet L-P, others. Th17/Treg ratio derived using DNA methylation analysis is associated with the late phase asthmatic response. *Allergy Asthma Clin Immunol* 2014;10:32.
9. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, James JJ, Maysuria M, Mitton JD, Oliveri P, Osborn JL, Peng T, Ratcliffe AL, Webster PJ, Davidson EH, Hood L. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* 2008;26:317–325.
10. NanoString Technologies, Inc. *nCounter Expression Data Analysis Guide*.
11. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol* 2005;67:301–320.
12. Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
13. Hughey JJ, Butte AJ. Robust meta-analysis of gene expression using the elastic net. *Nucleic Acids Res* 2015;43:e79–e79.
14. Van Vliet D, Smolinska A, Jöbsis Q, Rosias PPR, Muris JWM, Dallinga JW, van Schooten FJ, Dompeling E. Association between exhaled inflammatory markers and asthma control in children. *J Breath Res* 2016;10:16014.
15. Huang H-H, Liu X-Y, Liang Y, Chai H, Xia L-Y. Identification of 13 blood-based gene expression signatures to accurately distinguish tuberculosis from other pulmonary diseases and healthy controls. In: Liu F, Lee D-H, Lagoa R, Kumar S, editors. *Biomed Mater Eng* 2015;26:S1837–S1843.
16. Prosperi MC, Marinho S, Simpson A, Custovic A, Buchan IE. Predicting phenotypes of asthma and eczema with machine learning. *BMC Med Genomics* 2014;7:1.

17. Sin DD, Hollander Z, DeMarco ML, McManus BM, Ng RT. Biomarker Development for Chronic Obstructive Pulmonary Disease. From Discovery to Clinical Implementation. *Am J Respir Crit Care Med* 2015;192:1162–1170.
18. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 2013;14:128.
19. NanoString Technologies, Inc. *Strategies for Successful Gene Expression Assays*.