Application Note

Agilent
Trusted Answers

Detecting Systemic Interindividual Epigenetic Variation With the Agilent SureSelect CD Baylor Human CoRSIV Panel

Authors

C.J. Gunasekara
USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX

Yumei Li
Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

C. Coarfa
Department of Molecular and Cellular Biology and Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX

R.A. Waterland
USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

Summary

Target-capture bisulfite sequencing of correlated regions of systemic interindividual epigenetic variation (CoRSIVs) allows researchers to use easily biopsied tissues like peripheral blood to study systemic epigenetic differences among individuals and identify disease-associated epigenetic variation.

Introduction

There has been growing interest in the field of epigenetic epidemiology, defined as the study of the associations between interindividual epigenetic variation, specifically DNA methylation, and the risk of disease[1]. Compared to genetic epidemiology, epigenetic epidemiology presents major challenges. Much DNA methylation is cell type-specific, so it is difficult to generally 'epigenotype' an individual using peripheral blood DNA. Also, the disease process itself can induce epigenetic changes, compromising the ability to draw causal inferences. Tissue-independent systemic interindividual variants (SIV) in DNA methylation provide a major opportunity to advance the field of epigenome-wide association studies (EWAS). Because the establishment of DNA methylation at correlated regions of systemic interindividual variation (CoRSIVs) is sensitive to the periconceptional environment, CoRSIV-capture technology will also benefit studies of developmental programming.

Over the last decade, numerous studies have discovered and validated the existence of SIV in DNA methylation.[3,4] A 2019 study based on whole genome bisulfite sequencing identified 9,926 CoRSIVs.[2] Now, with the development of the Agilent SureSelect CD Baylor Human CoRSIV panel, researchers can optimally assay SIV in large population studies to discover associations with human diseases. By multiplexing at the capture step, this Agilent capture protocol provides a cost-effective solution, enabling genome-scale, deep bisulfite sequencing data at a price competitive with commercial methylation arrays.

Materials and Methods

SureSelect target enrichment bait design

SureSelect CoRSIV Capture designs were developed and evaluated iteratively by expanding content and optimizing Agilent SureDesign criteria. The capture designs were generated using the Agilent SureDesign custom design tool (https://earray. chem.agilent.com/suredesign/) with 1X tiling density. Design selections specific to each version are provided in Table 1.

Table 1. SureSelect Capture Design versions and selections.

V.	Agilent Design ID	Design Date	Included Regions	Agilent Design Selections	Size (Mbp)
1.0	S3163502	23 August 2018	4641 gene-associated CoRSIVs Sex-specific Chr Y regions	1X Tiling Density Least Stringent Masking Balanced Boosting	9.0
2.0	S3223244	16 July 2019	9926 CoRSIVs[2] SIV[3,4] ESS[4] Sex-specific Chr Y regions, imprinting control regions[5]	1X Tiling Density Least Stringent Masking Balanced Boosting	19.8

* Design ID S3449736 in the 'Published Designs' in the public SureDesign interface.

Target capture, library preparation, and sequencing

Target capture and library preparation were performed according to the Agilent SureSelect XT Methyl-Seq Library Preparation kit, with modifications. In brief, one microgram of genomic DNA was subject to shearing to 150 to 200 bp in size using a Covaris sonicator. After purification through AMPure XP beads (Beckman Coulter), end repair and A-tailing were performed. Then, five microliters of 15 µM methylated library adaptor (IDT) was ligated to each sample, and the product with a size of 250 to 450 bp was selected using AMPure XP beads.

Following the Agilent protocol, twelve libraries were multiplexed in equal proportions for each capture. The SureSelect XT Methyl-Seq target enrichment system was used for Illumina multiplexed sequencing. Target enrichment was performed using Agilent SureSelectXT CD Baylor Human CoRSIV panel v1.0. After hybridization with probes, Dynabeads MyOne Streptavidin T1 beads (Thermo Fisher Scientific) were used to bind the library. After several rounds of washes, the bound DNA was eluted in 0.1 N NaOH and subjected to bisulfite treatment using the EZ DNA Methylation Gold kit (Zymo Research). The final library was generated by amplification using SureSelect Methyl-seq PCR Master Mix and P5, P7 primers (Illumina). Sequencing was performed using an Illumina Novaseq 6000 (150 bp paired-end reads). On average, approximately 30 M reads were mapped per library (Figure 1A) with approximately 81% mapping efficiency (Figure 1B).

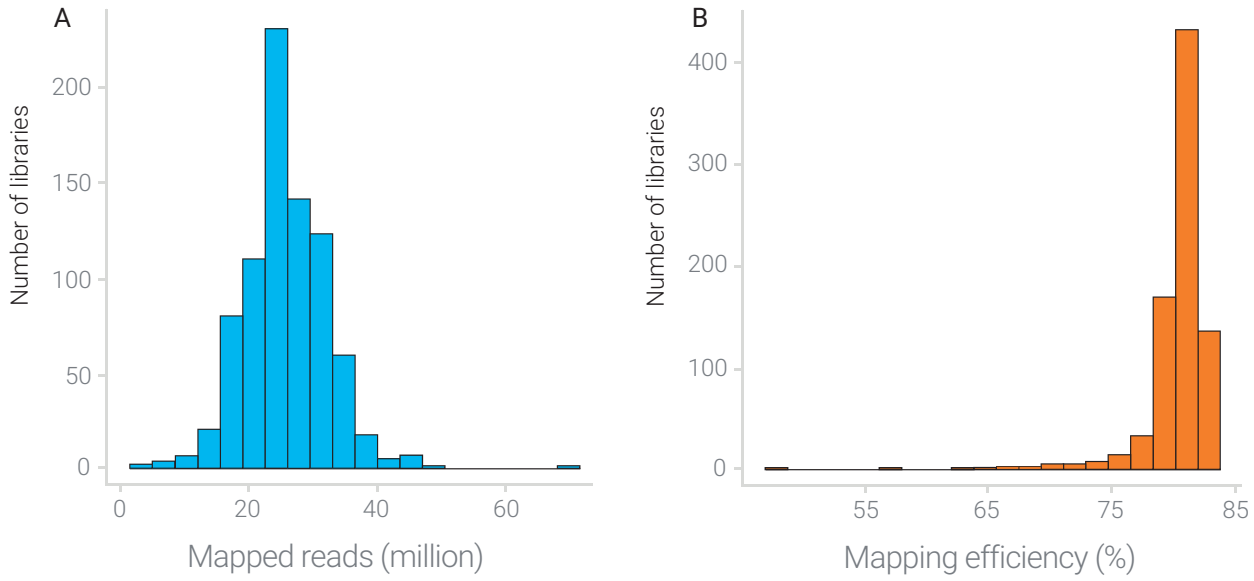

Figure 1. (A) Histogram of the number of mapped reads per library. (B) Bismark mapping efficiency per library (From Ref. 6).

Data analysis

Sequence reads were trimmed using Trim Galore (Babraham Bioinformatics), then mapped to the human genome build hg38 using the Bismark aligner (Babraham Bioinformatics). Uniquely mapped reads were retained for further analysis. As recommended for capture experiments by the Bismark manual, duplicate reads were not removed. CpG-level methylation was quantified using the Bismark pipeline. For each sample, average proportional DNA methylation was computed at each CoRSIV for which at least half of the CpGs are covered by at least five reads (for CoRSIVs with only two CpGs, both were required to be covered by at least five reads). To confirm the accuracy of the biological sex of each sample, the coverage of chromosome Y control regions was assessed.

Results

Multiple human tissue types from 188 GTEx donors

To confirm systemic DNA methylation at CoRSIVs, multiple tissues from 188 donors in the NIH Genotype-Tissue Expression (GTEx) program (807 samples total) were studied.[6] The tissues represented all three embryonic germ layer lineages (Figure 2).

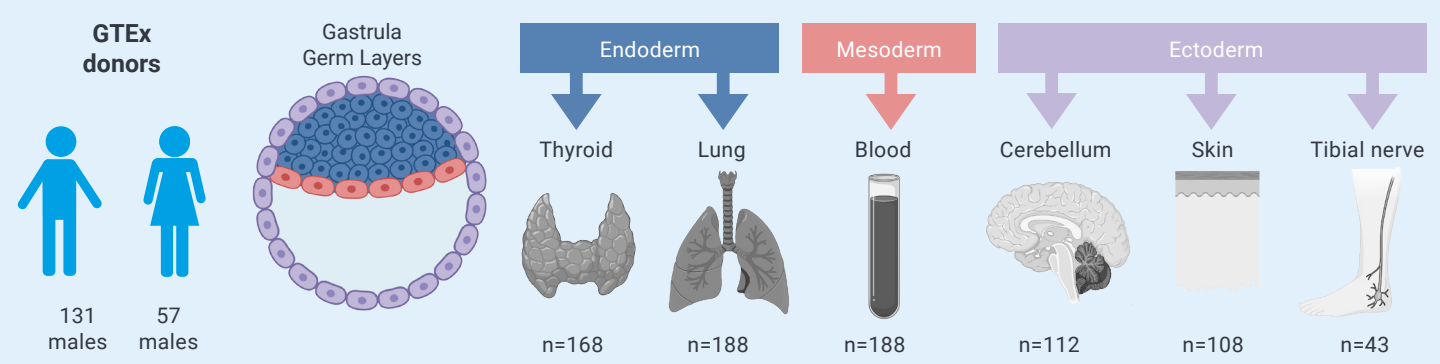

Figure 2. DNA samples were obtained from multiple tissues (representing the three embryonic germ layers) from each of 188 GTEx donors (From Ref. 8).

For this study, Agilent SureSelect CD Baylor Human CoRSIV panel V1.0 was used. To ensure adequate targeting, (out of the 9,926 CoRSIVs previously reported[2]) we filtered to include only the CoRSIVs within 3,000 base pairs from the body of a gene present in the Pubtator compendium[7] using BEDTOOLS software. This yielded 4,641 CoRSIVs as targets for capture. At each of the 4,641 CoRSIVs, the target region included flanking regions of 1,000 bp in each direction.

Target capture efficiency

To determine the proportion of on-target reads, only those that mapped completely within a target region were counted; capture efficiency was calculated as the fraction of on-target reads divided by all uniquely mapped reads. Across all 807 samples, the median capture efficiency was over 45% (Figure 3A). This level of enrichment enabled excellent coverage depth across the majority of CoRSIVs targeted (Figure 3B).

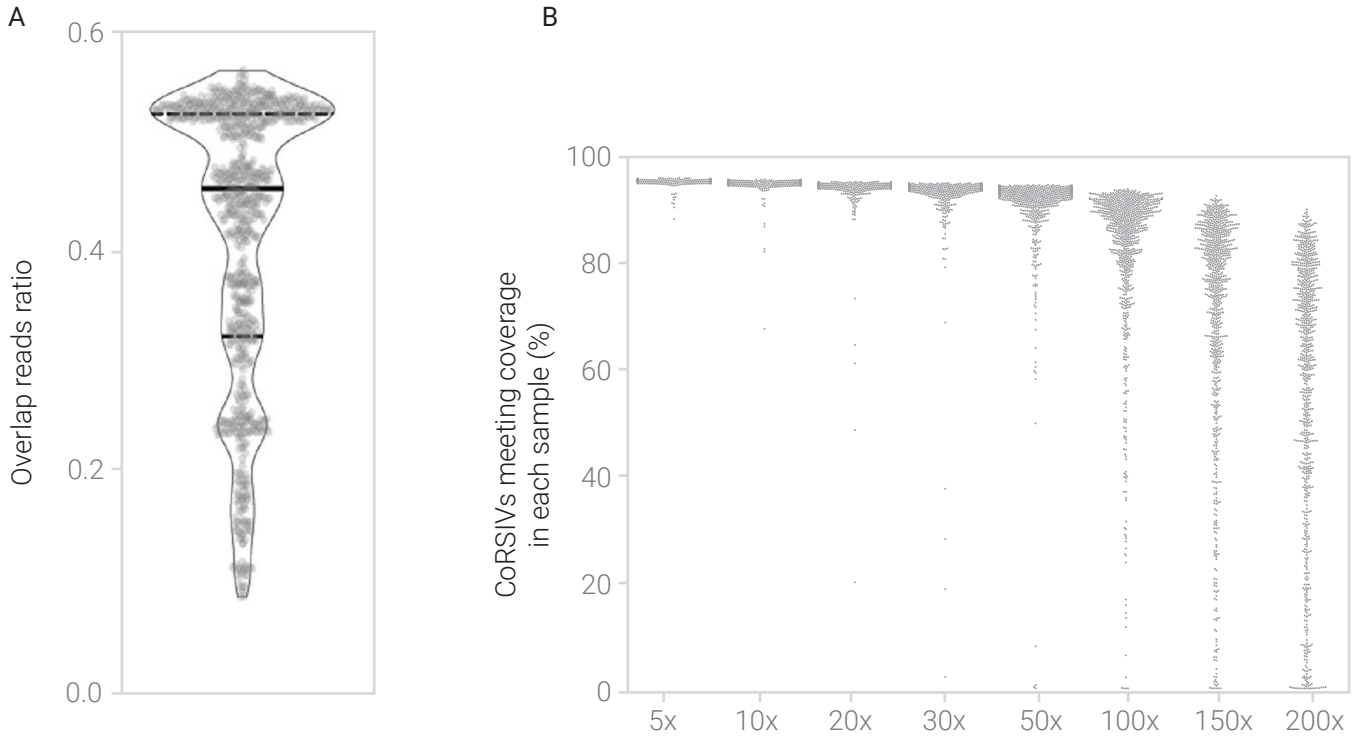

Figure 3. (A) Violin plot shows, for each of 807 samples, the proportion of reads that were on target (that is, completely within a target region). (B) Percentage of CoRSIVs for which target-capture bisulfite sequencing achieved various read depths; each point represents one of 807 samples (From Ref. 8).

Confirmation of systemic interindividual variation at CoRSIVs

A major rationale for studying CoRSIVs is that easily accessible tissues such as blood can be used to infer methylation in other internal tissues. High inter-tissue correlations are the hallmark of systemic interindividual variation. Scatter plots in Figure 4A represent an example of inter-tissue correlations at one CoRSIV at HPCAL1. Each point in the scatter plots represents an individual. Figure 4B summarizes all inter-tissue correlations evaluated in 4,086 CoRSIVs with adequate coverage in all tissues. The majority (~83%) of CoRSIVs show an average inter-tissue correlation (Pearson R) above 0.6.

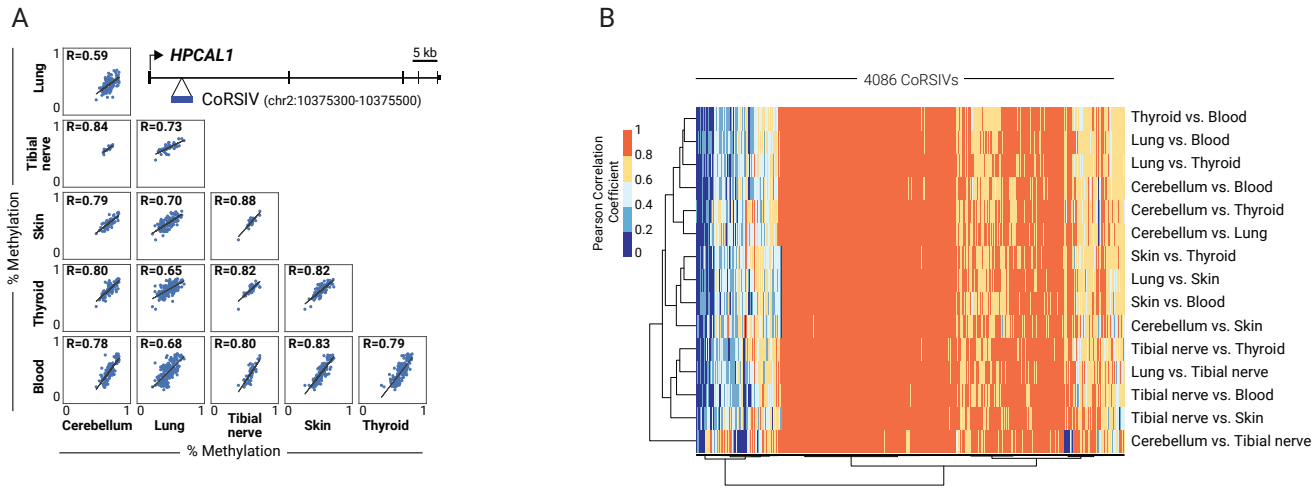

Figure 4. (A) Scatter plots between all possible tissue pairs illustrate high inter-tissue correlations at a CoRSIV within HPCAL1. Each point represent an individual (B) A heat map of inter-tissue correlations across 4,086 CoRSIVs shows generally high correlation coefficients between all possible tissue pairs (From Ref. 8).

DNA methylation in blood can be used as a proxy for DNA methylation in internal organs at CoRSIVs

At CoRSIVs, tissue-level methylation profiles from various individuals cluster by the individual, demonstrating an epigenetic level of individuality.[8] Accordingly, in these regions, blood can be used as a proxy for internal organs (Figure 5).

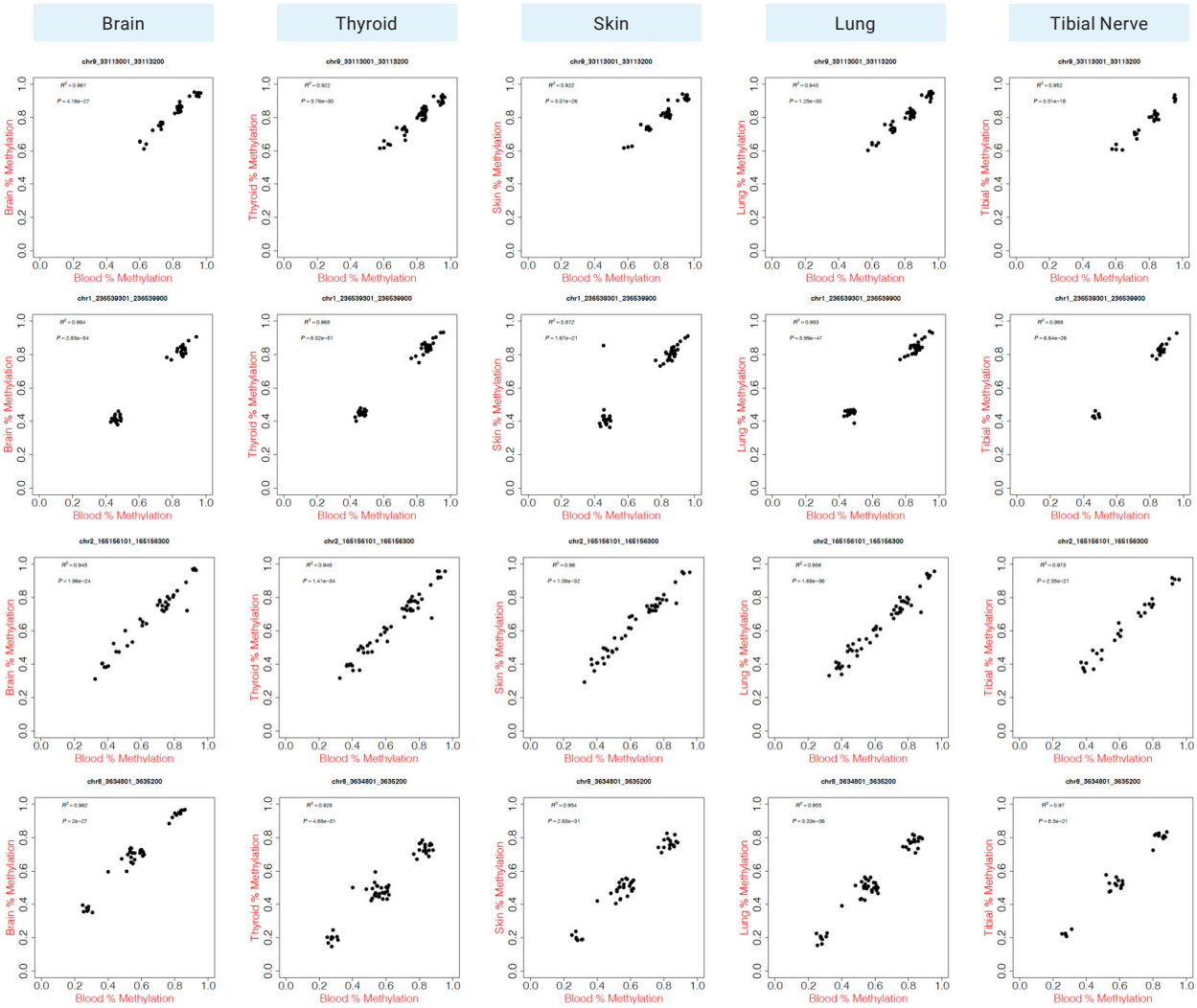

Figure 5. Each row represents DNA methylation data from one targeted CoRSIV. The X-axis is DNA methylation at the CoRSIV in whole blood and Y-axes indicate DNA methylation at the same CoRSIV measured in the brain, thyroid, skin, lung, and tibial nerve tissues. Each point represents one individual (From Ref. 8).

Conclusions

Agilent SureSelect CD Baylor Human CoRSIV panel is reliable and effective for measuring DNA methylation in CoRSIV regions. DNA methylation at CoRSIVs is established in early development and influenced by the periconceptional environment and maternal nutrition. Studying CoRSIVs on a population scale may provide a major opportunity to advance the field of epigenetic epidemiology linking early environmental exposures to later life health outcomes.

Validation results based on large population studies show that methylation at CoRSIVs in peripheral tissues correlates reliably with CoRSIV methylation in internal organs. Genetic influences on CoRSIVs are extremely strong[6], enabling a powerful new approach for mQTL analysis. Due to its excellent capture efficiency (~50%), this Agilent SureSelect panel can provide coverage of at least 20X sequencing depth at over 90% of targeted CoRSIVs, with only 30M 150 bp paired-end reads per library. Sample multiplexing at the capture level can further lower the per-library cost allowing a much larger sample size to boost statistical power. This panel has been refined over many iterations while constantly improving by the addition of more regions. Future discoveries of additional CoRSIVs will be added to subsequent versions.

References

1. Waterland, R. A.; Michels, K. B. Epigenetic Epidemiology of the Developmental Origins Hypothesis. *Annu. Rev. Nutr.* **2007**, *27* (1), 363–388. https://doi.org/10.1146/annurev.nutr.27.061406.093705.
2. Gunasekara, C. J.; Scott, C. A.; Laritsky, E.; Baker, M. S.; MacKay, H.; Duryea, J. D.; Kessler, N. J.; Hellenthal, G.; Wood, A. C.; Hodges, K. R.; Gandhi, M.; Hair, A. B.; Silver, M. J.; Moore, S. E.; Prentice, A. M.; Li, Y.; Chen, R.; Coarfa, C.; Waterland, R. A. A Genomic Atlas of Systemic Interindividual Epigenetic Variation in Humans. *Genome Biol.* **2019**, *20* (1), 105. https://doi.org/10.1186/s13059-019-1708-1.
3. Silver, M. J.; Kessler, N. J.; Hennig, B. J.; Dominguez-Salas, P.; Laritsky, E.; Baker, M. S.; Coarfa, C.; Hernandez-Vargas, H.; Castelino, J. M.; Routledge, M. N.; Gong, Y. Y.; Herceg, Z.; Lee, Y. S.; Lee, K.; Moore, S. E.; Fulford, A. J.; Prentice, A. M.; Waterland, R. A. Independent Genomewide Screens Identify the Tumor Suppressor VTRNA2-1 as a Human Epiallele Responsive to Periconceptional Environment. *Genome Biol.* **2015**, *16* (1), 118. https://doi.org/10.1186/s13059-015-0660-y.
4. Van Baak, T. E.; Coarfa, C.; Dugué, P.-A.; Fiorito, G.; Laritsky, E.; Baker, M. S.; Kessler, N. J.; Dong, J.; Duryea, J. D.; Silver, M. J.; Saffari, A.; Prentice, A. M.; Moore, S. E.; Ghantous, A.; Routledge, M. N.; Gong, Y. Y.; Herceg, Z.; Vineis, P.; Severi, G.; Hopper, J. L.; Southey, M. C.; Giles, G. G.; Milne, R. L.; Waterland, R. A. Epigenetic Supersimilarity of Monozygotic Twin Pairs. *Genome Biol.* **2018**, *19* (1), 2. https://doi.org/10.1186/s13059-017-1374-0.
5. Monk, D.; Morales, J.; den Dunnen, J. T.; Russo, S.; Court, F.; Prawitt, D.; Eggermann, T.; Beygo, J.; Buiting, K.; Tümer, Z.; the Nomenclature group of the European Network for Human Congenital Imprinting Disorders. Recommendations for a Nomenclature System for Reporting Methylation Aberrations in Imprinted Domains. *Epigenetics* **2018**, *13* (2), 117–121. https://doi.org/10.1080/15592294.2016.1264561.
6. Gunasekara, C. J.; MacKay, H.; Scott, C. A.; Li, S.; Laritsky, E.; Baker, M. S.; Grimm, S. L.; Jun, G.; Li, Y.; Chen, R.; Wiemels, J. L.; Coarfa, C.; Waterland, R. A. Systemic Interindividual Epigenetic Variation in Humans Is Associated with Transposable Elements and under Strong Genetic Control. *Genome Biol.* **2023**, *24* (2). https://genomebiology.biomedcentral.com/articles/10.1186/s13059-022-02827-3.
7. Wei, C.-H.; Kao, H.-Y.; Lu, Z. PubTator: A Web-Based Text Mining Tool for Assisting Biocuration. *Nucleic Acids Res.* **2013**, *41* (Web Server issue), W518-22. https://doi.org/10.1093/nar/gkt441.
8. Koch, L. A Map of Human Individuality. *Nat. Rev. Genet*. **2019**, *20* (8), 435. https://doi.org/10.1038/s41576-019-0149-8.

Acknowledgements

Identification of human CoRSIVs and development of these panels were conducted at Baylor College of Medicine in the USDA/ARS Children's Nutrition Research Center in Houston, Texas. Over the past 15 years, this work has also been supported by the March of Dimes, the UK Medical Research Council, the Bill & Melinda Gates Foundation, the National Institutes of Health, and the Cancer Prevention and Research Institute of Texas.

www.agilent.com

For Research Use Only. Not for use in diagnostic procedures.

PR7001-0703

Published in the USA, September 01, 2023
5994-5914EN

Agilent
Trusted Answers