**Supplementary Data 1**

**Methylation status of *VTRNA2-1*/*nc886* is stable across populations, monozygotic twin pairs and in majority of tissues**

***Clustering of individuals based on nc886 methylation status***

As all datasets in GEO are not available as raw data, we wanted to establish a reproducible way, irrespective of normalization method used, to cluster the individuals to *nc886* methylation status groups. DNA methylation data GSE157896 (1), available in GEO (2) was downloaded as raw idat-files and extracted with minfi package function read.metharray.exp. Four normalization methods with default settings were tested: SWAN from both minfi (SWAN M) and wateRmelon (SWAN W), quantile normalization from minfi as well as dasen from wateRmelon (3–5). In addition, raw beta values were extracted by preprocessRaw-function from minfi package.

The differentially normalized datasets, along with raw data, were clustered with either hierarchical or k-means clustering to 2, 3 and 4 groups. For each cluster, the methylation median of *nc886* locus was calculated. Clusters where median beta value of *nc886* locus was >0.40 were interpreted to be imprinted, clusters where median of *nc886* locus was < 0.15 were interpreted to be non-methylated and clusters where median of *nc886* locus was 0.15-0.40 were interpreted to be intermediately methylated. For each normalization and clustering method we then compared whether individuals were in the same category (imprinted, intermediately methylated or non-methylated or only imprinted and other, i.e. either intermediately methylated and non-methylated combined) for each data normalization method.

With three groups (imprinted, intermediately methylated and nonmethylated) there were considerable inconsistencies between non-methylated and intermediately methylated groups with both clustering methods, with up to 5% of individuals in different methylation status groups across different normalization methods. With two groups (imprinted and other) and hierarchical clustering, there were still inconsistencies across normalization methods with up to 4% of individuals clustered to different *nc886* methylation status groups. However, with k-means clustering, only 2 (out of 1019, 0.02%) individuals were grouped to different methylation status groups across all methods (Supplementary Data Table SD1).

Supplementary Data Table SD1. Reproducibility of nc886 status groups across different normalization methods. Dataset GSE157896 was normalized with four different methods, and individuals were clustered to three groups with k-means clustering. From these groups, we identified the imprinted individuals and grouped non-methylated and intermediately methylated as ‘other’. Across different normalization methods, only 2 individuals out of 1019 (0.02%) were clustered to different nc886 methylation status groups.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | quantile | | SWAN M | | SWAN W | | raw | |
|  |  | imprinted | other | imprinted | other | imprinted | other | imprinted | other |
| dasen | imprinted | 694 | 0 | 694 | 0 | 694 | 0 | 694 | 0 |
| other | **1** | 324 | **1** | 324 | **1** | 324 | **2** | 323 |
| quantile | imprinted |  |  | 695 | 0 | 695 | 0 | 695 | 0 |
| other |  |  | 0 | 324 | 0 | 324 | **1** | 323 |
| SWAN M | imprinted |  |  |  |  | 695 | 0 | 695 | 0 |
| other |  |  |  |  | 0 | 324 | **1** | 323 |
| SWAN W | imprinted |  |  |  |  |  |  | 695 | 0 |
| other |  |  |  |  |  |  | **1** | 323 |

To further confirm this, we repeated the different normalization methods to dataset GSE125105 (6) and clustered the individuals as described above with k-means clustering to imprinted and ‘other’. In this dataset, no more than 5 (out of 699, 0.72%) individuals were grouped to different *nc886* methylation status groups across all methods (Supplementary Data Table SD2).

Supplementary Data Table SD2. Reproducibility of nc886 status groups across different normalization methods in GSE125105. Across different normalization methods, only 5 individuals out of 699 (0.72%) were clustered to different nc886 methylation status groups.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | quantile | | SWAN M | | SWAN W | | raw | |
|  |  | imprinted | other | imprinted | other | imprinted | other | imprinted | other |
| dasen | imprinted | 515 | **5** | 519 | **1** | 519 | **1** | 519 | **1** |
| other | 0 | 179 | 0 | 179 | 0 | 179 | 0 | 179 |
| quantile | imprinted |  |  | 515 | 0 | 515 | 0 | 515 | 0 |
| other |  |  | **4** | 180 | **4** | 180 | **4** | 180 |
| SWAN M | imprinted |  |  |  |  | 519 | 0 | 519 | 0 |
| other |  |  |  |  | 0 | 180 | 0 | 180 |
| SWAN W | imprinted |  |  |  |  |  |  | 519 | 0 |
| other |  |  |  |  |  |  | 0 | 180 |

Therefore, all datasets utilized in this study were clustered with k-means clustering to three groups, from which the imprinted clusters (median *nc886* beta value > 0.40) were identified, other clusters for each data set were combined to category ‘other’.

**References**

1. Huang JY, Cai S, Huang Z, Tint MT, Yuan WL, Aris IM, et al. Analyses of child cardiometabolic phenotype following assisted reproductive technologies using a pragmatic trial emulation approach. Nat Commun. 2021 Sep 23;12(1):5613.

2. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002 Jan 1;30(1):207–10.

3. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014 May 15;30(10):1363–9.

4. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics. 2013 May 1;14(1):293.

5. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):R80.

6. Arloth J, Eraslan G, Andlauer TFM, Martins J, Iurato S, Kühnel B, et al. DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning. PLOS Computational Biology. 2020 Feb 3;16(2):e1007616.