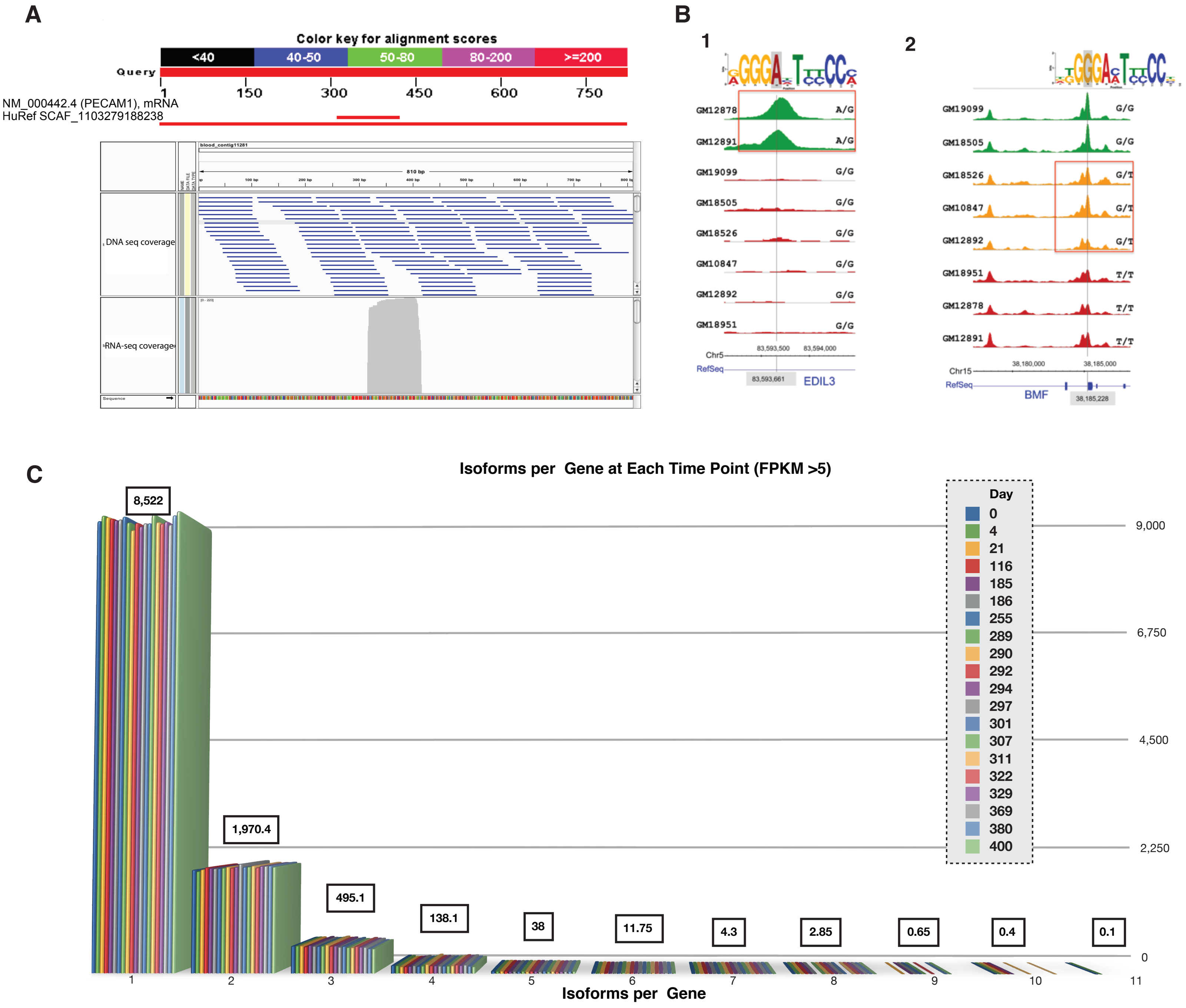
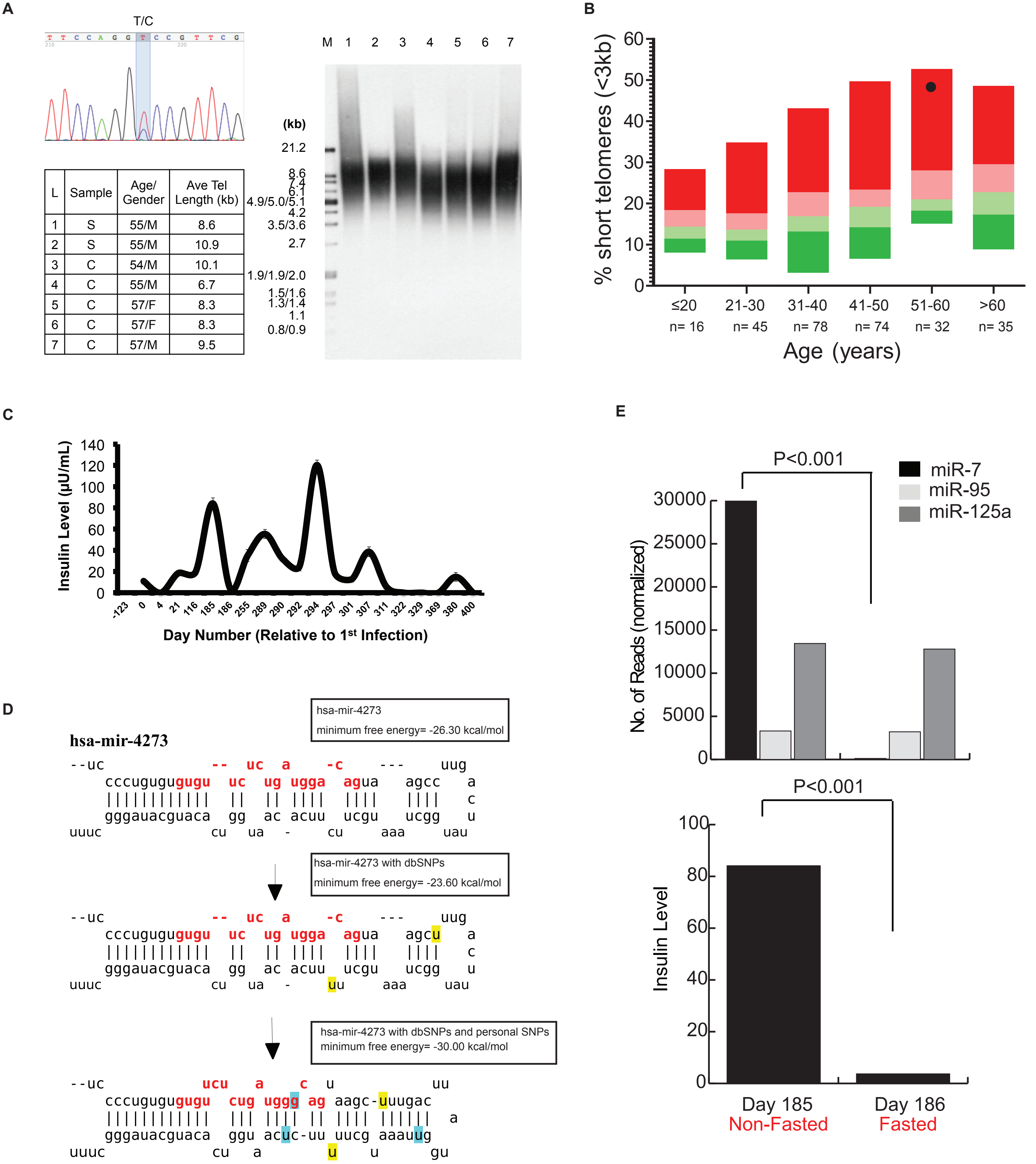

A
Color key for alignment scores
<40
40-50
50-80
80-200
>=200
Query
1
150
300
450
600
750
NM_000442.4 (PECAM1), mRNA
HuRef SCAF_1103279188238
blood_contig11281
810 bp
100 bp
200 bp
300 bp
400 bp
500 bp
600 bp
700 bp
800 bp
DNA seq coverage
RNA-seq coverage
Sequence
B
1
GM12878 A/G
GM12891 A/G
GM19099 G/G
GM18505 G/G
GM18526 G/G
GM10847 G/G
GM12892 G/G
GM18951 G/G
Chr5
83,593,500
83,594,000
RefSeq
83,593,661
EDIL3
2
GM19099 G/G
GM18505 G/G
GM18526 G/T
GM10847 G/T
GM12892 G/T
GM18951 T/T
GM12878 T/T
GM12891 T/T
Chr15
38,180,000
38,185,000
RefSeq
BMF
38,185,228
C
Isoforms per Gene at Each Time Point (FPKM >5)
8,522
1,970.4
495.1
138.1
38
11.75
4.3
2.85
0.65
0.4
0.1
Day
0
4
21
116
185
186
255
289
290
292
294
297
301
307
311
322
329
369
380
400
9,000
6,750
4,500
2,250
0
1
2
3
4
5
6
7
8
9
10
11
Isoforms per Gene



L	Sample	Age/ Gender	Ave Tel Length (kb)
1	S	55/M	8.6
2	S	55/M	10.9
3	C	54/M	10.1
4	C	55/M	6.7
5	C	57/F	8.3
6	C	57/F	8.3
7	C	57/M	9.5

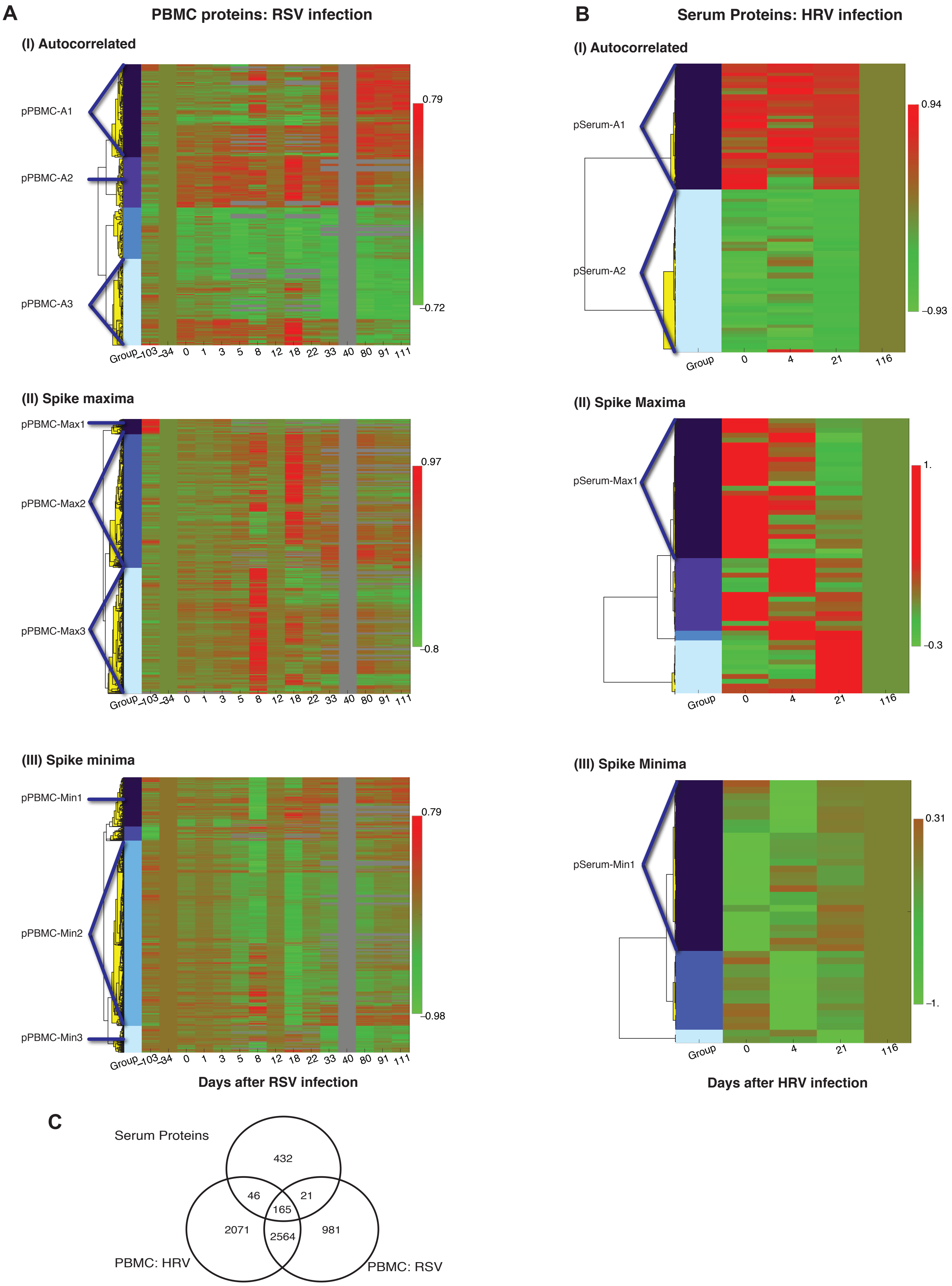
hsa-mir-4273

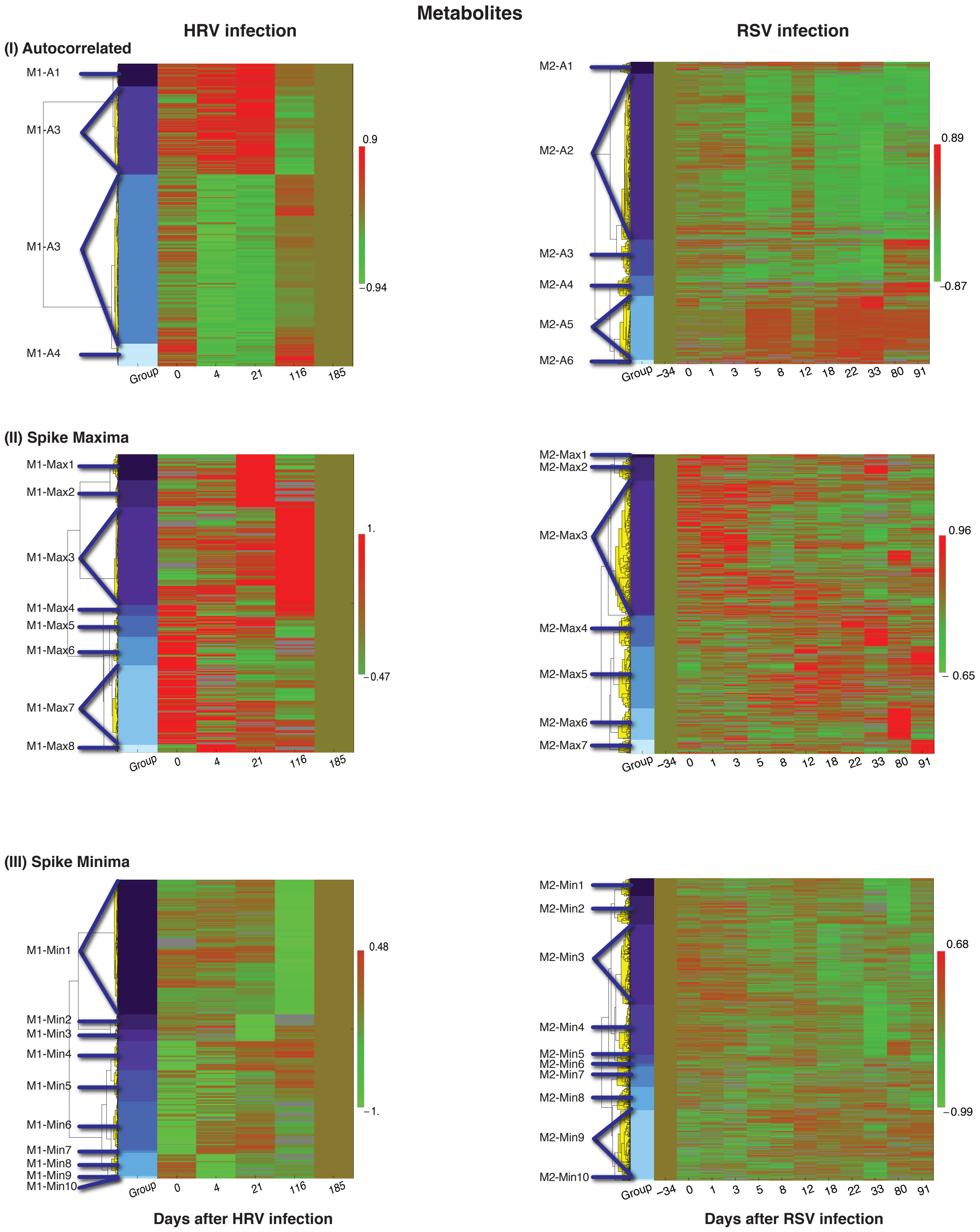
```
--uc            --  uc  a    -c    ---     uug
    cccugugugugu  uc  ug ugga  agua   agcc    a
    ||||||||||||  ||  || ||||  ||||   ||||    c
    gggauacguaca  gg  ac acuu  ucgu   ucgg    u
uuuc            cu  ua  -     cu    aaa     uau
```

```
--uc            --  uc  a    -c    ---     uug
    cccugugugugu  uc  ug ugga  agua   agcu    a
    ||||||||||||  ||  || ||||  ||||   ||||    c
    gggauacguaca  gg  ac acuu  ucgu   ucgg    u
uuuc            cu  ua  -     uu    aaa     uau
```

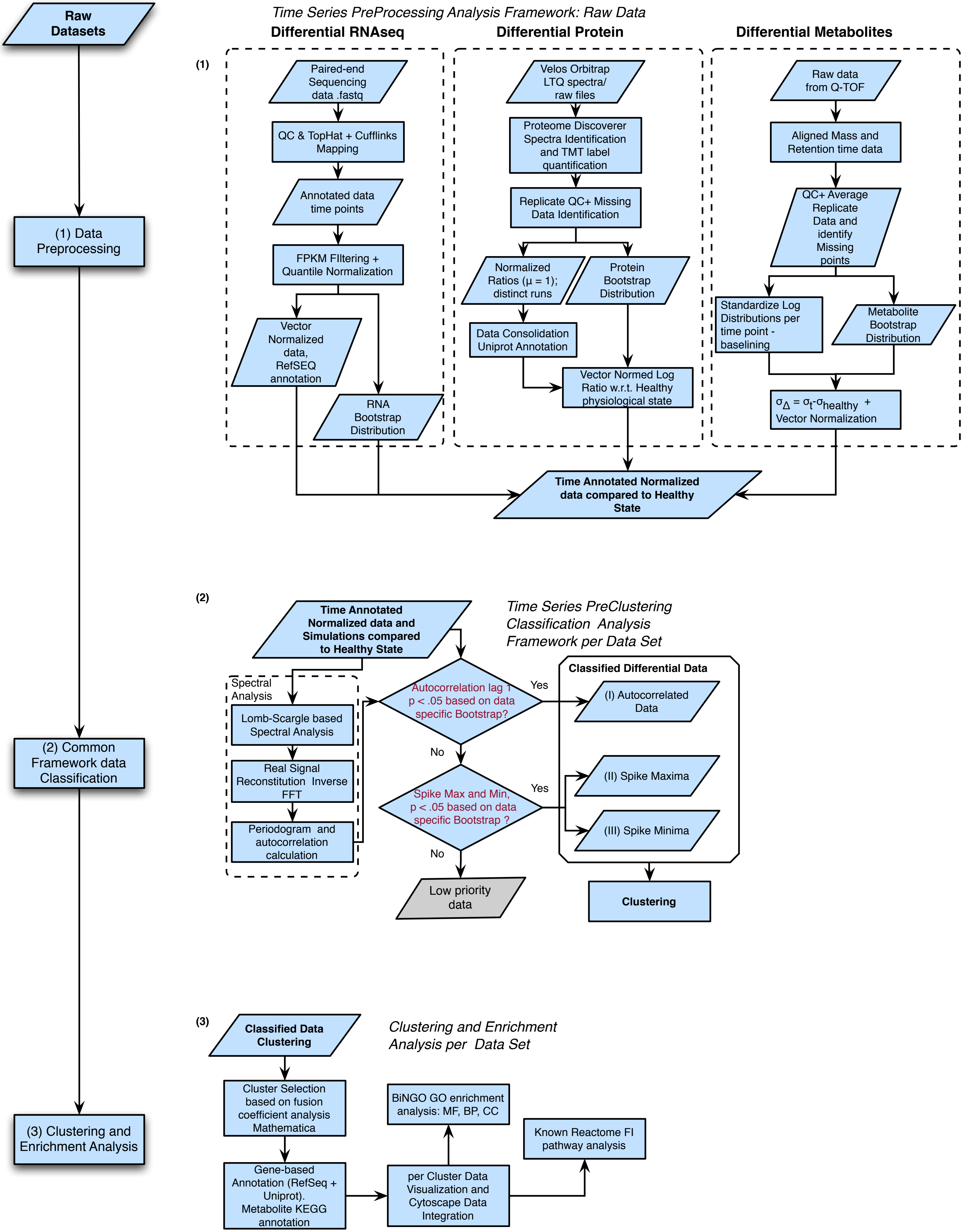
```
--uc            ucu  a    c  u           uu
    cccugugugugu  cug uggg ag aagc-uuugac
    ||||||||||||  ||| |||| || |||| ||||||  a
    gggauacguaca  ggu acuc-uu uucg aaauug
uuuc            cu  a       u    u       gu
```

Protein Classification and Clustering

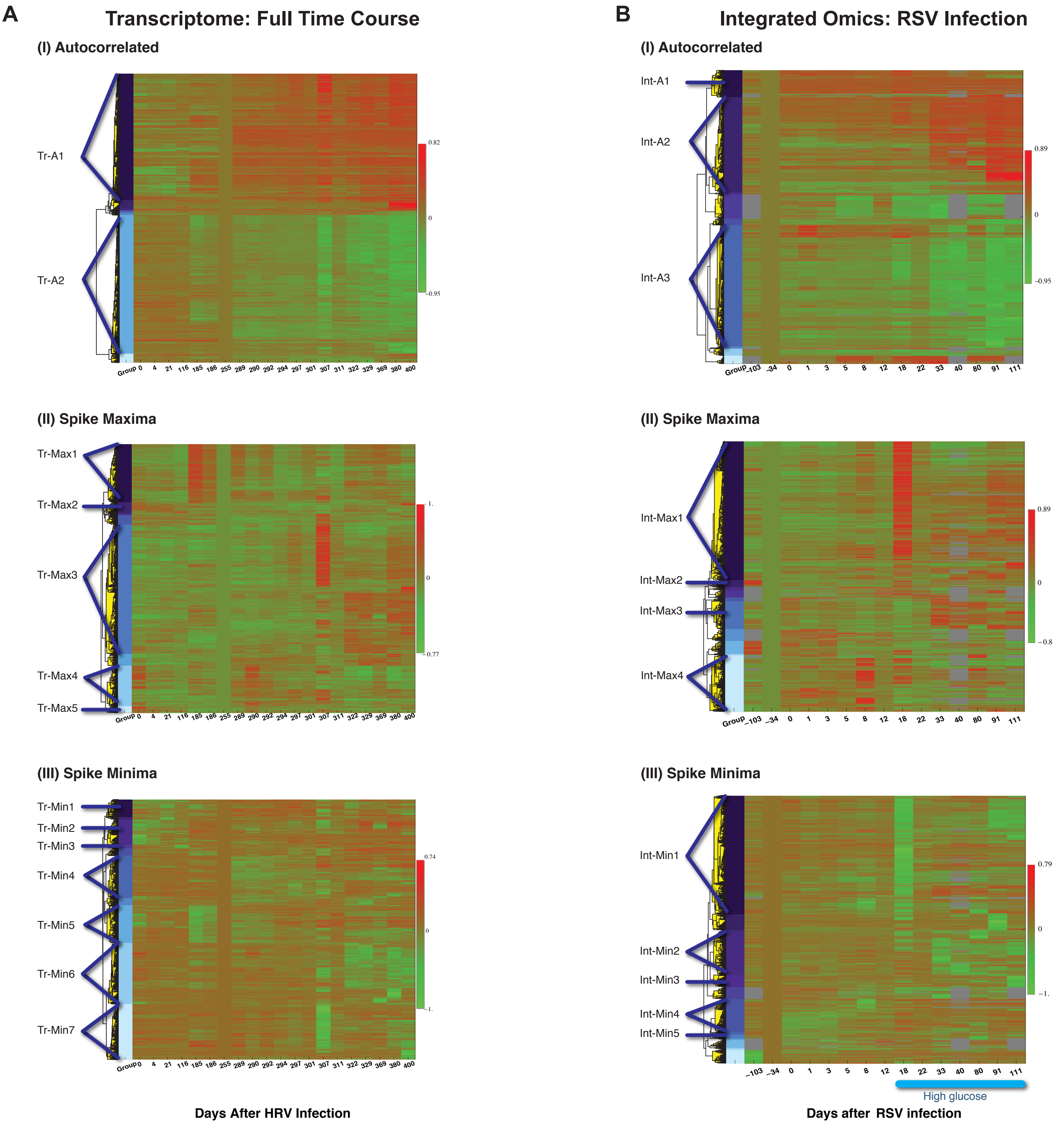



Metabolites
HRV infection
RSV infection
(I) Autocorrelated
M1-A1
M1-A3
M1-A3
M1-A4
0.9
−0.94
Group
0
4
21
116
185
M2-A1
M2-A2
M2-A3
M2-A4
M2-A5
M2-A6
0.89
−0.87
Group
−34
0
1
3
5
8
12
18
22
33
80
91
(II) Spike Maxima
M1-Max1
M1-Max2
M1-Max3
M1-Max4
M1-Max5
M1-Max6
M1-Max7
M1-Max8
1.
−0.47
Group
0
4
21
116
185
M2-Max1
M2-Max2
M2-Max3
M2-Max4
M2-Max5
M2-Max6
M2-Max7
0.96
− 0.65
Group
−34
0
1
3
5
8
12
18
22
33
80
91
(III) Spike Minima
M1-Min1
M1-Min2
M1-Min3
M1-Min4
M1-Min5
M1-Min6
M1-Min7
M1-Min8
M1-Min9
M1-Min10
0.48
−1.
Group
0
4
21
116
185
M2-Min1
M2-Min2
M2-Min3
M2-Min4
M2-Min5
M2-Min6
M2-Min7
M2-Min8
M2-Min9
M2-Min10
0.68
−0.99
Group
−34
0
1
3
5
8
12
18
22
33
80
91
Days after HRV infection
Days after RSV infection

Dynamic Data Analysis Framework

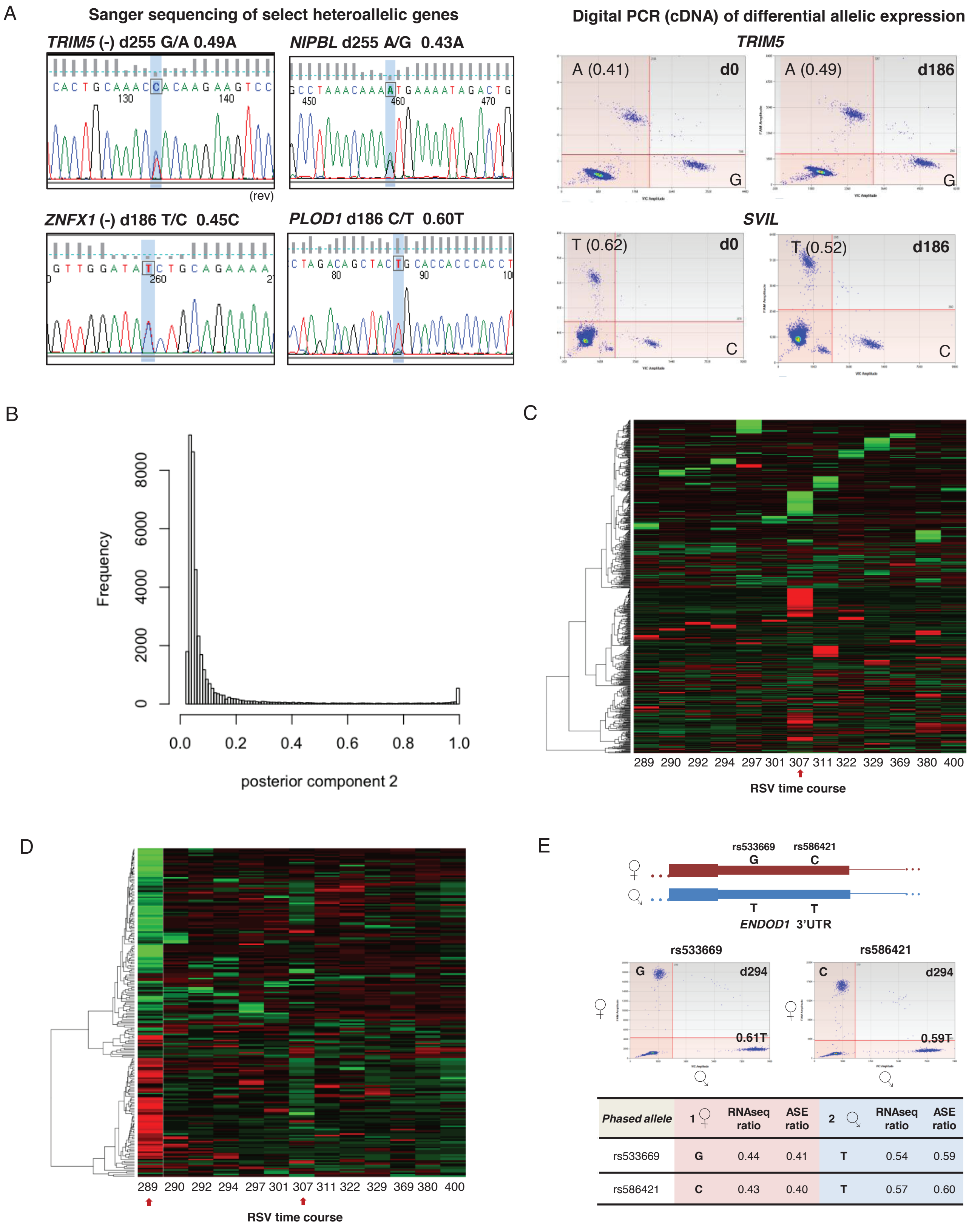

Raw Datasets
(1) Data Preprocessing
(2) Common Framework data Classification
(3) Clustering and Enrichment Analysis
Time Series PreProcessing Analysis Framework: Raw Data
Differential RNAseq
Differential Protein
Differential Metabolites
(1)
Paired-end Sequencing data .fastq
QC & TopHat + Cufflinks Mapping
Annotated data time points
FPKM FIltering + Quantile Normalization
Vector Normalized data, RefSEQ annotation
RNA Bootstrap Distribution
Velos Orbitrap LTQ spectra/ raw files
Proteome Discoverer Spectra Identification and TMT label quantification
Replicate QC+ Missing Data Identification
Normalized Ratios (μ = 1); distinct runs
Protein Bootstrap Distribution
Data Consolidation Uniprot Annotation
Vector Normed Log Ratio w.r.t. Healthy physiological state
Raw data from Q-TOF
Aligned Mass and Retention time data
QC+ Average Replicate Data and identify Missing points
Standardize Log Distributions per time point - baselining
Metabolite Bootstrap Distribution
$\sigma_{\Delta} = \sigma_{t}-\sigma_{healthy}$ + Vector Normalization
Time Annotated Normalized data compared to Healthy State
(2)
Time Annotated Normalized data and Simulations compared to Healthy State
Time Series PreClustering Classification Analysis Framework per Data Set
Spectral Analysis
Lomb-Scargle based Spectral Analysis
Real Signal Reconstitution Inverse FFT
Periodogram and autocorrelation calculation
Autocorrelation lag 1 p < .05 based on data specific Bootstrap?
Yes
No
Spike Max and Min, p < .05 based on data specific Bootstrap ?
Yes
No
Low priority data
Classified Differential Data
(I) Autocorrelated Data
(II) Spike Maxima
(III) Spike Minima
Clustering
(3)
Classified Data Clustering
Clustering and Enrichment Analysis per Data Set
Cluster Selection based on fusion coefficient analysis Mathematica
Gene-based Annotation (RefSeq + Uniprot). Metabolite KEGG annotation
per Cluster Data Visualization and Cytoscape Data Integration
BiNGO GO enrichment analysis: MF, BP, CC
Known Reactome FI pathway analysis

Classification and Clustering



A Transcriptome: Full Time Course

B Integrated Omics: RSV Infection



Phased allele	1 ♀	RNAseq ratio	ASE ratio	2 ♂	RNAseq ratio	ASE ratio
rs533669	G	0.44	0.41	T	0.54	0.59
rs586421	C	0.43	0.40	T	0.57	0.60

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Supplementary iPOP results (related to Figure 1) (A) Representative genomic region not present in the reference genome (hg19): An assembled genomic contig containing the gene *PECAM1* is shown. This region was discovered by contig assembly of the unmapped reads from WGS. Top track, WGS reads mapped to the contig; bottom track, RNA-Seq coverage of reads mapped uniquely to this contig. (B) Genomic variants affecting Transcription Factor binding sites: (B1) In *EDIL3,* the ancestral allele G, homozygous in 6 of the 8 samples and the subject, disrupts the motif whereas the allele A promotes binding of NFkB. (B2) In *BMF* the subject is also homozygous for a T allele disrupting the NFkB motif at rs539846 that lies in the first exon of the gene. (C) Isoforms per gene: At every time-point the number of isoforms detected for every Official Gene Symbol was computed.

Figure S2. Supplementary medically relevant results (related to Figure 2) (A) *TERT* A202T mutation and telomere length assay for PBMCs: Left top panel shows Sanger sequencing result displaying the heterozygous A202T mutation (T/C). Left bottom table shows Telomere length assay sample and result summary. Lanes 1 and 2, PBMC DNA from the subject at Day 255 and Day 292, respectively; Lanes 3-7, healthy controls that are free of this mutation. S, Subject (the volunteer subject); C, Control. Right panel shows Telomere length assay southern blot result. (B) Percentage of chromosomes with short telomere (< 3 kb) as determined by High-Throughput Q-FISH. The green, light green, peach and red colors represent the first, second, third and fourth quartile in each

age group, respectively. The black dot represents the test subject. (C) Insulin ELISA: Plasma insulin concentration at each time point was determined by ELISA. Day numbers were shown relative to the first day of the HRV infection. (D) Personal SNPs that lead to compensatory changes in hairpin in miR4273: For the pre-miRNA, *hsa-mir*-4273, the SNPs presented are from the dbSNP database. (E) Correlation of the levels of miR-7 with the plasma insulin, where miR-95 and miR-125a are shown as controls.

Figure S3. Protein Classification and Clustering, related to Figure 4. Dynamic protein data was grouped into (I) autocorrelated, spike maxima (II) and minima (III) classes and clustered hierarchically shown here for: (A) PBMC proteins, following the dynamics of the RSV infection and high glucose onset - for each labeled cluster, enrichment analyses may be found in Data S3. (B) Serum proteins, following the dynamics of the HRV infection - for each labeled cluster, enrichment analyses may be found in Data S5. (C) The overlaps between identified serum proteins (HRV infection time course) and PBMC proteins were determined (HRV and RSV infection time courses).

Figure S4. Clustering for metabolites related Figure 4. Metabolite data following separately the dynamics of the HRV and RSV infection and high glucose onset was grouped into (I) autocorrelated, spike maxima (II) and minima (III) classes and clustered hierarchically. For each labeled cluster, associated metabolites may be found in Data S4.

Figure S5. Integrated omics analysis framework, related to Figures 3-4. Different

omics data are analyzed accordingly with a view towards data integration through a common framework.

Figure S6. Supplementary Clustering Details, corresponding to Figures 3-4. Dynamic data was grouped into (I) autocorrelated, spike maxima (II) and minima (III) classes and clustered hierarchically shown here for: (A) Transcriptome data for the duration of the project; For each labeled cluster, enrichment analyses may be found in Data S6. (B) Integrated omics data (transcriptome, proteome and metabolome) following the dynamics of the RSV infection and high glucose onset. For each labeled cluster, enrichment analyses and associated metabolites may be found in Data S7.

Figure S7, related to Figure 5. Heteroallelic expression and editing in PBMCs.

(A) Sanger cDNA sequencing of selected heteroallelic expressed genes confirms heterozygosity but not the ratio of alternate allele (left), while digital PCR is utilized to validate differential allelic specific expression (ASE) alt/tot ratios across day 0 and day 186 time points (right). (B) Distribution of the posterior probability of allelic specific expression (ASE) based on the two-component beta binomial distribution model. The posterior probability is the observation that, X_{mt} is derived from the second component is interpreted as the strength of the ASE. (C) Heatmap of the RSV infection time course (min. 10 time points, 684 sites, posterior probability >0.75 at least at one time point) showing differential ASE with distinct patterning during onset of T2D, Day 307 (red arrow). (D) Heatmap of the RSV infection time course (min. 10 time points, 258 sites, posterior probability >0.75 at Day 307) showing differential ASE focused on Day 289 (onset of RSV infection - red arrow), onset of T2D at day 307 is also shown (red arrow). (E) Two adjacent allele-specific phased variants in *ENDOD*1 3'UTR show concordance in alt/tot expression using digital PCR of cDNA.

SUPPLEMENTAL TABLES

Table_S1. Genomic variants and validations - related to Figure 1

Table_S2. Medically Relevant Variants - related to Figure 2

Table_S3. Summary and Breakdown of RNAseq analysis in PBMCs - related to Figure 5

Table_S4. Summary of Phased variants in an individual – related to Figure 5

EXTENDED EXPERIMENTAL PROCEDURES

Sample Collection

The subject and mother in this study were recruited under the IRB protocol IRB-8629 at Stanford University. Whole blood samples were collected at each time point and Peripheral Blood Mononuclear Cells (PBMCs) were isolated by density gradient centrifugation at 400 x g for 25 minutes using the Lymphocyte Separation Media (MP Biomedicals). Serum and plasma were also collected for each time point. Genomic DNA and RNA were isolated from the PBMCs using the AllPrep DNA/RNA/Protein Mini Kit (QIAGEN). Protein was also prepared from lysed PBMCs for mass spectrometry with the Lysis Buffer (4% SDS, 100mM Tris-HCl pH7.6, 100 mM DTT).

Human Rhinovirus (HRV) and Respiratory Syncytial Virus (RSV) Detection

HRV and RSV were detected from upper respiratory swab samples from the subject at the Stanford Hospital and Clinics with standard assays (the Respiratory Viral Panel Test). Briefly, viral RNA was extracted from the swab samples, amplified with Reverse Transcription-Polymerase Chain Reaction, and the presence of a panel or respiratory viruses were detected using the Luminex® xTag^{TM} technology. For HRV infection, samples from Days 0, 4, and 21 were examined; and for RSV infection, samples from Days 289, 290, 292 and 294 were assayed.

Whole Genome Sequencing

Whole genome sequencing was performed at both Complete Genomics Inc. (Mountain View, CA) and Illumina, Inc. (San Diego, CA). Ten micrograms of genomic DNA was used for each platform. Paired-end 35b reads were used for Complete Genomics (CG) sequencing, and data were processed and variants (SNVs, Indels, SVs and CNVs) were called using the NCBI reference genome build 37 with the CG assembly software v1.10.1.32. For Illumina, 101b paired-end sequencing data were obtained using Illumina's HiSeq 2000 Sequencer. Illumina data were processed with the HugeSeq pipeline we developed for this project (Lam et al., submitted to Nature Biotechnology). This pipeline maps reads using Burrows-Wheeler Aligner (Li and Durbin, 2009) (BWA), calls SNVs, indels, and SVs using the algorithms presented in the text. SVs detected by two or more methods were called high confidence.

Whole Exome Sequencing

Whole exome sequencing was performed using three available platforms: the Agilent SureSelect All Exon 50Mb, the Nimblegen SeqCap EZ Exome Library v2.0, and the Illumina TruSeq Exome Enrichment Kit (Clark et al., 2011). Three micrograms of genomic DNA was used for each enrichment platform. The enriched sequencing libraries were prepared according to the manufacturer's protocols with slight modifications as stated below, and were each subjected to Illumina sequencing on one lane of the HiSeq 2000 sequencer.

For the Agilent platform, genomic DNA was sheared with the Covaris S2 system; the DNA fragments were end-repaired, extended with an "A" base on the 3' end, ligated with paired-end adaptors, and amplified (4 cycles). Exome-containing adaptor-ligated libraries were hybridized for 24 hours with biotinylated oligo RNA baits, and enriched with Streptavidin-conjugated magnetic beads. The final libraries were further amplified for 11 cycles with Polymerase Chain Reaction (PCR).

For the Nimblegen SeqCap EZ-Exome Library, Illumina sequencing library was made following Nimblegen's protocol with the following improvements: in Chapter 4 Steps 1-4 of the protocol two PCR reactions were set up for each sample with 15 microliters of each unenriched sample library as template, and 2 micrograms of amplified sample library was used for each sample in the hybridization step described in Chapter 5 Step 2. These modifications ensure that we obtain sufficient material from PCR for the hybridization, and by doubling the amount of amplified sample library we make the most use of the enrichment probes. Briefly, DNA fragmented with the Covaris S2 system was concentrated with ethanol precipitation, end-repaired with the Epicentre End-It^{TM} DNA End-Repair Kit, a deoxyadenosine was added at the 3' end of the fragments with the Klenow 3'->5' exo- enzyme (New England Biolabs), and ligated with Illumina's Paired-End Adaptor Oligo Mix (Part# 1001782). The ligated libraries were size selected for an average insert size of 250 bp (2 mm gel slice) by agarose gel excision and extraction, amplified for 8 cycles by Pre-Capture LM-PCR, and hybridized for 72 hours with biotinylated oligo DNA baits for exome-containing libraries. The hybridized libraries were enriched with Streptavidin-conjugated magnetic beads and washed and amplified by PCR (18 cycles), and the quality of the libraries was checked by qPCR as described in the protocol.

For the Illumina TruSeq Exome Enrichment Kit, Pre-enrichment DNA libraries were constructed following Illumina's TruSeq DNA Sample Preparation Guide. A 300-400bp band was gel selected for each library and exome enrichment was performed according to Illumina's TruSeq Exome Enrichment Guide. Two 20-hour biotinylated bait-based hybridization were performed with each followed with Streptavidin Magnetic Beads binding and a washing step and an elution step. A 10-cycle PCR enrichment was performed after the second elution and the enriched libraries were subjected to Illumina sequencing after quality check on one lane of HiSeq 2000.

Sanger DNA Sequencing

Sanger DNA PCR and sequencing primers were designed manually and with the Optimus Primer software (http://op.pgx.ca/), and were synthesized at Integrated DNA Technologies (Coralville, IA). DNA sequencing was performed at ELIM BIOPHARM (Hayward, CA). Sequencing results were visualized with the CodonCode Aligner software (http://www.codoncode.com/aligner/).

Whole Transcriptome Sequencing (mRNA-Seq)

Strand-specific RNA-Seq libraries were prepared as described previously (Parkhomchuk et al., 2009). Briefly, 9 micrograms of total RNA isolated from PBMCs were used and mRNA was enriched with the Dynal Oligo (dT) beads (Invitrogen). The isolated mRNA was fragmented using the RNA Fragmentation Reagents (Ambion) and cDNA containing dUTP in the second strand was synthesized. The cDNA molecules were end-repaired with the Epicentre End-It^{TM} DNA End-Repair Kit, a deoxyadenosine was added at the 3' end of the fragments with the Klenow 3'->5' exo- enzyme (New England Biolabs), and ligated with Illumina's Paired-End Adaptor Oligo Mix (Part# 1001782). The ligated libraries were size selected for an average insert size of 250 bp (2 mm gel slice) by agarose gel excision and extraction, and the dUTP-containing second strands were digested with Uracil-DNA Glycosylase (New England Biolabs). The treated libraries were then amplified by Polymerase Chain Reaction at the following conditions: 98°C 30 sec, 15 cycles of (98°C 10 sec, 65°C 30 sec, 72°C 30 sec), 72°C 5 min. Each prepared library was sequenced on 1-3 HiSeq 2000 lanes to obtain an average of 123 million uniquely mapped reads (20 time points). The TopHat package (Trapnell et al., 2009) was used to align the reads to the hg19 reference genome, followed by Cufflinks (Trapnell et al., 2010) for transcript assembly and RNA expression analysis. The number of redundant reads was low (7.78%). The Samtools package (Li et al., 2009) was used to identify variants including single nucleotide variants (SNV) and indels.

Small RNA Sequencing (microRNA-Seq)

MicroRNA were isolated from 10 million PBMCs at five time points (Days 4, 21, 116, 185 and 186 from HRV infection) with the mirVana^{TM} miRNA Isolation Kit (Ambion). microRNA-Seq libraries were prepared from 1 microgram of isolated miRNA according to Illumina's Small RNA v1.5 Sample Preparation Guide. Each library was sequenced with 36b single-end reads on 1 lane of Illumina's GAIIx sequencer.

The human pre-miRNAs, miRNAs sequences were extracted from miRBase release17 [hg19]. The SOAP program (Li et al., 2008) was used to map sequence reads with a maximum of 2 bp mismatches to the hairpin sequences. miRanda algorithm (John et al., 2004) and TargetScan version 5 (Lewis et al., 2005) were used for targets prediction. For miR-7, 323 targets were predicted with TargetScan program, and 240 of 323 were expressed. 65 expressed mRNAs fit the profile of miRNA expression along each time points tested. There are at least 108 additional mRNAs targets that were associated with diabetes predicted with miRanda. DAVID program (Dennis et al., 2003; Huang et al., 2008) was used for pathway enrichment analysis. To examine the significance of gene–term enrichment, the program uses a modified Fisher's exact test (EASE score). The enrichment P-values are globally corrected for multiple hypothesis testing using Benjamini (Huang et al., 2008). Cluster 3 (Eisen et al., 1998) was used to perform the Hierarchical cluster categories of mRNA targets. The Java TreeView program (Saldanha, 2004) was then used to visualize these clusters.

PBMC and Serum Shotgun Proteome Profiling

A. Protein extraction and labeling using TMT

The PBMC cell pellets were lysed in 10x volume of buffer containing 4% SDS and 100 mM dithiotreitol in 100 mM tris-HCl pH 8.0. Lysates were incubated at 95 °C for 5 min and briefly sonicated. Detergent was removed from the lysates using the FASP protocol (reference) using YM-30 microcon filter units (Cat No. MRCF0R030, Millipore). In brief, 200 μL of 8 M urea in 0.1 M Tris/HCl, pH 8.5 was added and samples were centrifuged at 14 000xg at 20 °C for 15 min. This step was repeated 3 times. Then 50 μL of 0.05 M iodoacetamide in 8 M urea was added to the filters and the samples were incubated in darkness for an hour. Sample was washed 3 times with 100 μ L of 200 mM ThAB. Protein concentration was measured using Bradford method. Finally, trypsin (Promega, Madison, WI) was added at protein to enzyme ratio of 50:1. Samples were incubated overnight at 37 °C. Peptides were collected by centrifugation and labeled using TMT 6plex reagent. Immediately before use, equilibrate the TMT label reagents to room temperature. For the 0.8 mg vials, 41 µl of anhydrous acetonitrile were added to each tube and 41 µl of the TMT Label Reagent was then added to each 25-100 µg sample. The reaction was incubated for 1 hour at room temperature. To quench the reaction, 8 µl of 5% hydroxylamine was added to the sample and incubated for 15 minutes. Samples were combined at equal amounts and dried by speed vac.

For serum proteome, the 14 most abundant proteins were depleted using an Agilent Mars human 14 column (4.6 mm x 50mm). The unbound fraction from the column was collected for further proteome analysis. The protein sample was then processed as described above.

B. Peptide separation

A highly reproducible online Waters 2D liquid chromatography (Waters NanoAquity 2D nLC) was used for peptide separation. The protein sample was first resuspended in 100mM ammonium formate at pH10 and then loaded to the LC system. Peptides were separated by reverse phase chromatography at high pH in the first dimension, followed by an orthogonal separation at low pH in the second dimension. An online dilution of the effluent was performed after the first dimension to ensure no peptides were lost prior to the second dimension. In the first dimension the mobile phases were buffer A: 20mM ammonium formate at pH10 and buffer B: Acetonitrile. Peptides were separated on a Xbridge 300 μ m x 5 cm C18 5.0 μ m column (Waters) using 14 discontinuous step gradient at 2 μ l/min. Acetonitrile concentration for each step was adjusted to ensure nearly equivalent peptide load and MS intensity for each second-dimension run. Since peptide fractions eluted from the first dimension column was at high pH and differing Acetonitrile concentrations, they were not compatible with the second dimension separation. To maximize peptide recovery the fractions were diluted online using 0.1% formic acid in water at 20 μ l/min and then trapped by Symmetry 180 μ m x 2cm C18 5.0 μ m trap column (waters). In the second dimension, peptides were loaded to a in-house packed 75 μ m ID/15 μ m tip ID x 20cm C18-AQ 3.0 μ m resin column with buffer A (0.1% formic acid in water). Peptides were separated with a linear gradient from 5% to 30% buffer B (0.1% formic acid in acetonitrile) at a flow rate of 300 nl/min in 180 min. Each sample separation was repeated 3 times.

C. Proteomics MS analysis

The LC system was directly coupled in-line with a linear trap quadrupole (LTQ)-Orbitrap Velos instrument (Thermo Fisher Scientific) via Thermo nanoelectrospray source. The source was operated at 2.2-2.4 kV to optimize the nanospray, with the ion transfer tube at 200 °C. The mass spectrometer was run in a data dependent mode. One survey scan acquired in the Orbitrap mass analyzer with resolution 60,000 at m/z 400 was followed by MS/MS of the 10 most intense peaks with charge state ≥ 2 and above an intensity threshold of 5000. MS/MS fragmentation was done in the high collisional Cell (HCD) with normalized collision energy of 40% and activation time of 0.1s. The MS/MS scan was acquired in the Orbitrap at resolution of 7,500. For all sequencing events dynamic exclusion was enabled to minimize repeated sequencing. Peaks selected for fragmentation more than once within 30s were excluded from selection (10 ppm. window) for 60s.

D. Proteomics data processing and analysis

The raw data acquired were processed with the Proteome Discoverer (Thermo). IPI human database, v. 3.75 (Kersey et al., 2004) was used. Mass tolerance of 10ppm was used for precursor ion and 0.02 Dalton for fragment ions for the database search. The search included cysteine carbamidomethylation as a fixed modification. N-terminal and lysine TMT 6plex modification and methionine oxidation were used as variable modifications. Up to two missed cleavages were allowed for trypsin digestion. Only unique peptides with minimum 6 amino acid length were considered for protein identification. The false discovery rate (FDR) was set as less than 1% and we required two unique peptides per protein for identification. For peptide quantitation, only unique peptides with reporter ion mass tolerance of less than 10ppm were used. The median value of different peptide ratios was used for protein quantitation. Downstream analysis of proteomics is described below.

Serum Metabolome Profiling

A. Serum metabolite extraction

100 ul of serum sample was used for metabolomics study. Metabolites were extracted by adding 4 times volume of equal volume mixture of methanol, acetonitrile and acetone that were pre-chilled at -20°C. To maximize metabolites extraction, samples were vortex at 4°C for 15 min at 2 min interval. Proteins were precipitated by incubating the sample

at -20°C for 2 hours. Samples were then centrifuged at 10,000 rpm at 4° for 10min. The supernatant was collected and dried for metabolomics analysis. For each time point, 3 of the 100 ul samples were analyzed in triplicate.

B. Metabolomics LC MS analysis

An Agilent 1260 Liquid Chromatography system was directly coupled in-line with an Agilent 6538 accurate mass Q-TOF MS with electrospray ionization (ESI) operated at positive and negative mode. The LC mobile phases consisted of 0.2% acetic acid in water (buffer A) and 0.2% acetic acid in methanol (buffer B). The extract was resuspended in 50% methanol and sonicated for 5min. Sample was loaded to an Agilent SB-aq 1.8 μ m, 2.1 x 50 mm analytical column with a SB-C8 3.5 μ M, 2.1 x 30 mm guard column in front. Columns were heated to 60°C with a flow rate of 0.6 ml/min. A linear gradient from 2% to 98% buffer B in 13min was used for metabolites separation. To assure the mass accuracy of the recorded ions, continuous internal calibration ions were infused in-line through the dual ESI source using an isocratic pump at flow rate of 0.05ml/min. Internal calibrants at m/z 121.0509 and 922.0098 were used in positive ion mode and m/z of 119.0362 and m/z 980.0164 were used in negative ion mode.

The Q-TOF was operated at source condition of 3,750V with drying gas 9 L/min and nebulizer gas 45 psi at 300°C. The instrument was run at extended mass range to1700 m/z. The fragmentor voltage was set at 125V and skimmer at 47V. The data was acquired at scan rate of 1.5spectra/sec for MS. MS/MS was run at targeted mode at scan rate of 3 spec/sec with 10 spec/sec for MS. Collision energy of 20V and a fixed isolation window of 4 m/z and retention time window of 0.25 min were used for the targeted

MS/MS. Each sample was run at MS mode first at both positive and negative modes and the differentially expressed metabolites were selected for MS/MS experiment.

C. Metabolomics data processing and analysis

MassHunter Workstation software (Agilent Technologies), including Qualitative Analysis (version 3.01) and Mass Profiler Professional (MPP version B.02) were used to process both MS and MS/MS data. The Molecular Feature Extractor (MFE) in Qualitative analysis software were used to search for features that have common elution profile and groups ions into one or more compounds containing m/z values that are related (peaks in the same isotope cluster, different adducts or charge states of the same entity). The results were exported as files in Compound Exchange Format (CEF files) for further analysis in MPP. MPP was used to align data from different samples, filter data for statistical analysis and database search. For the chromatography alignment, only ions with intensity above 5,000 and retention time window within 0.2 min were selected. If ions were not present in all the files, they were filtered out. For samples from the same time point, the median value was used for that time point. Further statistical analysis was done to find the differentially expressed compounds as described below. METLIN human metabolites database was used for the database search. Mass tolerance was set at 10ppm.

Serum C-Reactive Protein and Plasma Insulin Enzyme-Linked ImmunoSorbant Assays

Serum C-Reactive Protein (CRP) levels were quantitated with the hsCRP ELISA Kit from Abnova following the manufacturer’s instructions. Plasma Insulin levels were measure with the Human Insulin (Animal Serum Free) ELISA Kit (Millipore) according to the manufacturer's protocol.

Serum Cytokine Profiling

Serum cytokine profiling was performed with the Luminex 51-plex Human Cytokines bead-based assay at the Stanford Human Immune Monitoring Center with the Luminex 200 Instrument. The analytes are listed in Figure 2F plus IL-6 (which was not detected in all the samples for 2 repeated runs). One hundred microliters of serum were used for each time point.

Blood Glucose, Glycated HbA1c and Triglyceride Measurement

Blood glucose, Glycated HbA1c, and triglyceride levels were measured at the Laboratory of the Stanford Hospital and Clinics, if not otherwise stated, along with other standard lipid and chemistry profiles not covered in this manuscript. Glucose levels were measured with the ACCU-CHEK system for Days 363-602 (except Days 369, 476, 532, 546 and 602). The moving average was shown with a window of 15 days (7 days prior and post each time point) in Figure 2D. Duplicate measurements were taken for 13 time points using ACCU-CHEK as well as for Days 322 and 369, with a variance typically less than 3% and never more than 5%.

Autoantibodyome Profiling

Autoantibodyome profiling was completed for 4 time points (Days -123, 0, 4 and 21) using the Invitrogen ProtoArray Protein Microarray v5.0 (which contain 9,483 unique human proteins spotted in duplicate), according to the manufacturer's instructions and as described previously (Hudson et al., 2007). Thirty-four healthy plasma samples were used as controls. Plasma samples were diluted 1:100 in 5 ml Washing Buffer (1X PBS, 0.1% Tween 20, 1X Roti-Block) for the autoantibodyome profiling. The probed protein microarray chips were dried and scanned with the Genepix 4200AL Microarray Scanner (Molecular Devices, Sunnyvale, CA) the Genepix Pro 6.1 software. The arrays were scanned to obtain signal location, intensity quantification and identification information (.gpr format) using GenePix Pro 6.1 (Molecular Devices). For each array inter-array normalization was performed via the ProCAT algorithm (Zhu et al., 2006) (sliding window of length 15). The arrays were then quantile normalized (Bolstad et al., 2003) and a comparison of intensities of probes was carried out between the subject and the healthy control group using a two-tailed Mann-Whitney non-parametric test, $p<0.01$, in *Mathematica 8.0* and using Benjamini-Hochberg (Benjamini and Hochberg, 1995) to correct for multiple hypothesis tests, adjusted $p<0.01$ (Data S2). Biological replicates were compared for reproducibility showing a high degree of correlation across slides with $R^2>0.894$, and the Coefficient of Variation (CV) across slides had median value 0.0656 and 96.6% of spots having CV <1 (Data S11C.I.1-2). For protein spots duplicated on the arrays we found $R^2 =0.99$ and median CV 0.04 with 96.5% of signals having CV <1. (Data S11C.I.3-4).

Telomere Length Assay

Telomere length in the PBMCs of the volunteer subject was measured and calculated with both Southern Blotting and the High-throughput Q-FISH. Southern Blotting of telomeres was performed using the Telo TAGGG Telomere Length Assay Kit (Roche) following the manufacturer's instructions. Telomere length at two time points was investigated (Days 255 and 292) to reveal potential telomere length differences for healthy and infected states. X-ray images were digitized with the Typhoon scanner (GE Healthcare) and analyzed with the ImageJ software (Abramoff et al., 2004).

High-Throughput Q-FISH (HT Q-FISH) was performed using mononuclear cells isolated from peripheral blood using a ficoll separating solution (LymphoprepTM). The cells were then plated on a clear bottom black-walled 96-well plate, and HT-QFISH was performed as previously described (Canela et al., 2007). Telomere length values were analyzed using individual telomere spots corresponding to the specific binding of a Cy3 labeled telomeric probe (subject: 5.41kb compared to age group median: 5.95 kb). Fluorescence intensities were converted into kilobases as previously described (Canela et al., 2007; McIlrath et al., 2001). Each median telomere length value was calculated and plotted. Linear regression analysis was used to assess the correlation between age and median telomere length or percentage of nuclei with telomeres <3kb in lymphocytes of the donors. Median telomere length values and percentage of telomeres <3kb (short telomeres) of donors in the indicated age groups were calculated. The number of samples of each group is indicated (n). The minimum, 25th percentile, median, 75th percentile and the maximum values from each age group were calculated and used to create four equal groups, each representing a fourth of the distributed sampled population. GraphPad Prism has been used for data calculation. See Figure S2A-B.

Genome Phasing

Single nucleotide variants (obtained from a minimum 2 platforms) and indels (from 3 platforms) of the individual's DNA were phased as summarized in Data S11D [see also Figure S7E, Table S4, Data S10]. This variant list was augmented with maternal sequence and genotype data, as well as with the phased CEU haplotypes from the 1000 Genomes Project. For variants that are observed in both the subject and in 1000 Genome haplotypes, phasing was achieved using the program BEAGLE (Browning and Browning, 2007). The maternal genotype is provided to BEAGLE only if the call is high confidence (also from minimum of 2 platforms), otherwise the data is considered as missing. Novel variants not observed in the 1000 Genome haplotypes are phased based on a Mendelian inheritance pipeline and the maternal genotype alone. The two datasets are then merged, followed by correction by any experimental data (including data from Complete Genomics on haplotyping and Paired-End Sequencing if available). The inferred maternally- and paternally-derived haploid genome was then analyzed with programs Polyphen-2 (Adzhubei et al., 2010) to identify the biological impact of the phased variants. A secondary pipeline was developed to identify compound heterozygous variants, which tags the genes that accumulate variants (SNPs and indels) found on different alleles. This study focused on compound missense and nonsense mutations, which may potentially be damaging. Those identified genes are further categorized into three compound heterozygous types: Type 1- Genes with at least one heterozygous variant on each allele, Type 2- Genes with both homozygous and accumulated heterozygous variants on each allele, and Type 3: Homozygous variants with additional heterozygous variant(s) on only allele 1 (Type 3A) or on only allele 2 (Type 3B) (See also Table S4).

Variants identified in RNA (Heteroallelic expression and RNA editing)

Variants in RNAseq data were identified using Samtools (Li et al., 2009), as described above, and compared against the hg19 reference genome. The RITE-2-seq (RNA Identifier Tool for Expression and Edits) pipeline was developed to identify RNA variants as summarized in Data S11A. A minimum of 40 unique reads (as well as 10 unique reads) were obtained at a variant position and compared to the high and low single nucleotide genomic calls (as described above). Those variants that matched DNA were subsequently characterized as heterozygous or homozygous (Table S3) and heterozygous calls were analyzed for differential allelic specific expression (ASE). Variants that were not in the genome were deemed as candidate RNA edits, and were further filtered to remove false positives due to misalignments (multigene families and pseudogenes), as well as 'close proximity' variants (errors likely due to an alignment to an uncharacterized novel isoform; mapping errors accumulated within a window of 10 bp were removed). These candidates were also re-compared to both low and high confidence exome data (described above), as to remove any extra DNA based variants (high confidence candidate RNA edits summarized in Data S8). Polyphen-2 and ANNOVAR (Wang et al., 2010), as well as in-house developed callers, were used to localize the variants to genic regions, and those identified as missense calls were further used in this omics study to validate corresponding variant transcripts at the protein level (further described below). RNA realignment with the corrected personalized genome and corrected transcriptome (see Data Dissemination Section below for availability) will aid particularly in improving mapping reference bias.

To evaluate differential allele-specific expression (ASE) at each site, we used a two-component beta binomial mixture model (similar to that used in Skelly et al. 2011). Under this model, the number of observed non-reference allele, X_{mt}, given the total read depth, N_{mt}, is assumed to have a binomial distribution, Binom(N_{mt},p_{mt}). With probability $1-\pi$, p_{mt} is drawn from a beta distribution, Beta(α,α), and with probability π, it is drawn from a second beta distribution, Beta(δ,δ), such that $\delta<\alpha$. The parameters α, δ and π are estimated by maximizing the likelihood function. For the first infection cycle, $\hat{\alpha}=78$, $\delta=4$, and $\pi=0.11$; the second infection cycle is more overdispersed, $\hat{\alpha}=45$, $\hat{\delta}=2.4$, $\pi=0.17$. The posterior probability that the observation, X_{mt} is derived from the second component is interpreted as the strength of allelic-specific expression. The distribution of this posterior probability is shown in Figure S7B and though most sites reveal no differential ASE, a few sites and time points show convincing evidence that the ratio is not (50%, 50%). We also estimated a shrunk ASE ratio (alternate allele count / total count ratio; alt/tot) for each data point by a weighted average under the two components. For Figure 5C-D, alt/tot ratios from infection states (Day 0 and Day 289) were compared to uninfected states Days 116-255 and Days 311-400, respectively. All heatmaps and histogram analysis were performed using the rescaled shrunken ASE ratios, with a minimum coverage of 40 reads (RNA-seq) across a minimum of 5 and 13 time points for HRV and RSV infections, respectively.

Heatmaps examining differential ASE were generated using R program (version 2.13.1), where missing data points were imputed using row means (for multiple points) and the k nearest neighbour (for single points) method. Single and average linkage hierarchical clustering with application of the Pearson correlation distance metric was performed for the heatmaps. These figures contain all variant positions, including missense, synonymous and UTR locations. All heatmaps are based on the ratio of the alternative allele or edited nucleotide to total expression (alt/tot). Genes with differentially expressed alleles were further investigated for functional clustering utilizing DAVID [Database for Annotation, Visualization, and Integrated Discovery (Huang et al., 2008)], and those with KEGG pathways and GO patterns of Benjamini $p<0.05$ values were of particular interest during this time course study.

The RNA editing expression was analysed using RNA-Seq data from the 20 time points (minimum of 7 time points / infection course), with the binomial test (log transformed modification) performed on reads with a minimum of 40 coverage (RNA-Seq), selecting $p<0.001$ as a cutoff for candidates with RNA edited expression. The DNAnexus, Inc. genomic browser was used to view the location of the variants relative to the gene [from NCBI RefSeq database (Pruitt et al., 2007)]). Chromas, Technelysium Pty Ltd., version 2.33 was used to view Sanger sequencing of cDNA generated from RNA at corresponding time points, were used for validation of heteroallelic expression and RNA edits (Figures 5-6 and S7). Candidates for differential ASE and editing were also validated via digital droplet PCR quantification utilizing the QuantaLife^{TM} Droplet Reader (Bio-Rad Laboratories, Inc.). Here, cDNA at the respective time point was prepared, followed by emulsion droplet preparation consisting of FAM and VIC variant-specific probes, gene-specific primers and an emulsion PCR pool (Hindson et al., 2011).

Variants identified in Proteins

For variants (SNVs and edits) identified in the genome we created a workflow to identify such changes at the protein level (Data S11B). After analyzing the data with Polyphen 2 (Adzhubei et al., 2010) the identified missense information and protein locations were obtained. A protein database was created using the available information based on the IPI (Kersey et al., 2004) human database v. 3.83. For each variant a corresponding protein sequence was constructed and added to a modified database. Additionally, a database of masses for modified peptides was calculated, post *in silico* digestion, to obtain a mass list for targeted identification experiments. The obtained spectra for both the targeted and untargeted experiments were searched independently against both the modified and unmodified (original sequence) databases containing the proteins of interest. Variant peptide candidates were identified in Proteome Discoverer ® (Thermo Scientific) using the built-in SEQUEST (Eng et al., 1994) algorithm by searching against the constructed databases augmented with a reversed database search (Elias et al., 2004; Gygi et al., 1999; Peng et al., 2003) [with a False Discovery Rate (FDR) < 0.01 and requiring 1 unique peptide per protein for identification]. The identified peptides were then filtered and were selected if they matched only to the modified database and additionally successfully aligned to the original database using local Smith-Waterman algorithm (Smith and Waterman, 1981) to verify exact matching with a single mismatch of the input modified amino acid. Peptide variants corresponding to SNVs without an entry in dbSNP were classified as private. Furthermore heterozygous peptide candidates were identified for the proteins that matched the modified database if they also aligned to the original database and showed single mismatch to the modified database. All candidate peptides were then separated into high and low confidence variant candidate lists if they exactly matched unique proteins, with a single mismatch of the input modified amino acid, after searching the IPI database sequences via BLAST (Altschul et al., 1997). Results are summarized in Data S9.

Von Willebrand factor (VWF) cleaving protease activity assay

Protease activity was assayed by the scanning densitometry of dimers of the 176 kd fragment of VWF generated by the addition of VWF substrate treated with guanidine hydrochloride to subject plasma. The percent protease activity was calculated as the percentage of VWF cleaving activity compared that measured in plasma from pooled normal controls (defined as 1 U ml^{-1}) (Tsai and Lian, 1998).

General -omics Analysis Framework and Result Summaries

A. Analysis framework

During the investigation multiple –omics data were collected at each time point. Each data set was analyzed as outlined in Figure S5. Namely, datasets were: (1) preprocessed using methods appropriate to the –omics type, towards a similar goal of ultimately integrating the different omics platforms, (2) spectrally analyzed and classified into significant categories (3) assessed for biological significance through enrichment analysis. Most statistical analysis was performed using *Mathematica 8.0 (Wolfram*

Research, 2010) (except as indicated below). In this section we provide further details pertaining to the analysis framework of each data set, as well as a summary of the methodology at each step.

(1) Data Preprocessing

After completion of the various experiments the initial raw data was first preprocessed for each –omics as outlined below to obtain a vector normalized set of time points for each constituent:

(a) Transcriptome: Illumina reads (.fastq files) were mapped to hg19 (Genome Reference Consortium GRCh37 using the Tophat (Trapnell et al., 2009), followed by Cufflinks (Trapnell et al., 2010) for transcript assembly and expression levels using RefSeq) (Pruitt et al., 2009; Pruitt et al., 2007) annotation. Data across the different time points was matched to accession, and Quality Control (QC) filtering was performed, requiring that at minimum one data point per accession displayed expression levels > 5 FPKM (Trapnell et al., 2010). The filtered datasets were then quantile normalized across all data points. Log-2 ratios of expression with respect to (w.r.t.) healthy timepoints (day 255) were then vector normalized (Euclidean metric) to one for each accession-number set. Concurrently, a bootstrap distribution of n>100,000 timed sample sets was obtained (non-parametric sampling with replacement for each time point) for statistical comparison (see part (2) below).

(b) Proteome: Spectra were obtained from three TMT (Tandem Mass Tag) labeled samples (with three technical replicates each) for relative quantitation analysis.

As described above, protein identification was carried out using Proteome Discoverer ®, with FDR <0.01 and requiring two unique peptides per protein for identification. For relative quantitation each time point was compared to a healthy time point, Day 255, and all ratios were normalized by Proteome Discoverer so that the average ratio per sample is one. Post protein identification, the three sets were matched using a replicated common ratio present in all three (namely, in this investigation, for PBMC proteins using the 131/126 ratio for the intensities of tags with masses 126 amu to 131 amu corresponding to Days 255 and 301 respectively, showing high reproducibility, with correlations $R^2 > 0.72$, Data S11C.II). Additionally, QC assessment required a coefficient of variation (CV) < 0.13 for the replicated ratio (corresponding to excluding outliers > 3 standard deviations from the median CV); that the reference (day 255) mass tag be always present in all three samples; and a minimum of 2/3 points be present for all proteins identified. The log-2 relative ratios were again vector normalized to one (Euclidean metric), and again a non-parametric bootstrap distribution (n>100,000 samples) was constructed by sampling each time point with replacement.

(c) Metabolome: Spectra from profiling at each time point were obtained with 3 technical replicates each and aligned for mass and retention times using MassHunter ® (Agilent Technologies) as described above. The aligned spectra information was filtered for a minima of 2/3 time points being present for each mass identified in the mass spectrometry sets, for which the median of the replicates was calculated, retaining data displaying a CV < 0.4. The log-2 distribution of each time-point set was standardized (baselining) to the median and average median deviation of its own distribution. Additionally a non-parametric bootstrap distribution, of 100,000 samples was constructed by sampling each time-point set with replacement. For both simulated and original data, the difference, $\sigma_{\Delta} = \sigma_{t}$-$\sigma_{healthy}$, was computed, comparing the median deviation, $\sigma_{t,}$ of each mass at time-point (t) from its distribution median to the median deviation of each mass, $\sigma_{healthy}$, at healthy time-point (Day 255) from its own distribution median. Finally, the set of differences was vector normalized (Euclidean metric) for each mass.

(2) Common Framework Data Classification

After all data had been vector normalized, it was analyzed to determine trends that dynamically emerge for each transcript, protein or metabolite. As the data sampling was uneven in time, a spectral analysis approach is adopted. For each time-series curve a periodogram was constructed through oversampling the frequency space by using a Lomb-Scargle(Lomb, 1976; Scargle, 1982, 1989) [Fourier] transformation - which has been successfully applied in astronomy (Gregory, 2005; Van Dongen et al., 1999; Van Dongen et al., 2001; Yang et al., 2011) for unevenly sampled time series data and implemented in various forms for biological problems (Abramoff et al., 2004; Ahdesmaki et al., 2007; Parkhomchuk et al., 2009; Schimmel, 2001; Van Dongen et al., 2001; Yang et al., 2011; Zhao et al., 2008). Briefly the Lomb-Scargle method is equivalent to performing a linear least-squares fit of harmonic functions for a given time-series. Namely for a time series $X(t_j), j \in \{1,2,...\mathrm{N}\}$, sampled at an arbitrary N points, the periodogram can be written as (Van Dongen et al., 1999)

$$P_{\mathrm{X}}(\omega) = \frac{1}{2}\left\{ \frac{\left[\sum_j X(t_j)\cos[\omega(t_j-\tau)]\right]^2}{\sum_j \cos^2[\omega(t_j-\tau)]} + \frac{\left[\sum_j X(t_j)\sin[\omega(t_j-\tau)]\right]^2}{\sum_j \sin^2[\omega(t_j-\tau)]} \right\}, \qquad (1.1)$$

where τ is given by

$$\tan[2\omega\tau] = \frac{\sum_j \sin[2\omega t_j]}{\sum_j \cos[2\omega t_j]}. \tag{1.2}$$

After obtaining the periodogram, the original time-series was reconstructed using an inverse Fourier transform and evenly resampling frequencies/times (as discussed by Scargle and Hocke et al(Hocke, 1998; Hocke and Kämpfer, 2009; Scargle, 1982, 1989)). This allowed us to reconstruct the series so that standard time-series analysis methods could be applied, and to fill in gaps in a robust fashion given that the spectral approach considers the entire time-series data as a whole, in contrast to other local linear or spline interpolation methods. The data was then classified into three groups: (I) After reconstructing, each time-series curve, $Y(t_j)$, we considered autocorrelation, $\rho_k = \sum_{j=1}^{N-k}(Y(t_j)-\mu_Y)(Y(t_{(j+k)}-\mu_Y)/\sum_{j=1}^{N}(Y(t_j)-\mu_Y)$, at lag k=1, as a check of non-randomness. A class of *autocorrelated* signals was selected (p<0.05 cutoff, one-tailed, based on obtaining a distribution of the autocorrelations from the bootstrap distributions constructed for each dataset. As an example, for transcriptome data for the duration of the time course this corresponds to $\rho_1 > 0.25$, Data S11C.III, in good agreement with theoretical values (Anderson, 1942) for the length of data, N=20). After removal of the autocorrelated signals from the set, the remaining signals were checked for aberrant spikes, significantly high or low signal instances compared to what would be expected in a random distribution. Signals that displayed aberrant high signals (p<0.05, one-tailed by comparison to analysis of randomly simulated distribution of normalized time signals of corresponding length N for each time-series) were classified as (II) *spike maxima*, while signals that displayed aberrant low signals (p<0.05, one-tailed) were classified as (III) *spike minima*. Thus three classes of significant trends were selected for each of the input –omics datasets.

(3) Clustering

The classified datasets from (2) above were clustered using the hierarchical agglomerative algorithm in *Mathematica 8.0*, with correlation distance and average linkage. Once the clustering was determined, the number of clusters per agglomerated dataset was ascertained by inspection of the fusion coefficients of their respective dendrograms. To assess the biological significance for each of the obtained clusters,, gene-based pathway and ontology enrichment and network analysis was performed using Cytoscape (Cline et al., 2007; Shannon et al., 2003; Smoot et al., 2011). Namely, the Reactome (Croft et al., 2011; Joshi-Tope et al., 2005; Matthews et al., 2007; Matthews et al., 2009; Vastrik et al., 2007) Functional Interaction (FI) plugin was used to assess membership of genes to Reactome and KEGG (Kanehisa and Goto, 2000) pathways and to calculate enrichment ($p<0.05$, FDR <0.05). Furthermore, Gene Ontology (Ashburner et al., 2000) (GO) analysis was performed using the BiNGO (Maere et al., 2005) plugin for Cytoscape, for significantly enriched membership ($p<0.05$ and Benjamini-Hochberg (Benjamini and Hochberg, 1995) adjusted $p<0.05$) in each of Cellular Component (CC), Molecular Function (MF) and Biological Process (BP) categories.

B. Results summaries and file guide

In this section we provide results summaries for the dynamical analysis following the analysis framework outlined above. In particular the results from the main text and relevant tables are also included in the supplemental tables following the naming conventions in associated figure as outlined below. Based on the criteria indicated in part

(A) above, all data below is grouped into classes and assessed for biological significance through enrichment analysis:

1. Transcriptome: Entire Time Course.

Expression levels for 19,714 distinct isoforms from RNA-seq data analysis were consistently tracked from day 0 to day 400 of the study, covering the onset of both HRV and RSV infections (see IFigure1C for isoform distributions). Of these isoforms, 4,922 were grouped in the autocorrelation class, while 3,718 were categorized as spike maxima and 7,891 as spike minima. Clustering and significant results from the enrichment analysis are shown Figure S6A and associated Data S6.

2. Proteome PBMC: RSV Infection

Relative expression levels for 3,731 PBMC proteins were consistently tracked from day 186 to day 400 of the study, covering the onset of RSV infection after day 289. Of the tracked proteins, 257 were grouped in the autocorrelation class, while others displayed significant aberration from the median response, namely 1,240 showing spike maxima and 1,194 showing spike minima. Clustering and significant results from the enrichment analysis are shown Figure S3A and associated Data S3.

3. Proteome Serum: HRV Infection

Relative expression levels for 664 serum proteins were consistently tracked from day 0 to day 116 of the study, covering the onset of HRV infection (for this part of the analysis day 116 was used for TMT ratios, corresponding to ratios w.r.t. the 130 amu tag in the spectra). Ninety-four were grouped in the autocorrelation class, 57 categorized as spike maxima and 40 as spike minima. Clustering and significant results from the enrichment analysis are shown Figure S3B and associated Data S5.

4. Metabolome: HRV and RSV infections

For the HRV infection (Days 0-185), 6,862 distinct serum metabolite m/z intensities were tracked. Of these, 385 were grouped in the autocorrelation class, 506 categorized as spike maxima and 748 as spike minima. For the RSV infection, 4,228 distinct serum metabolite m/z intensities were tracked (Days 255-400); 475 were grouped in the autocorrelation class, 577 categorized as spike maxima and 884 as spike minima. Given the modest number of identified metabolites based solely on mass (~20%) enrichment analysis did not yield significant pathways and further pathway associations will be discussed elsewhere. Clustering and overlap results are found in Figure S4 and associated Data S4.

5. Integrated Proteome Transcriptome and Metabolome for RSV infection

The different omics data set classes were clustered together for the transcriptome, PBMC proteome and serum metabolome for Days 186 to 400 of the study, covering the onset of RSV infection and high glucose levels in the latter stages of the investigation. Additional clustering and overlap results are found in Figure S6B and associated Data S7.

Data Dissemination

All omics data are being deposited in public databases. Transcriptome data (FASTQ files) and Protein Array data (GPR files) are being submitted to the GEO database. Whole genome sequences [Complete Genomics, Illumina (whole genome and Exome) for both the subject and mother] are being submitted to SRA. Proteome and Metabolome

Mass Spectra data are being submitted to TRANCHE (http://www.proteomecommons.org). Accession numbers will be available no later than acceptance of the manuscript.

SUPPLEMENTAL TEXT

VWF Cleavage Assay by ADAMTS13

We also identified a rare missense SNV in exon 24 of the gene encoding a Von Willebrand factor cleaving protease, (ADAMTS13 p.A1033T, A disintegrin and metalloproteinase with thrombospondin motifs 13 isoform 1 preproprotein). This variant occurs with a minor allele frequency of 0.033 in the CEU population and multiple sequence alignment of 46 vertebrate species identified the corresponding codon as highly evolutionarily conserved. Losses of function mutations in this gene have been associated with hereditary thrombotic thrombocytopenic purpura (TTP) (Kokame et al., 2002; Levy et al., 2001; Upshaw, 1978). This mutation has been previously reported as a polymorphism, though its effects on VWF-cleaving protease activity have not been investigated (Levy et al., 2001). Subsequent assay of von-Willebrand factor cleaving protease activity yielded a value of 63% of that of general population controls, consistent with values reported for heterozygous carriers of ADAMTS13 mutations associated with impaired VWF cleaving protease activity (Kokame et al., 2002; Levy et al., 2001). Thus, our results suggest this rare polymorphism may be a susceptibility factor for reduced VWF-cleaving protease activity. These findings may have implications for the risk of development of acquired TTP in response to autoimmunity, medications, and infection in heterozygous individuals.

REFERENCES

Abramoff, M.D., Magalhaes, P.J., and Ram, S.J. (2004). Image Processing with ImageJ. Biophotonics International *11*, 36-42.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods *7*, 248-249.

Ahdesmaki, M., Lahdesmaki, H., Gracey, A., Shmulevich, l., and Yli-Harja, O. (2007). Robust regression for periodicity detection in non-uniformly sampled time-course gene expression data. BMC Bioinformatics *8*, 233.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research *25*, 3389-3402.

Anderson, R.L. (1942). Distribution of the serial correlation coefficient. The Annals of Mathematical Statistics *13*, 1-13.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological), 289-300.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185-193.

Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet *81*, 1084-1097.

Canela, A., Vera, E., Klatt, P., and Blasco, M.A. (2007). High-throughput telomere length quantification by FISH and its application to human population studies. Proc Natl Acad Sci U S A *104*, 5300-5305.

Clark, M.J., Chen, R., Lam, H.M., Karczewski, K.J., Euskirchen, G., and Snyder, M. (2011). Exome DNA Sequencing: A Comparison of Enrichment Technologies. Nat Biotechnol *Accepted.*

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B.*, et al.* (2007). Integration of biological networks and gene expression data using Cytoscape. Nat Protoc *2*, 2366-2382.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B.*, et al.* (2011). Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Research *39*, D691-697.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences of the United States of America *95*, 14863-14868.

Elias, J.E., Gibbons, F.D., King, O.D., Roth, F.P., and Gygi, S.P. (2004). Intensity-based protein identification by machine learning from a library of tandem mass spectra. Nature Biotechnology *22*, 214-219.

Eng, J.K., McCormack, A.L., and Yates III, J.R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry *5*, 976-989.

Gregory, P.C. (2005). Bayesian logical data analysis for the physical sciences: a comparative approach with Mathematica support (Cambridge Univ Pr).

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. Mol Cell Biol *19*, 1720-1730.

Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C., et al. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem 83, 8604-8610.

Hocke, K. (1998). Phase estimation with the Lomb-Scargle periodogram method (European Geophysical Society).

Hocke, K., and Kämpfer, N. (2009). Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram. Atmos Chem Phys *9*, 4197-4206.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protocols *4*, 44-57.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Research *37*, 1-13.

Hudson, M.E., Pozdnyakova, I., Haines, K., Mor, G., and Snyder, M. (2007). Identification of differentially expressed proteins in ovarian cancer using high-density protein microarrays. Proceedings of the National Academy of Sciences of the United States of America *104*, 17494-17499.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. (2004). Human MicroRNA targets. PLoS Biol *2*, e363.

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L.*, et al.* (2005). Reactome: a knowledgebase of biological pathways. Nucleic Acids Research *33*, D428-432.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

Kersey, P.J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004). The International Protein Index: an integrated database for proteomics experiments. Proteomics *4*, 1985-1988.

Kokame, K., Matsumoto, M., Soejima, K., Yagi, H., Ishizashi, H., Funato, M., Tamai, H., Konno, M., Kamide, K., Kawano, Y.*, et al.* (2002). Mutations and common polymorphisms in ADAMTS13 gene responsible for von Willebrand factor-cleaving protease activity. Proceedings of the National Academy of Sciences of the United States of America *99*, 11902-11907.

Levy, G.G., Nichols, W.C., Lian, E.C., Foroud, T., McClintick, J.N., McGee, B.M., Yang, A.Y., Siemieniak, D.R., Stark, K.R., Gruppo, R.*, et al.* (2001). Mutations in a member of the ADAMTS gene family cause thrombotic thrombocytopenic purpura. Nature *413*, 488-494.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008). SOAP: short oligonucleotide alignment program. Bioinformatics *24*, 713-714.

Lomb, N. (1976). Least-squares frequency analysis of unequally spaced data. Astrophysics and space science *39*, 447-462.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448-3449.

Matthews, L., D'Eustachio, P., Gillespie, M., Croft, D., de Bono, B., Gopinath, G., Jassal, B., Lewis, S., Schmidt, E., Vastrik, I., *et al.* (2007). An Introduction to the Reactome Knowledgebase of Human Biological Pathways and Processes. . Bioinformatics Primer, NCI/Nature Pathway Interaction Database.

Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., de Bono, B., Garapati, P., Hemish, J., Hermjakob, H., Jassal, B., *et al.* (2009). Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Research *37*, D619-622.

McIlrath, J., Bouffler, S.D., Samper, E., Cuthbert, A., Wojcik, A., Szumiel, I., Bryant, P.E., Riches, A.C., Thompson, A., Blasco, M.A., *et al.* (2001). Telomere length abnormalities in mammalian radiosensitive cells. Cancer Res *61*, 912-915.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic acids research *37*, e123.

Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., and Gygi, S.P. (2003). Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. Journal of proteome research *2*, 43-50.

Pruitt, K.D., Tatusova, T., Klimke, W., and Maglott, D.R. (2009). NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Research *37*, D32-36.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research *35*, D61-65.

Saldanha, A.J. (2004). Java Treeview--extensible visualization of microarray data. Bioinformatics *20*, 3246-3248.

Scargle, J.D. (1982). Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data. The Astrophysical Journal *263*, 835-853.

Scargle, J.D. (1989). Studies in astronomical time series analysis. III-Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. The Astrophysical Journal *343*, 874-887.

Schimmel, M. (2001). Emphasizing difficulties in the detection of rhythms with Lomb-Scargle periodograms. Biol Rhythm Res *32*, 341-345.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498-2504.

Skelly, D.A., Johansson, M., Madeoy, J., Wakefield, J., and Akey, J.M. (2011). A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. Genome research 21, 1728-1737.

Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. Journal of Molecular Biology *147*, 195-197.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics *27*, 431-432.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Tsai, H.M., and Lian, E.C. (1998). Antibodies to von Willebrand factor-cleaving protease in acute thrombotic thrombocytopenic purpura. N Engl J Med *339*, 1585-1594.

Upshaw, J.D., Jr. (1978). Congenital deficiency of a factor in normal plasma that reverses microangiopathic hemolysis and thrombocytopenia. N Engl J Med *298*, 1350-1352.

Van Dongen, H.P., Olofsen, E., VanHartevelt, J.H., and Kruyt, E.W. (1999). A procedure of multiple period searching in unequally spaced time-series with the Lomb-Scargle method. Biol Rhythm Res *30*, 149-177.

Van Dongen, H.P., Ruf, T., Olofsen, E., VanHartevelt, J.H., and Kruyt, E.W. (2001). Analysis of problematic time series with the Lomb-Scargle Method, a reply to 'emphasizing difficulties in the detection of rhythms with Lomb-Scargle periodograms'. Biol Rhythm Res *32*, 347-354.

Vastrik, I., D'Eustachio, P., Schmidt, E., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., Matthews, L.*, et al.* (2007). Reactome: a knowledge base of biologic pathways and processes. Genome Biol *8*, R39.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research *38*, e164.

Wolfram Research, I. (2010). Mathematica, Version 8.0 (Champaign Illinois, Wolfram Research, Inc.).

Yang, R., Zhang, C., and Su, Z. (2011). LSPR: an integrated periodicity detection algorithm for unevenly sampled temporal microarray data. Bioinformatics *27*, 1023-1025.

Zhao, W., Agyepong, K., Serpedin, E., and Dougherty, E.R. (2008). Detecting Periodic Genes from Irregularly Sampled Gene Expressions: A Comparison Study. EURASIP Journal on Bioinformatics and Systems Biology *2008*.

Zhu, X., Gerstein, M., and Snyder, M. (2006). ProCAT: a data analysis approach for protein microarrays. Genome Biol *7*, R110.