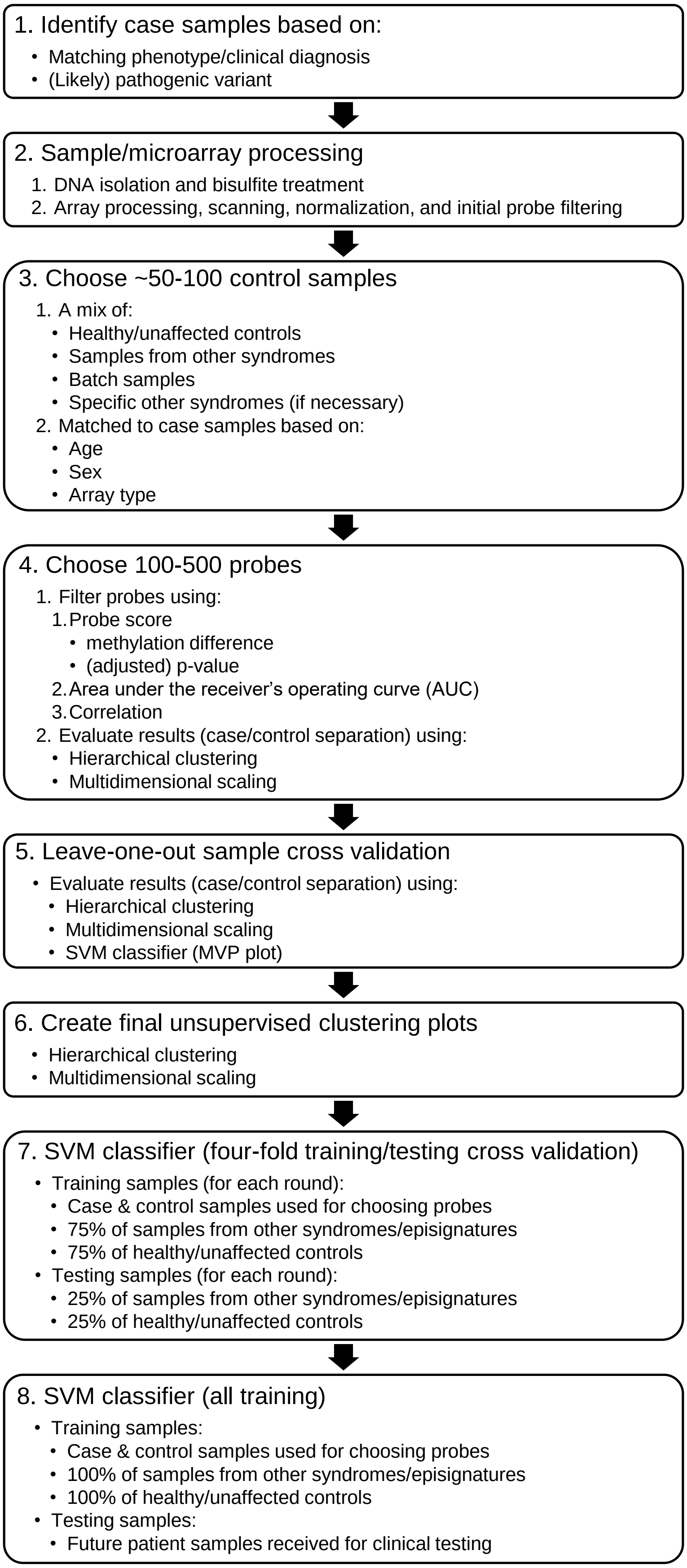
HGGA, Volume *3*

Supplemental information

Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders

Michael A. Levy, Haley McConkey, Jennifer Kerkhof, Mouna Barat-Houari, Sara Bargiacchi, Elisa Biamino, María Palomares Bralo, Gerarda Cappuccio, Andrea Ciolfi, Angus Clarke, Barbara R. DuPont, Mariet W. Elting, Laurence Faivre, Timothy Fee, Robin S. Fletcher, Florian Cherik, Aidin Foroutan, Michael J. Friez, Cristina Gervasini, Sadegheh Haghshenas, Benjamin A. Hilton, Zandra Jenkins, Simranpreet Kaur, Suzanne Lewis, Raymond J. Louie, Silvia Maitz, Donatella Milani, Angela T. Morgan, Renske Oegema, Elsebet Østergaard, Nathalie Ruiz Pallares, Maria Piccione, Simone Pizzi, Astrid S. Plomp, Cathryn Poulton, Jack Reilly, Raissa Relator, Rocio Rius, Stephen Robertson, Kathleen Rooney, Justine Rousseau, Gijs W.E. Santen, Fernando Santos-Simarro, Josephine Schijns, Gabriella Maria Squeo, Miya St John, Christel Thauvin-Robinet, Giovanna Traficante, Pleuntje J. van der Sluijs, Samantha A. Vergano, Niels Vos, Kellie K. Walden, Dimitar Azmanov, Tugce Balci, Siddharth Banka, Jozef Gecz, Peter Henneman, Jennifer A. Lee, Marcel M.A.M. Mannens, Tony Roscioli, Victoria Siu, David J. Amor, Gareth Baynam, Eric G. Bend, Kym Boycott, Nicola Brunetti-Pierri, Philippe M. Campeau, John Christodoulou, David Dyment, Natacha Esber, Jill A. Fahrner, Mark D. Fleming, David Genevieve, Kristin D. Kerrnohan, Alisdair McNeill, Leonie A. Menke, Giuseppe Merla, Paolo Prontera, Cheryl Rockman-Greenberg, Charles Schwartz, Steven A. Skinner, Roger E. Stevenson, Antonio Vitobello, Marco Tartaglia, Marielle Alders, Matthew L. Tedder, and Bekim Sadikovic



1. Identify case samples based on:
   - Matching phenotype/clinical diagnosis
   - (Likely) pathogenic variant

2. Sample/microarray processing
   1. DNA isolation and bisulfite treatment
   2. Array processing, scanning, normalization, and initial probe filtering

3. Choose ~50-100 control samples
   1. A mix of:
      - Healthy/unaffected controls
      - Samples from other syndromes
      - Batch samples
      - Specific other syndromes (if necessary)
   2. Matched to case samples based on:
      - Age
      - Sex
      - Array type

4. Choose 100-500 probes
   1. Filter probes using:
      1. Probe score
         - methylation difference
         - (adjusted) p-value
      2. Area under the receiver’s operating curve (AUC)
      3. Correlation
   2. Evaluate results (case/control separation) using:
      - Hierarchical clustering
      - Multidimensional scaling

5. Leave-one-out sample cross validation
   - Evaluate results (case/control separation) using:
     - Hierarchical clustering
     - Multidimensional scaling
     - SVM classifier (MVP plot)

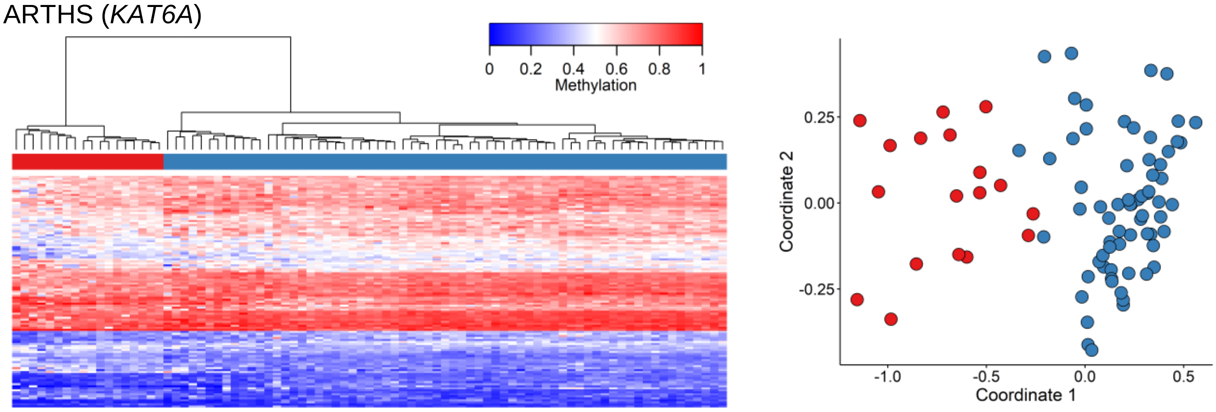
6. Create final unsupervised clustering plots
   - Hierarchical clustering
   - Multidimensional scaling

7. SVM classifier (four-fold training/testing cross validation)
   - Training samples (for each round):
     - Case & control samples used for choosing probes
     - 75% of samples from other syndromes/episignatures
     - 75% of healthy/unaffected controls
   - Testing samples (for each round):
     - 25% of samples from other syndromes/episignatures
     - 25% of healthy/unaffected controls

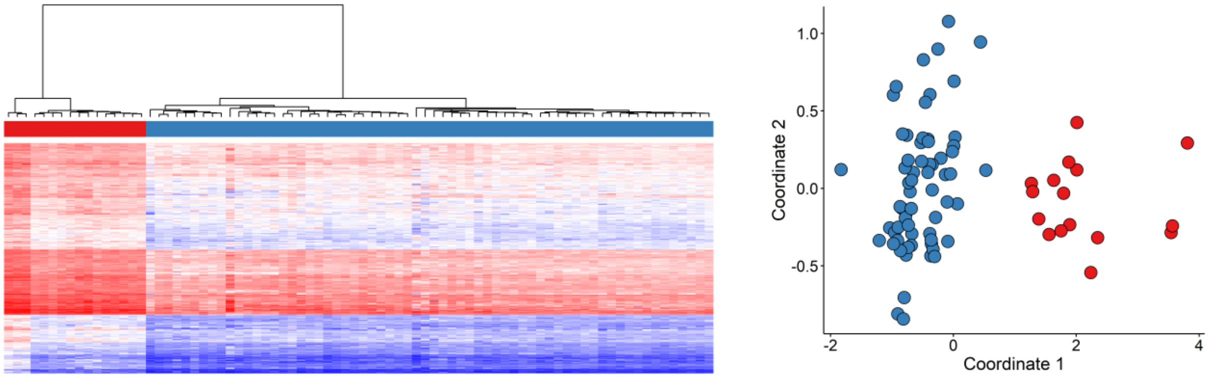
8. SVM classifier (all training)
   - Training samples:
     - Case & control samples used for choosing probes
     - 100% of samples from other syndromes/episignatures
     - 100% of healthy/unaffected controls
   - Testing samples:
     - Future patient samples received for clinical testing

Supplementary Figure 1: Episignature discovery pipeline used to identify the 19 new episignatures. A summary of the steps involved in sample processing and data analysis for identification and validation of a new episignature. Ordered steps are numbered, unordered elements are bulleted.

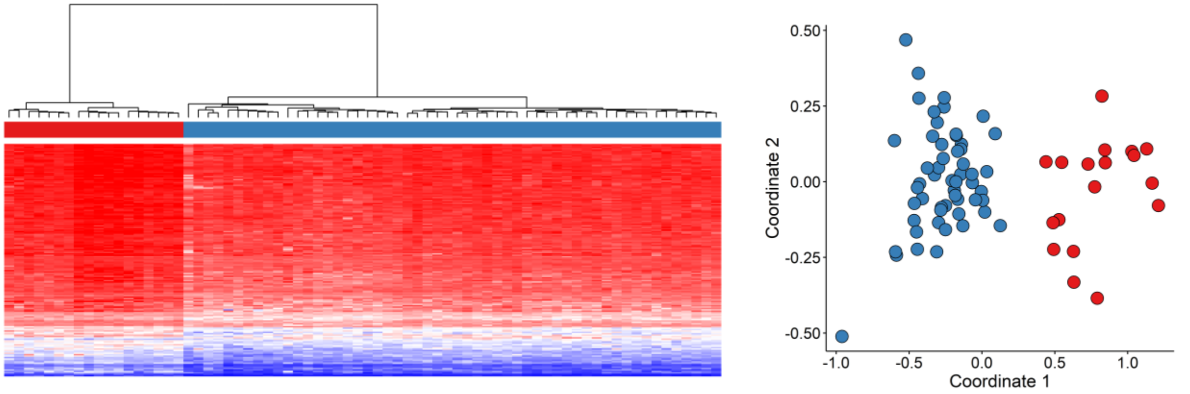
A. ARTHS (*KAT6A*)



B. BEFAHRS (*TET3*)

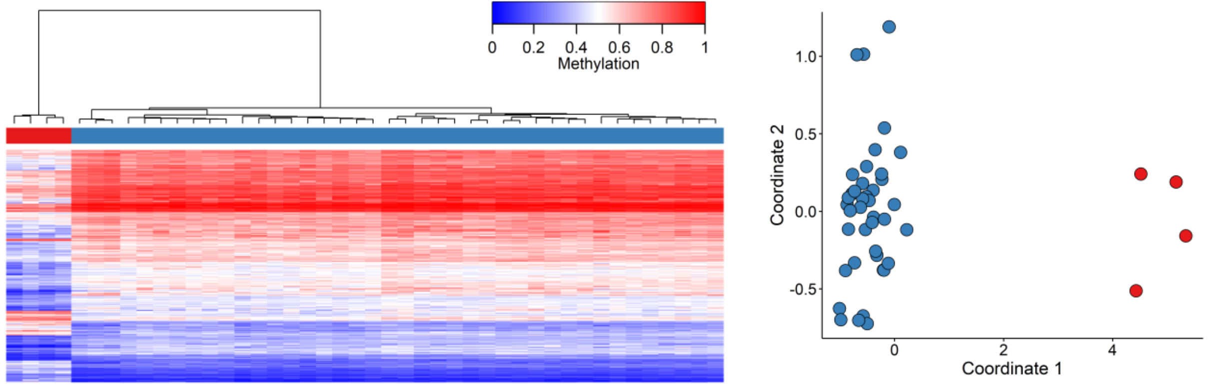


C. Chr16p11.2del

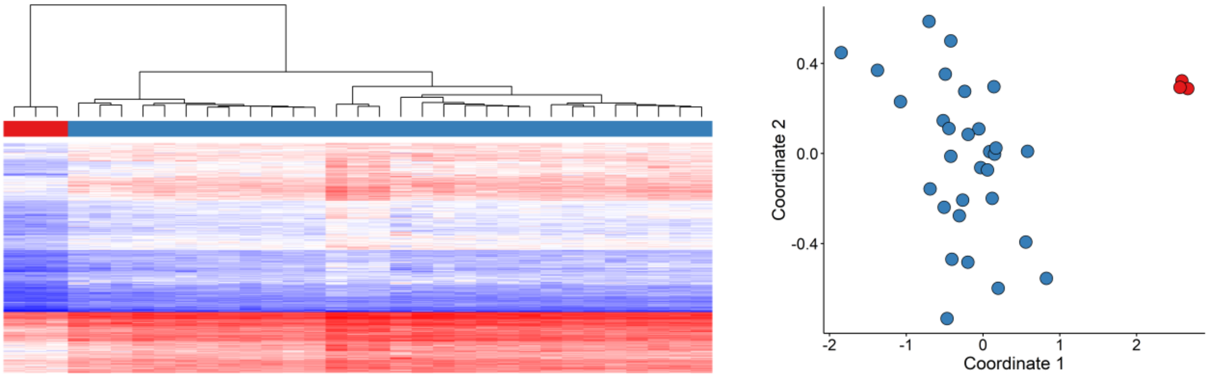


Supplementary Figure 2: Unsupervised clustering using the selected probes for the 19 new episignatures. For each episignature, hierarchical clustering (left) and MDS plots (right) are shown. Unless indicated, red are case samples and blue are controls. The methylation scale bar in A. applies to the heatmap portion each panel.

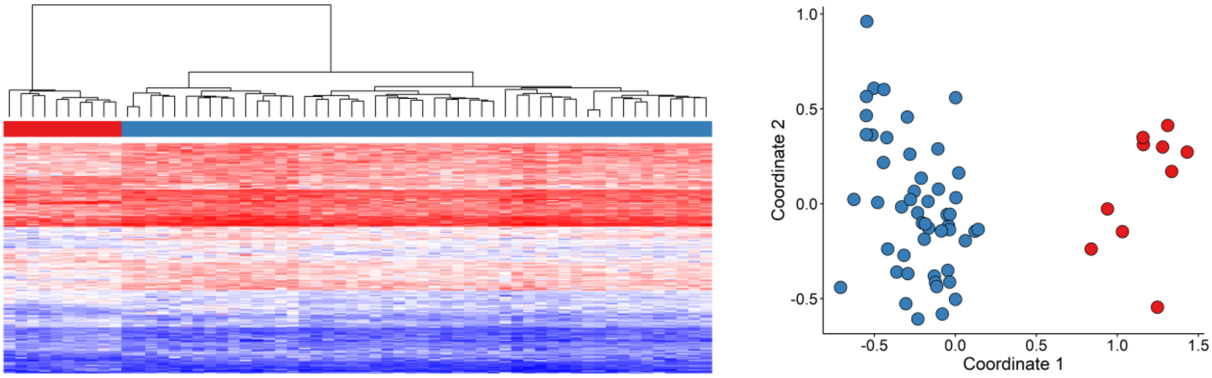
D. CSS1/2 (*ARID1A/B*) c.6200



E. CSS4 (*SMARCA4*) c.2656



F. CSS9 (*SOX11*)



G. DYT28 (*KMT2B*)

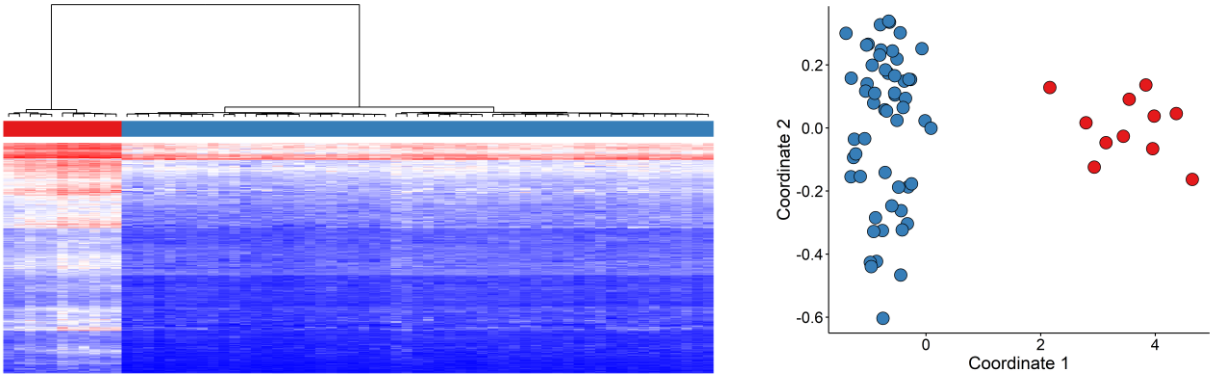
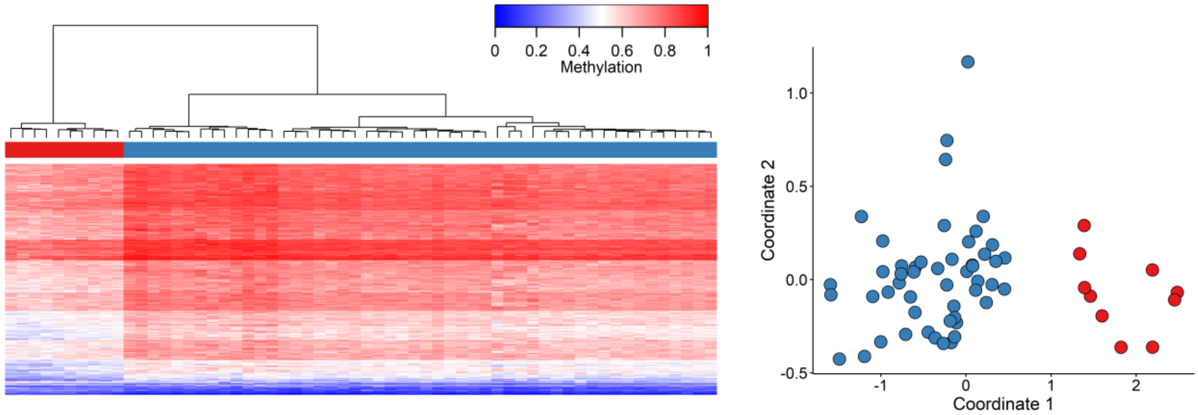
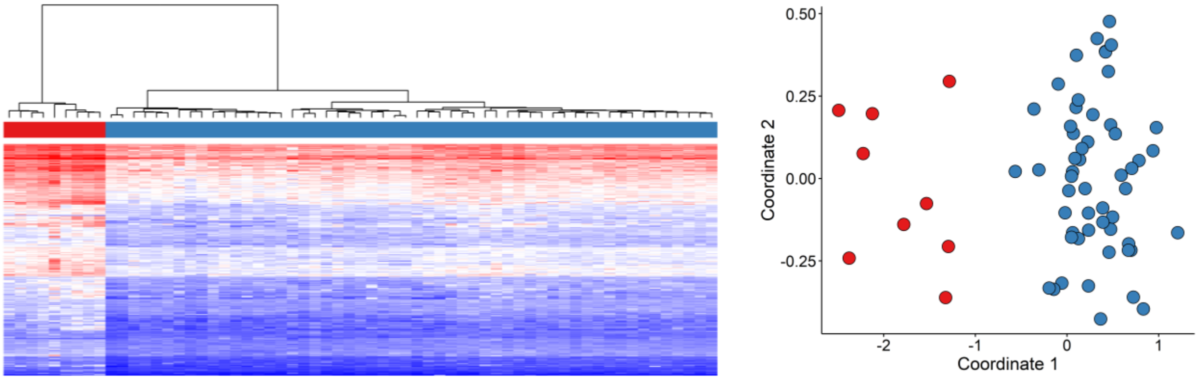


Figure S2 continued

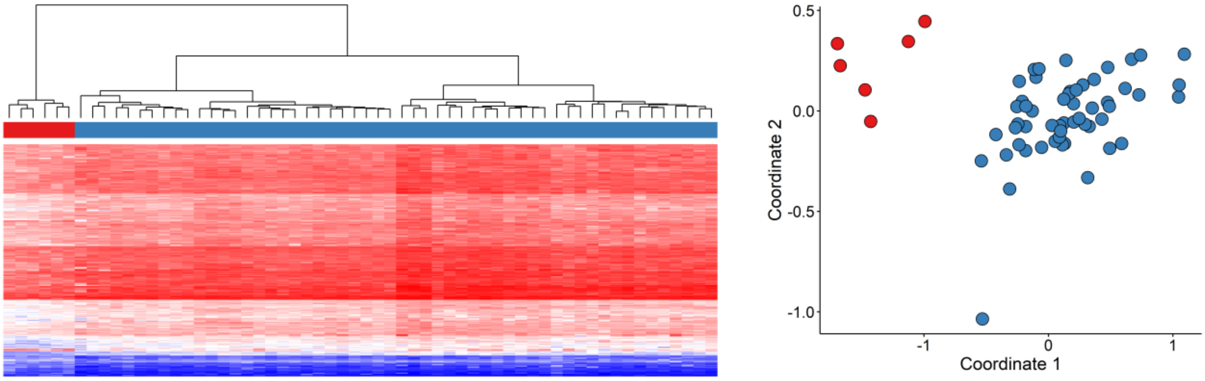
H. GADEVS (*YY1*)



I. KDM2B



J. KDM4B



K. LLS (*SETD2*)

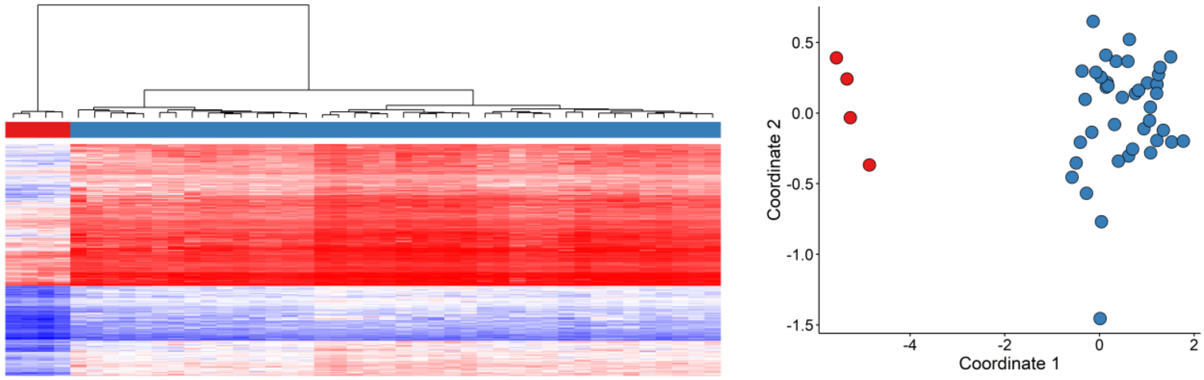
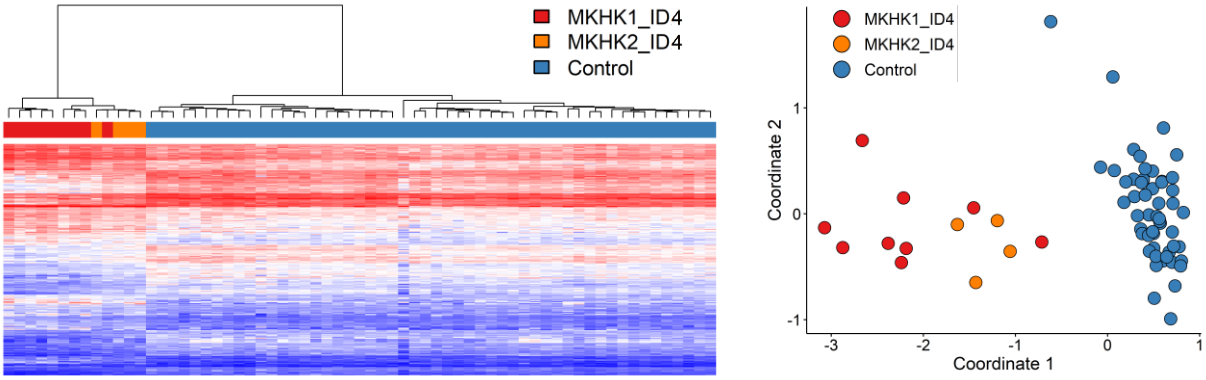
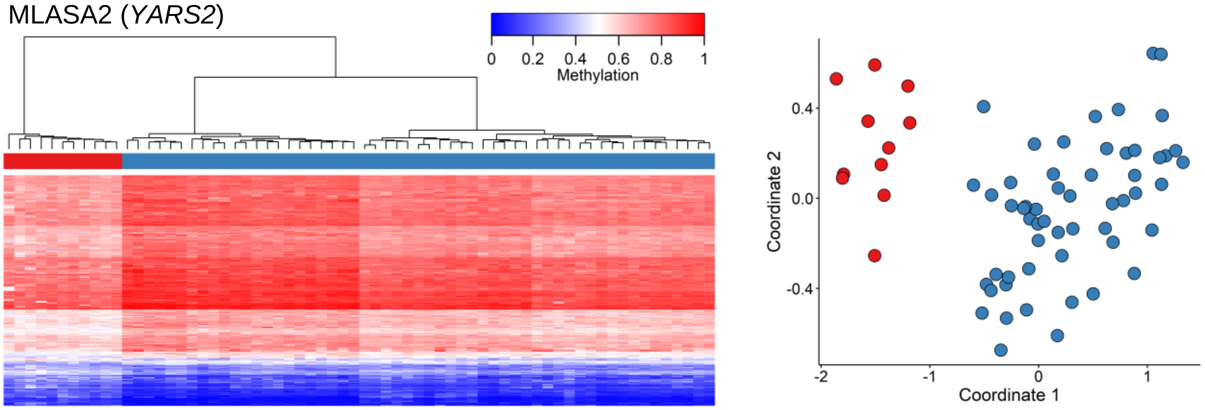


Figure S2 continued

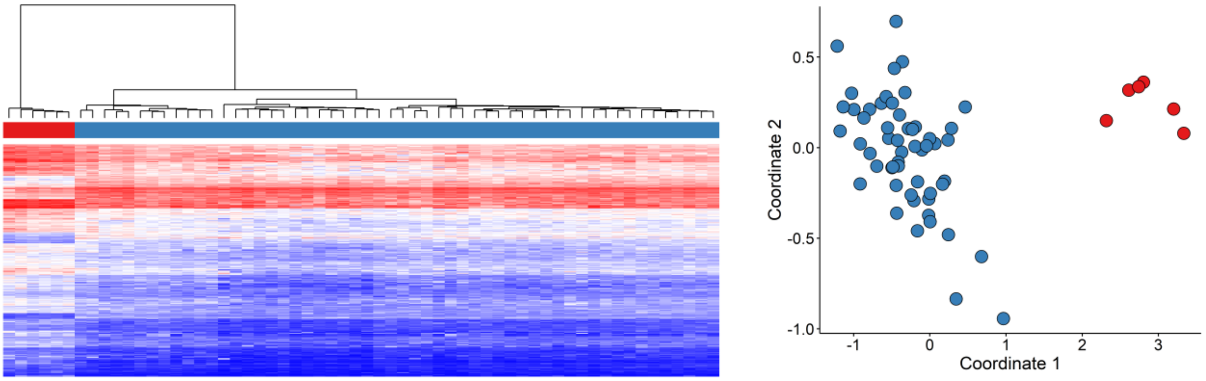
L. MKHK1/2 (*CREBB/EP300*) ID4



M. MLASA2 (*YARS2*)



N. MRXSA (*FAM50A*)



O. PHMDS (Chr22q13.3del)

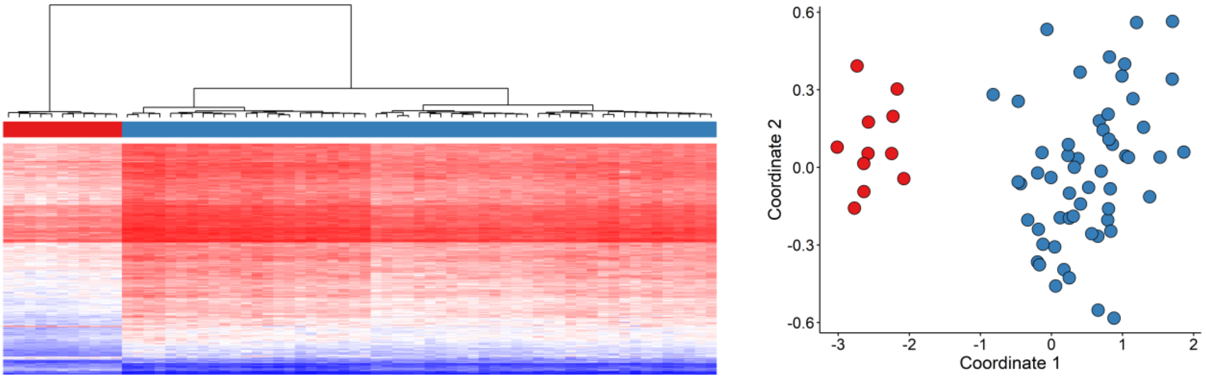
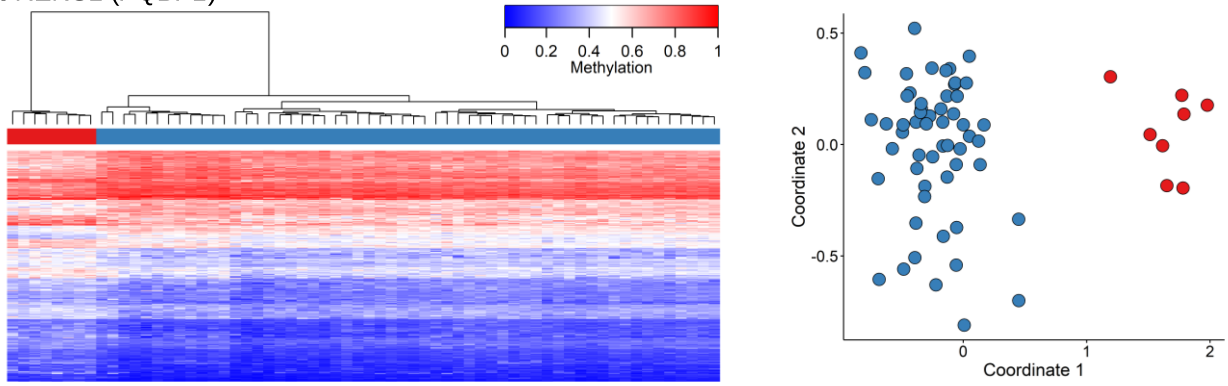
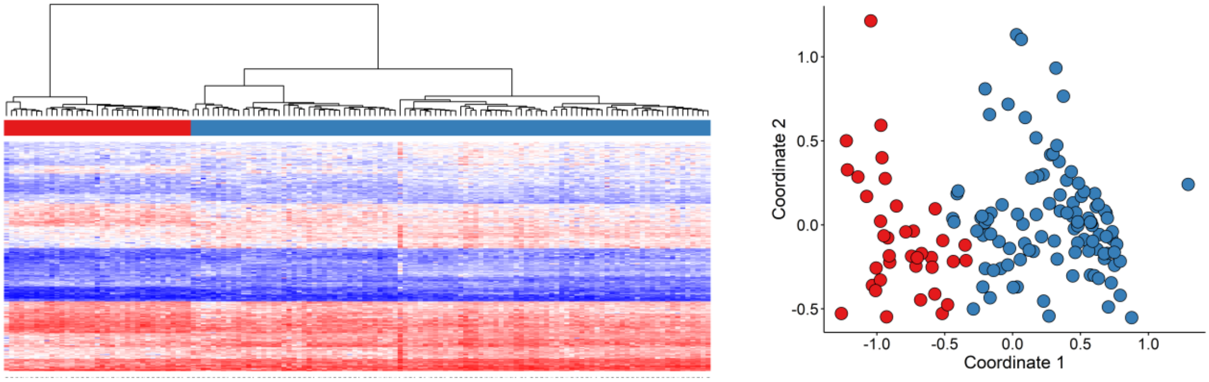


Figure S2 continued

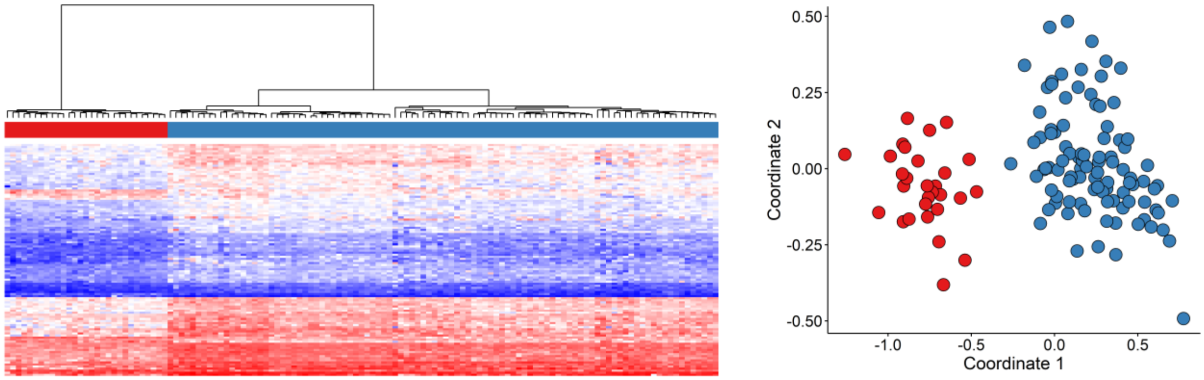
P. RENS1 (*PQBP1*)



Q. RSTS1 (*CREBBP*)



R. RSTS2 (*EP300*)



S. VCFS (Chr22q11.2 del)

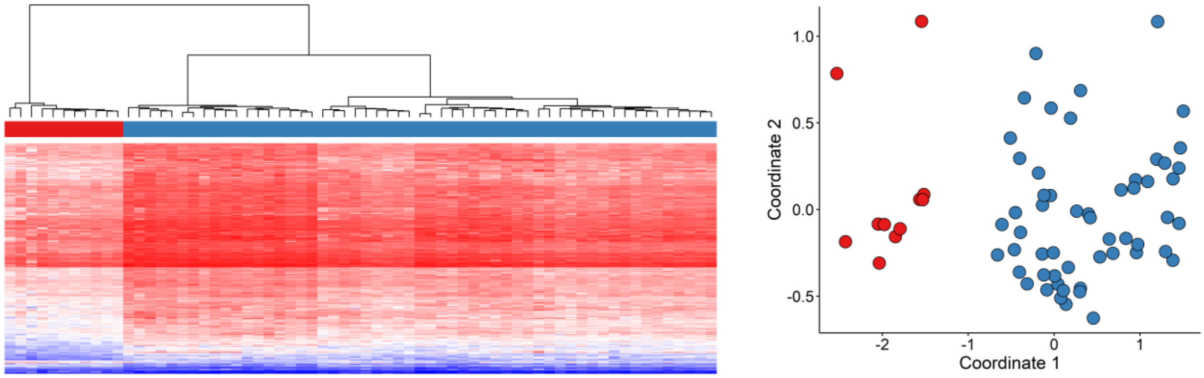
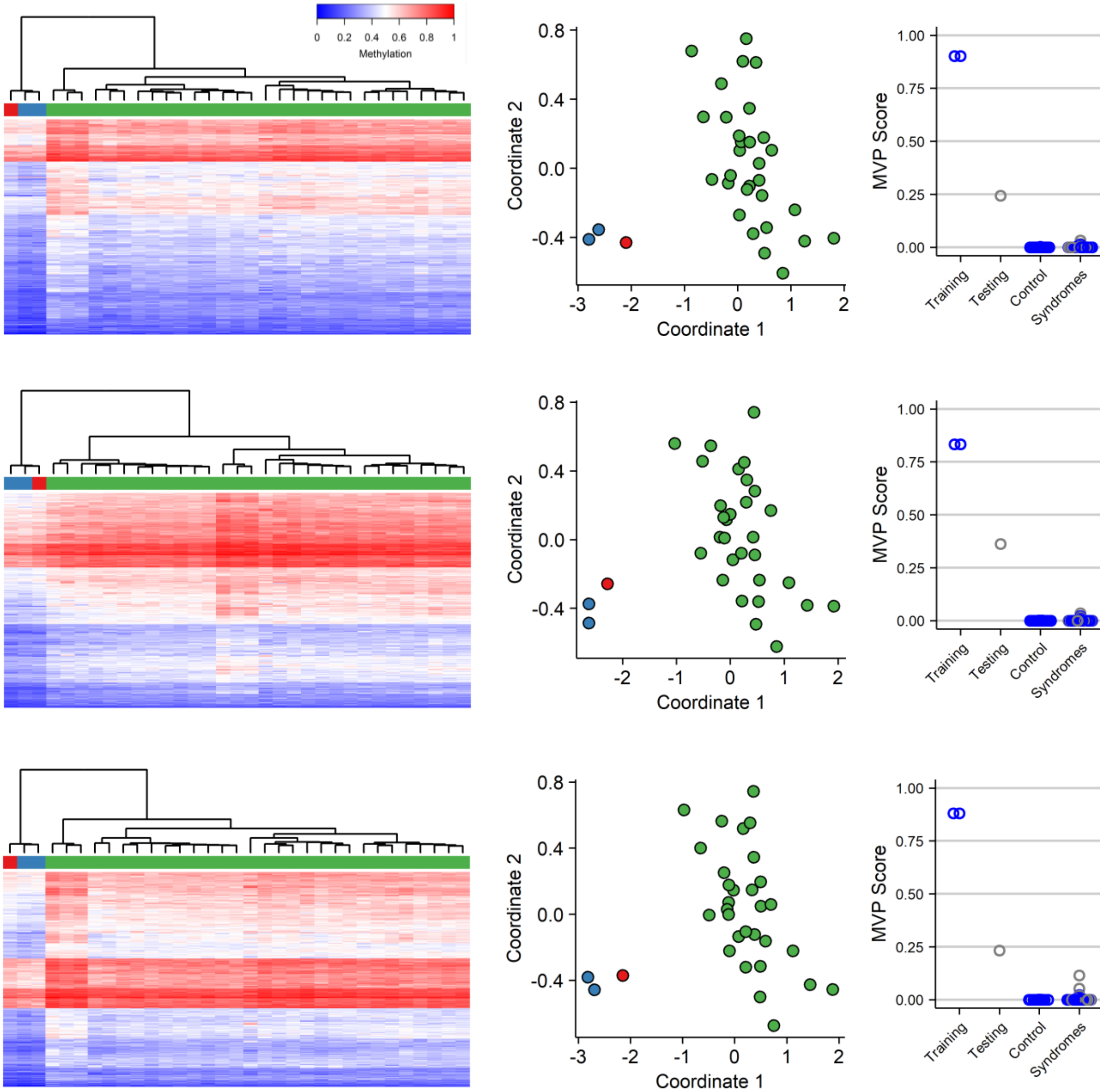
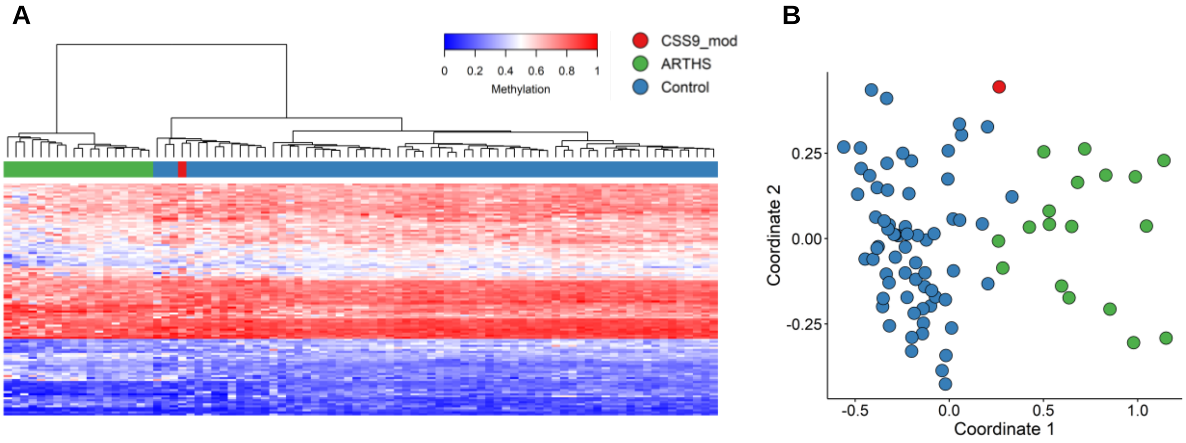


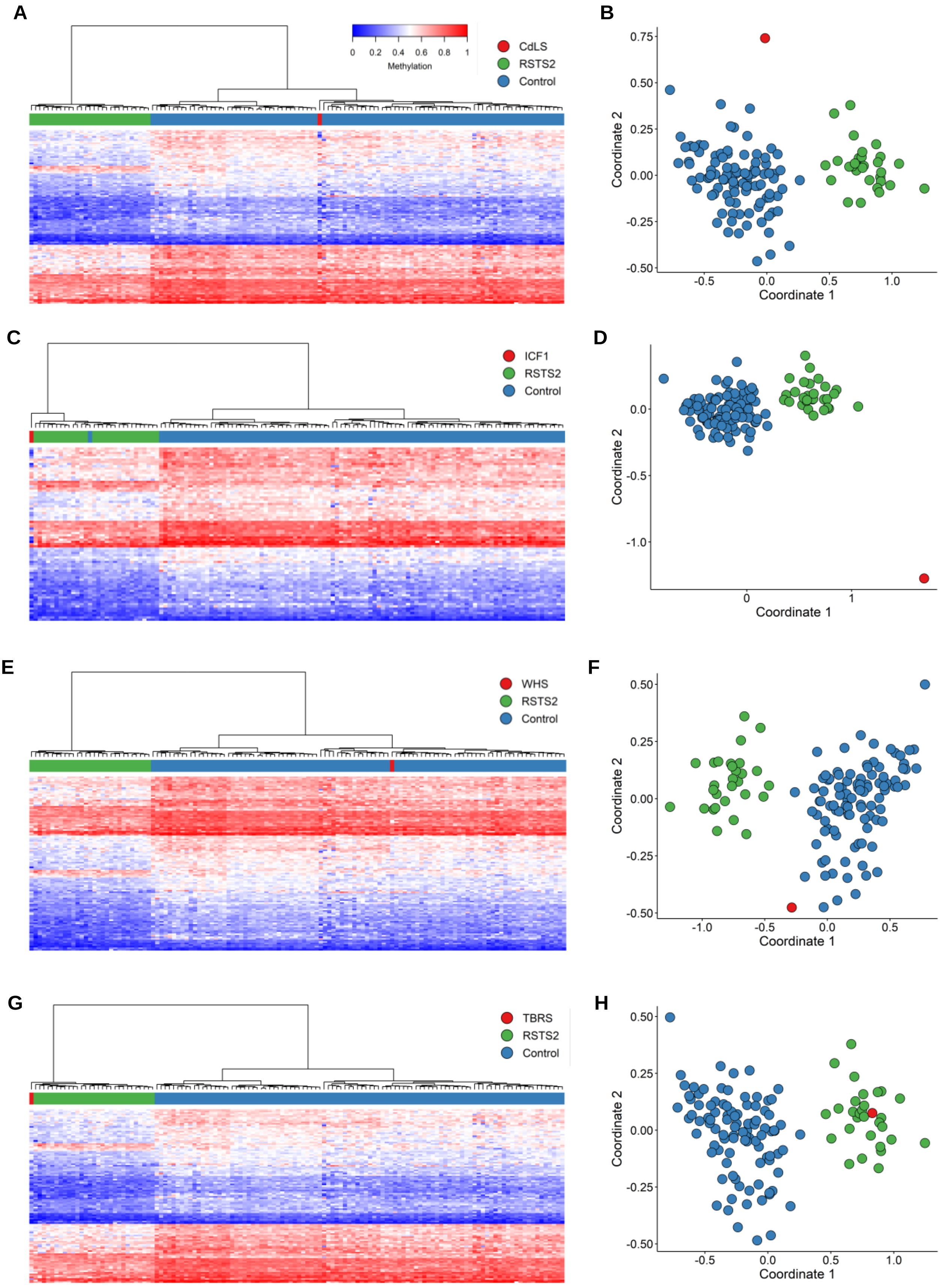
Figure S2 continued



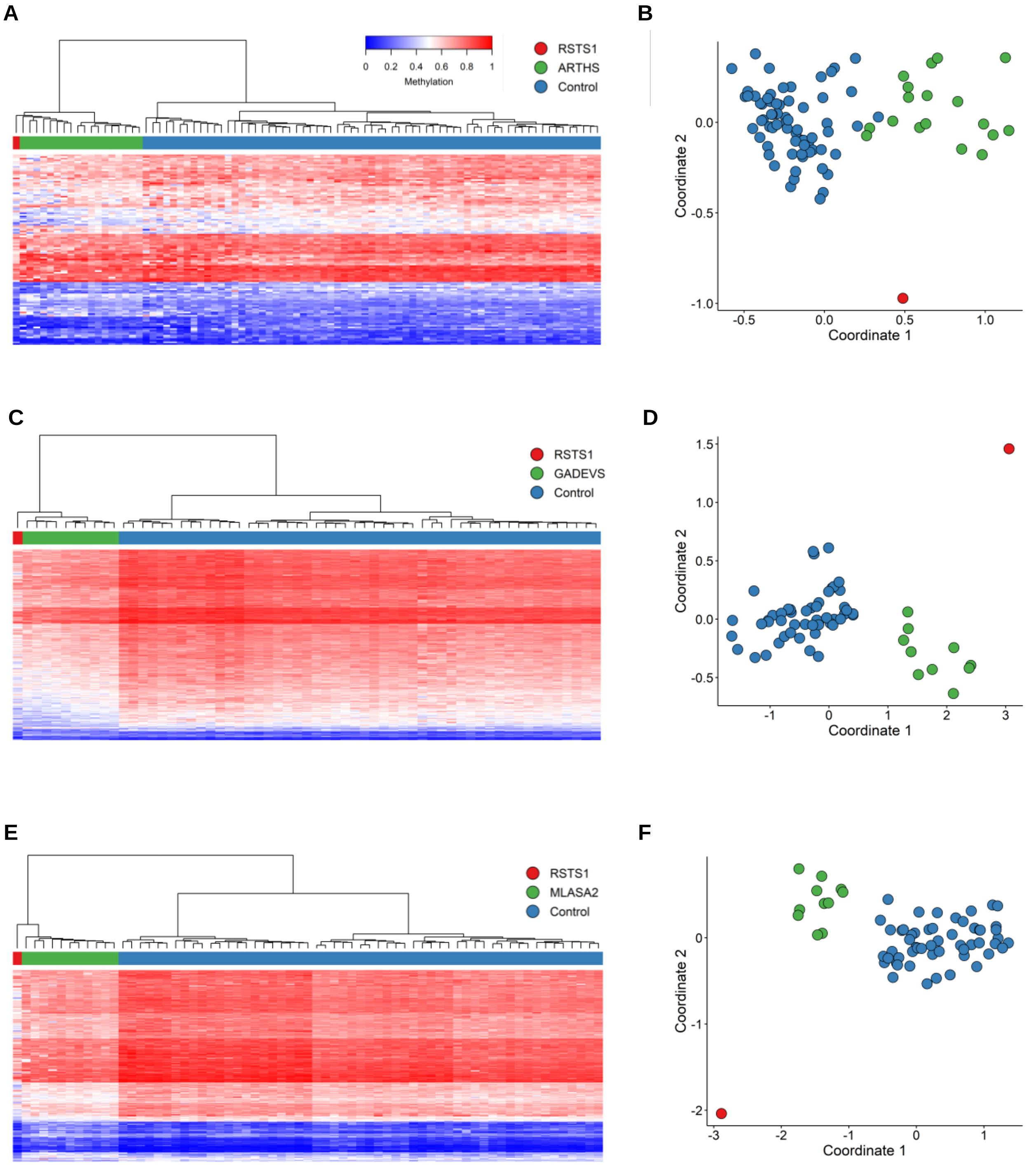
Supplementary Figure 3: Leave-one-out cross-validation of the three CSS4_c.2650 samples. For each round of cross validation two samples were used for probe selection and classifier training and the third sample was used for testing.



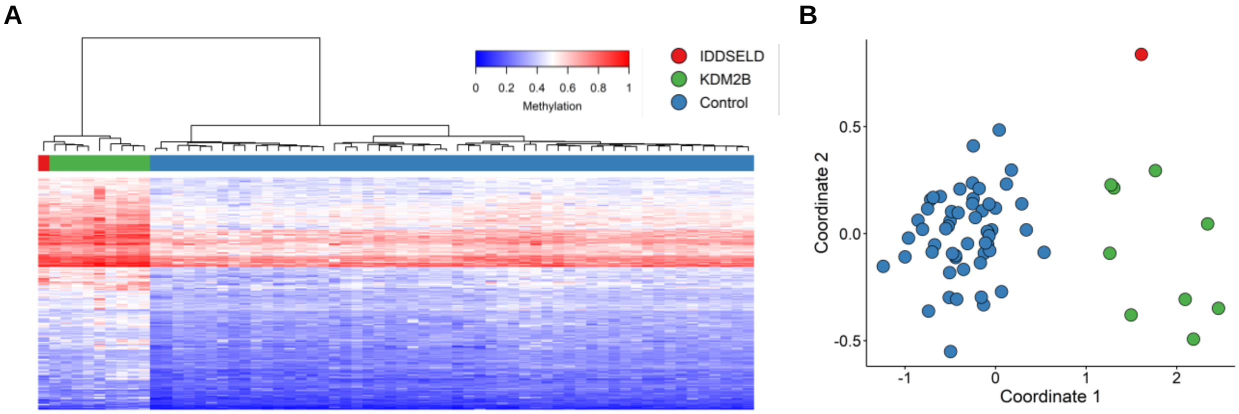
Supplementary Figure 4: Hierarchical clustering and MDS plots for the assessment of sample 1_CSS9 for ARTHS.



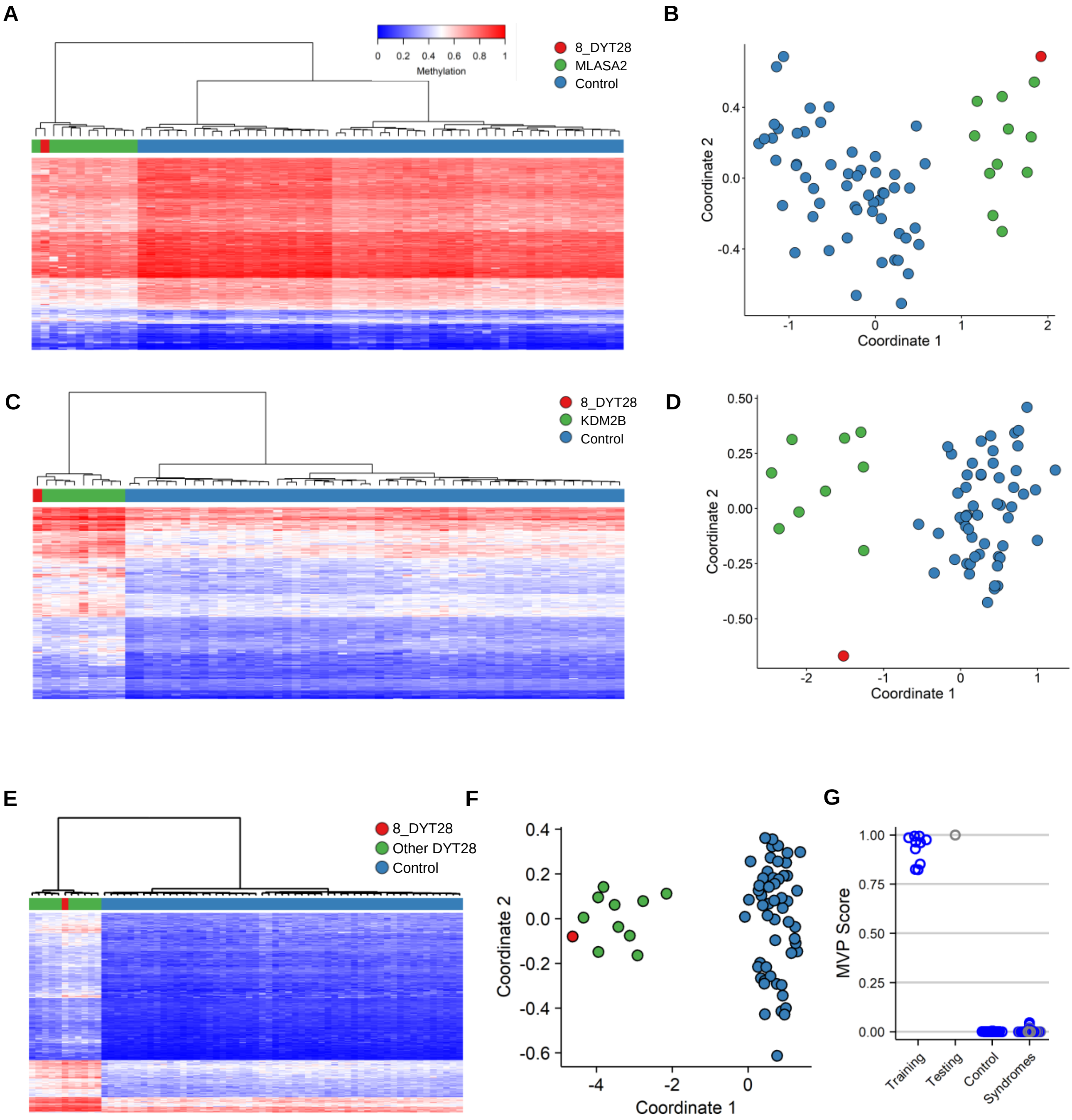
Supplementary Figure 5: Hierarchical clustering and MDS plots for the assessment of four purportedly non-RSTS2 samples with elevated RSTS MVP scores. **A,B.** Sample 2_CdLS. **C,D.** Sample 4_ICF1. **E,F.** Sample 5_WHS. **G,H.** Sample 7_TBRS.



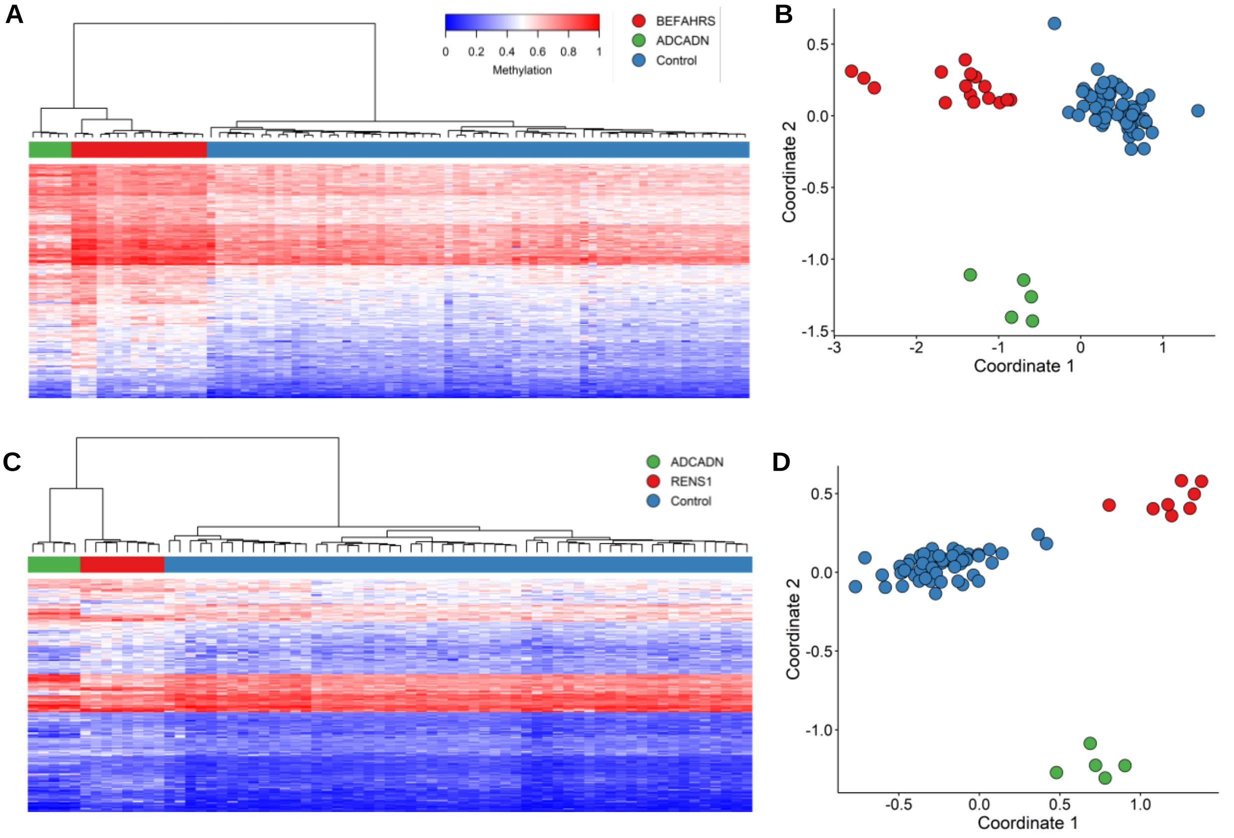
Supplementary Figure 6: Hierarchical clustering and MDS plots for the assessment of sample 3_RSTS1 for ARTHS, RSTS1, and MLASA2.



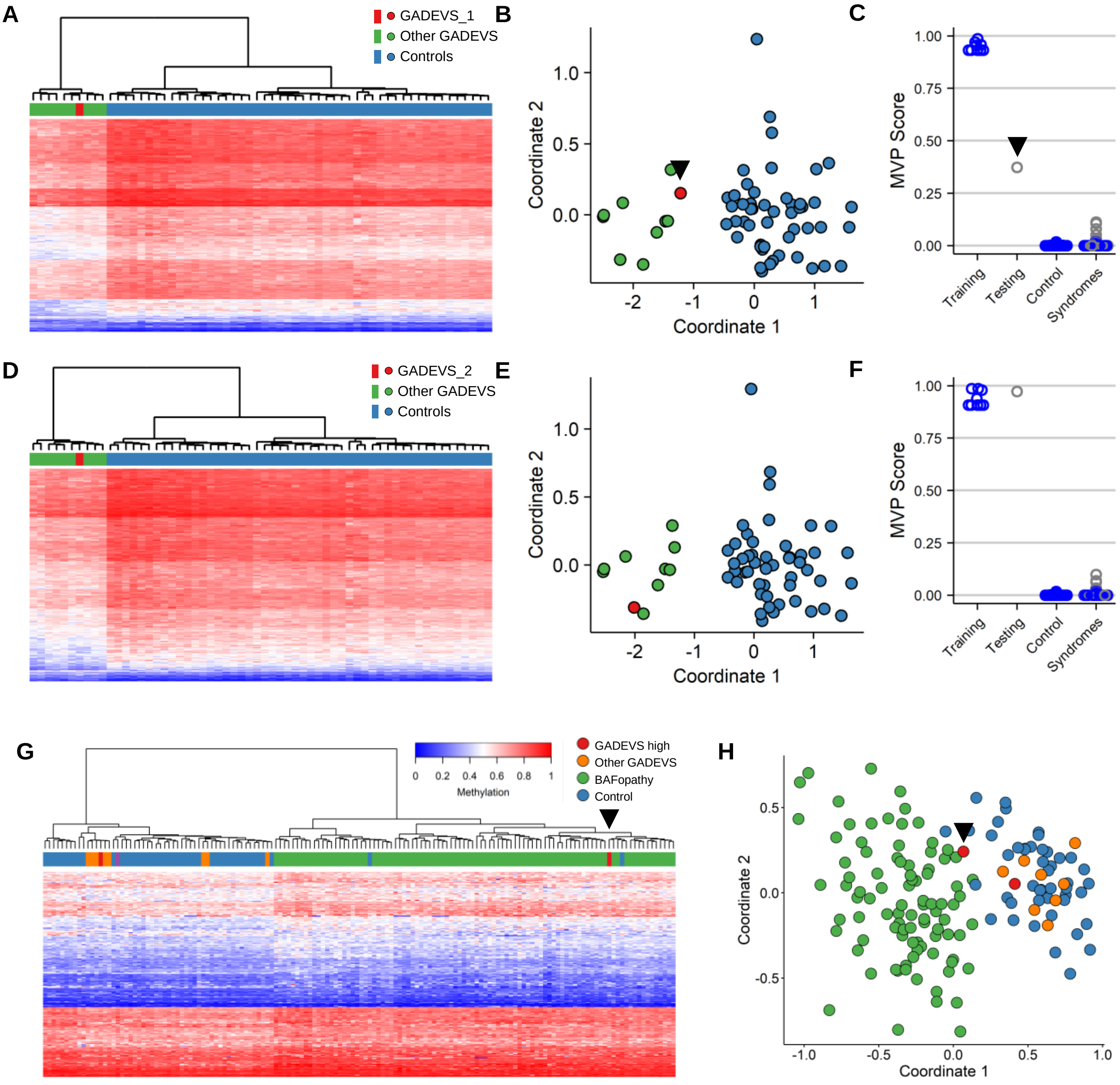
Supplementary Figure 7: Hierarchical clustering and MDS plots for the assessment of sample 6_IDDSELD for KDM2B-related syndrome.



Supplementary Figure 8: Assessment of sample 8_DYT28 for MLASA2 and DYT28. A. Hierarchical clustering and **B.** MDS plot for sample 8_DYT28 plotted using the MLASA2 episignature probes. **C.** Hierarchical clustering and **D.** MDS plot for sample 8_DYT28 plotted using the KDM2B episignature probes. **E-G.** Leave-one-out sample cross-validation of the DYT28 episignature. Sample 8_DYT28 is used as the testing sample, the 10 other DYT28 samples were used for training. **E.** Hierarchical clustering. **F.** MDS plot. **G.** SVM classifier results. Syndromes are samples from the 56 other neurodevelopmental syndromes with episignatures. Blue circles are samples used for classifier training (75% of samples), grey circles are samples used for classifier testing (25% of samples).



Supplementary Figure 9: Hierarchical clustering and MDS plots of ADCADN samples. A, B. using the BEFAHRS episignature probes**. C, D.** using the RENS1 episignature probes.



Supplementary Figure 10: Analysis of two GADEVS samples which had high MVP scores for BAFopathy. A-F. GADEVS leave-one-out sample cross-validation results, probe selection was repeated each time using nine samples with the one indicated sample used for testing. Sample GADEVS_1 is marked with a black arrowhead. **A,D.** Hierarchical clustering. **B,D.** MDS plot. **C,F.** SVM classifier results. Syndromes are samples from the 56 other neurodevelopmental syndromes with episignatures. Blue circles are samples used for classifier training (75% of samples), grey circles are samples used for classifier testing (25% of samples). **G**, Hierarchical clustering and **H,** MDS plot of all GADEVS samples using the BAFopathy episignature probes. Sample GADEVS_1 is marked with a black arrowhead as in A-C.