

Long-Read Sequencing: A New Frontier for Estimating Epigenetic Age

Shijia Zhu, Rob Warren, Chandra Muller, Eric Grodsky, Adam Brickman, Jennifer Manly, Stefani Thomas, Eileen Crimmins, Jessica Faul, Bharat Thyagarajan.

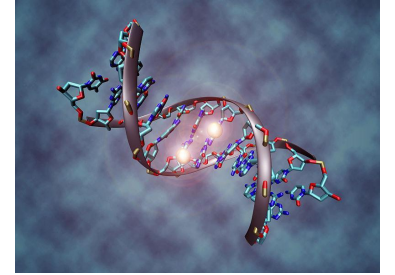
BACKGROUND: EPIGENETICS

Epigenetics: Changes in gene expression that do not involve change in genetic sequence.

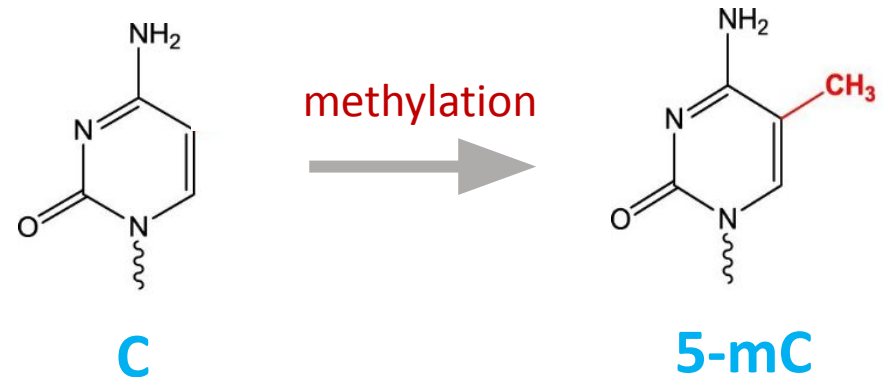
DNA Methylation: A crucial epigenetic modification where a methyl group is added to a cytosine base, often in CpG dinucleotides.

- This plays a vital role in gene regulation, development, and disease.

DNA sequences
G, A, T, C



Chemical modifications
to DNA



METHODS: SEQUENCE VARIANTS VS. METHYLATION

DNA sequences (G/A/T/C)

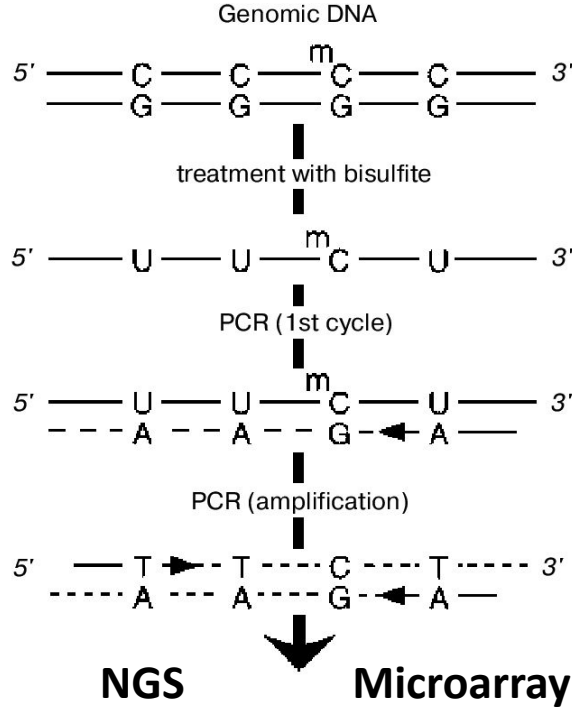


Next Gen Sequencing (NGS)

DNA methylation (5mC)



Bisulfite conversion



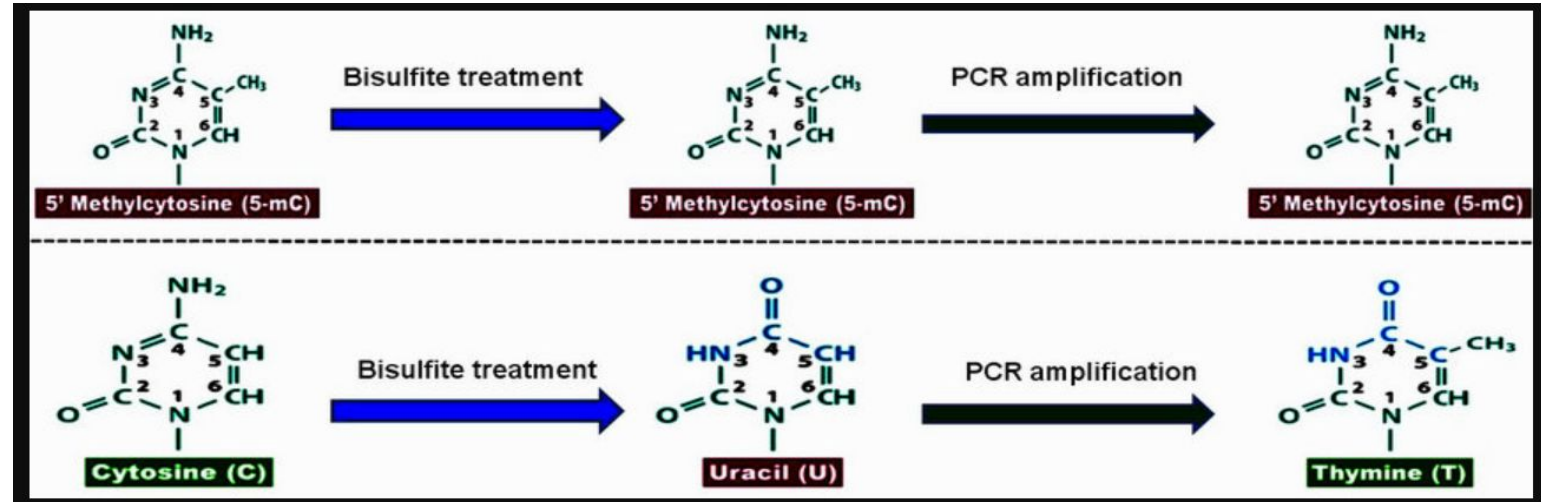
OR



METHYLATION EPIC BEADCHIPS

• PRINCIPLE

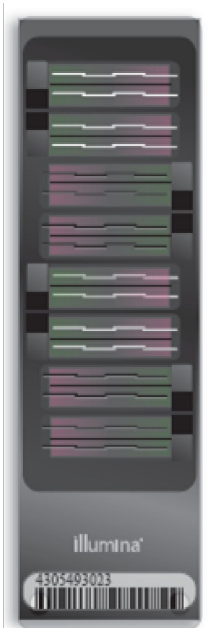
- Sodium Bisulfite: Converts unmethylated cytosines to uracil while methylated Cytosines remain unchanged.



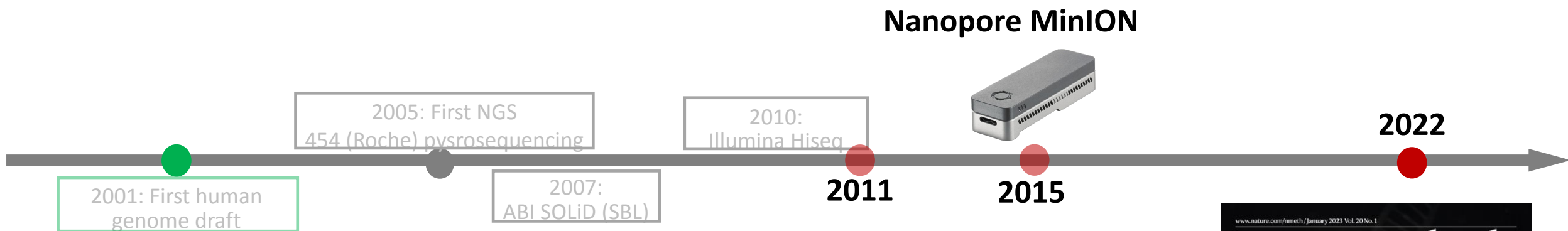
- > **850k** methylation sites at single-nucleotide resolution

- Promoters/enhancers of FANTOM5, ENCODE, Dnase, miRNA
- Tumor, human stem cell, and tissue-specific CpG sites
- widely used for EWAS

MethylationEPIC
BeadChip



LONG READ SEQUENCING (LRS)



2001: First human genome draft

2005: First NGS 454 (Roche) pyrosequencing

2007: ABI SOLiD (SBL)

2010: Illumina HiSeq

PacBio RS



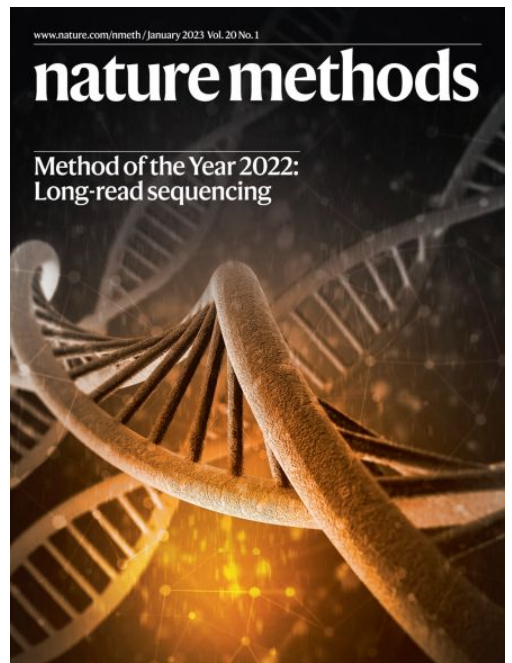
2011

2015

Nanopore MinION

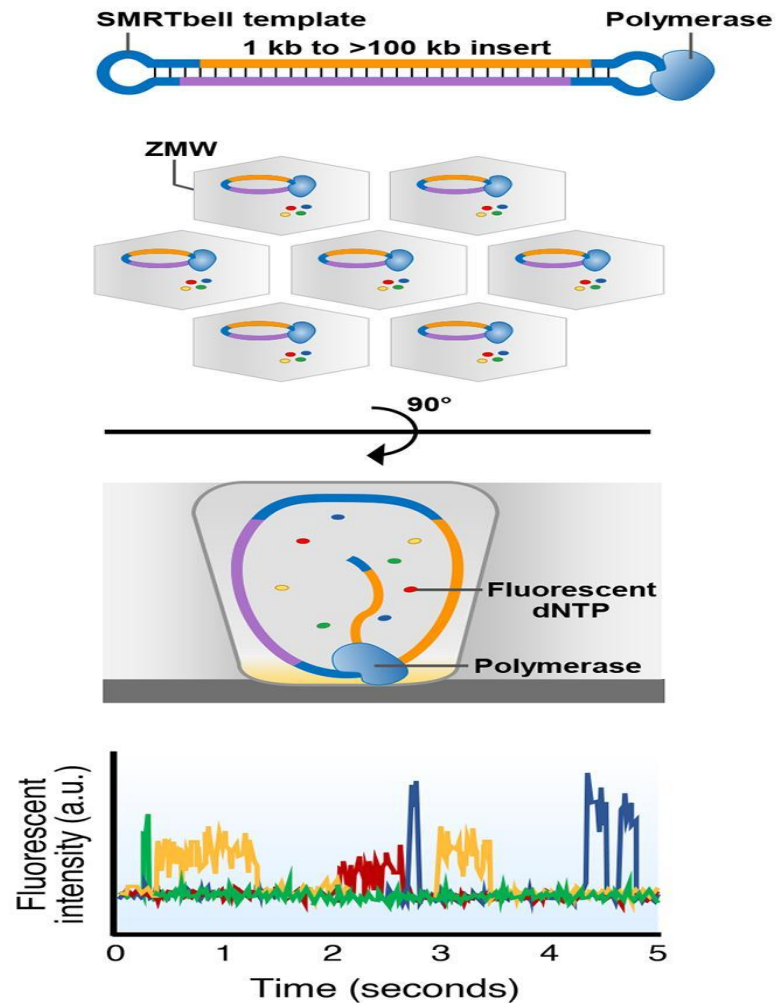


2022

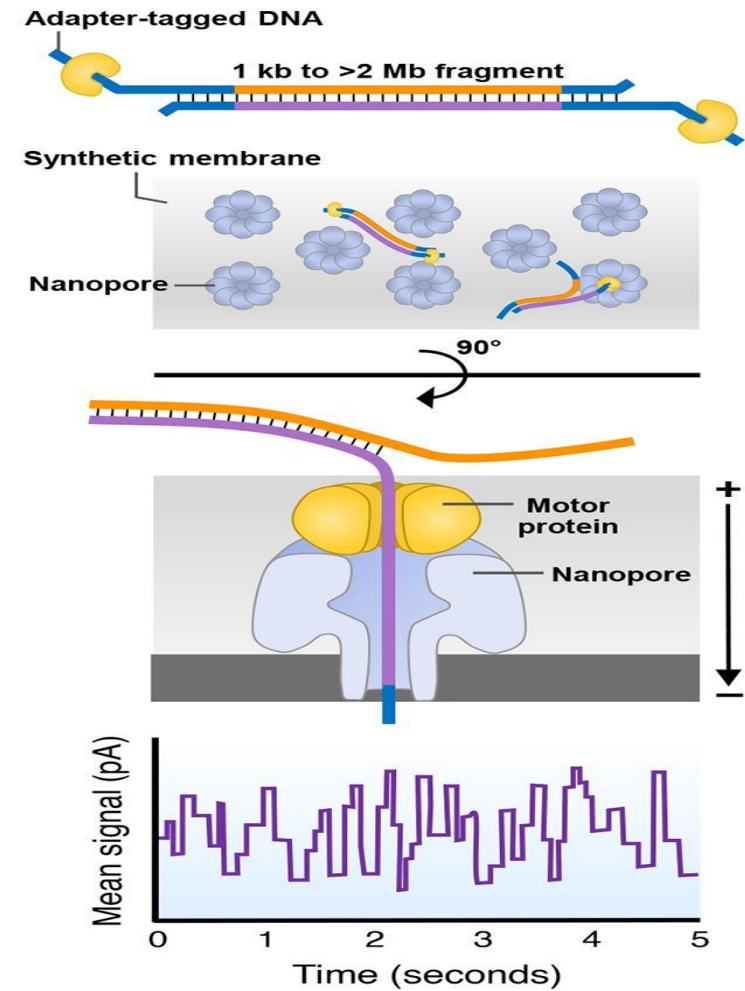


LONG READ SEQUENCING (LRS) TECHNOLOGIES

a Pacific Biosciences SMRT sequencing



b Oxford Nanopore Technologies sequencing



COMPARISON OF LRS VS. ARRAYS

| Feature | Array-Based Methods | Oxford Nanopore Sequencing |
|-----------------------|--|---------------------------------------|
| Principle | Bisulfite conversion & Hybridization | Direct electrical signal measurement |
| Methylation Detection | Indirect (after conversion) | Direct (native DNA) |
| Cytosine Modification | Cannot distinguish 5mC from 5hmC | Can distinguish various modifications |
| Coverage | Limited to pre-selected sites (changes over time) | Genome-wide |
| Throughput | High | Lower (per run) |
| Cost (large studies) | Generally lower | Higher (per sample, can vary) |
| DNA Input | Lower | Can be higher, but improving |
| Data Analysis | Relatively straightforward | More complex |
| Haplotype Phasing | Limited | Excellent due to long reads |
| Structural Variation | Not directly assessed | Well-suited for detection |
| Repetitive Regions | Poorly covered | Well-covered |

**CAN LRS BE USED IN
POPULATION STUDIES TO STUDY
DNA METHYLATION CHANGES?**

STUDY DESIGN

50 samples from Health and Retirement Study (HRS)
275 samples from High School and Beyond (HS&B)

HRS samples analyzed using EPIC v.1.0 array
HS&B samples analyzed using EPIC v2.0 array

All samples were analyzed using Long
Read Sequencing (LRS): Oxford Nanpore

DNA methylation at 850K CpG sites

DNA methylation at 75M CpG sites

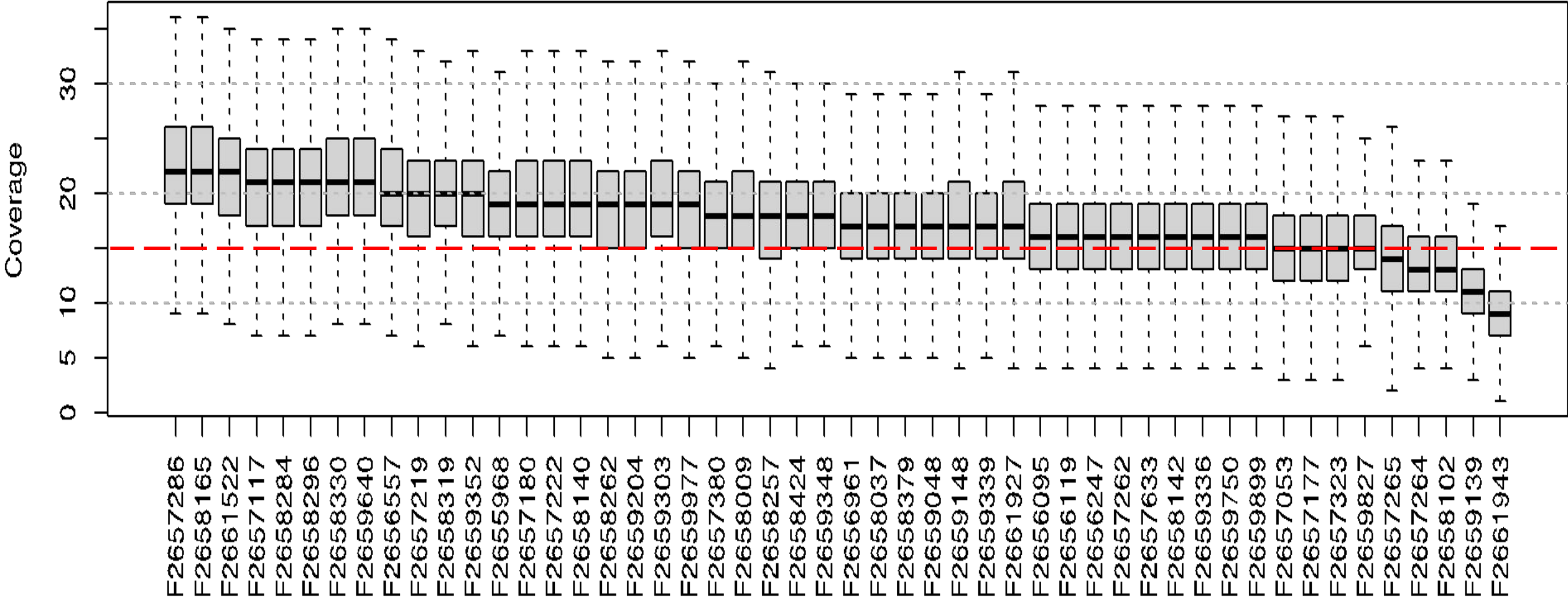
Correlation among the common CpG sites
Compare Epigenetic Age using both methods

LRS reproducibility: 6 duplicate samples

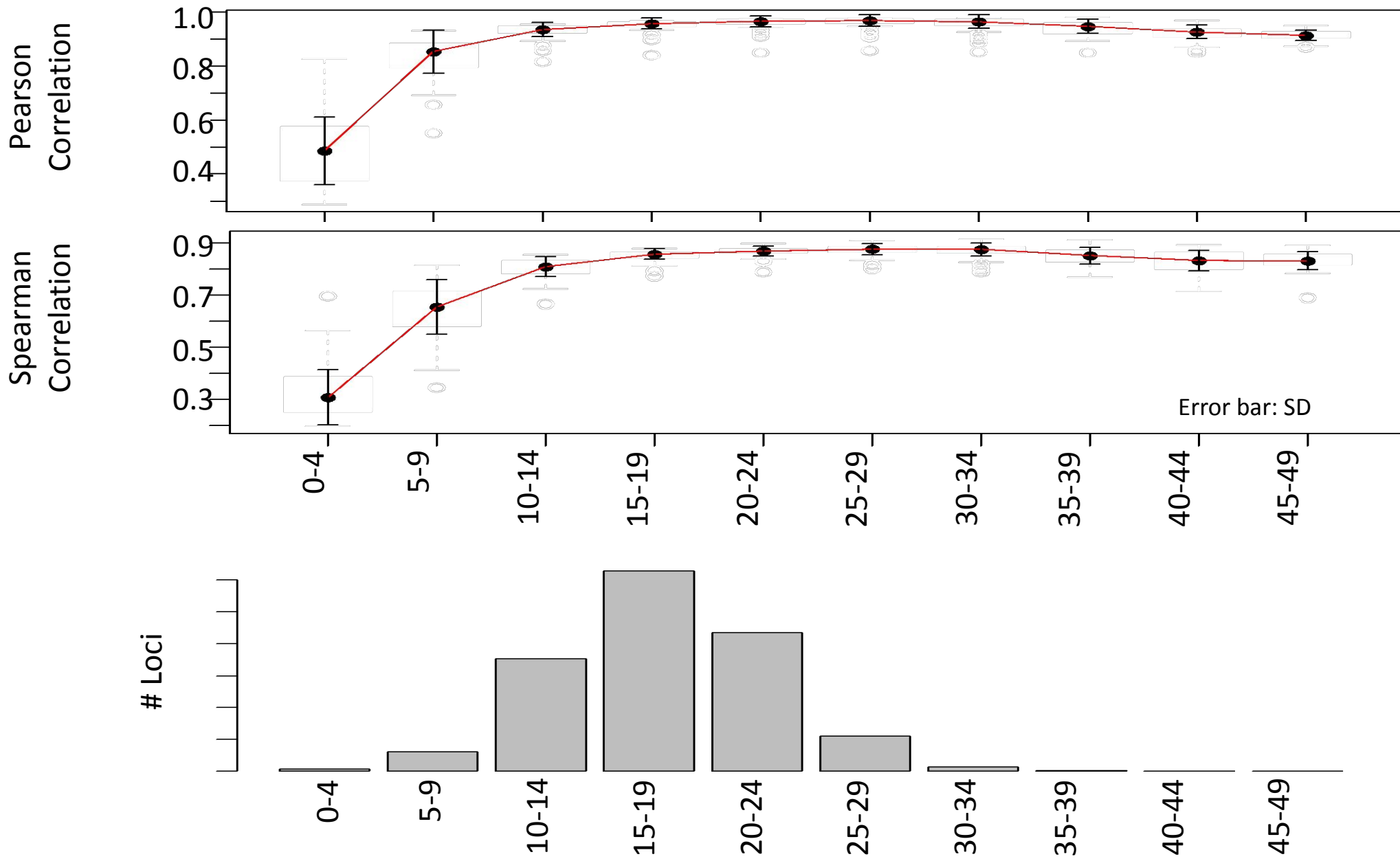
METHODS

- We used Dorado to do basecalling, and next used Modkit to call DNA methylation
- For both LRS and EPIC array data, we created 5 epigenetic clocks using the Principal Component (PC) clocks for comparison
 - Horvath 1
 - Horvath 2
 - Hannum
 - Phenoage
 - GrimAge2
- We evaluated the correlation in epigenetic age between the two methods in HRS and HS&B separately

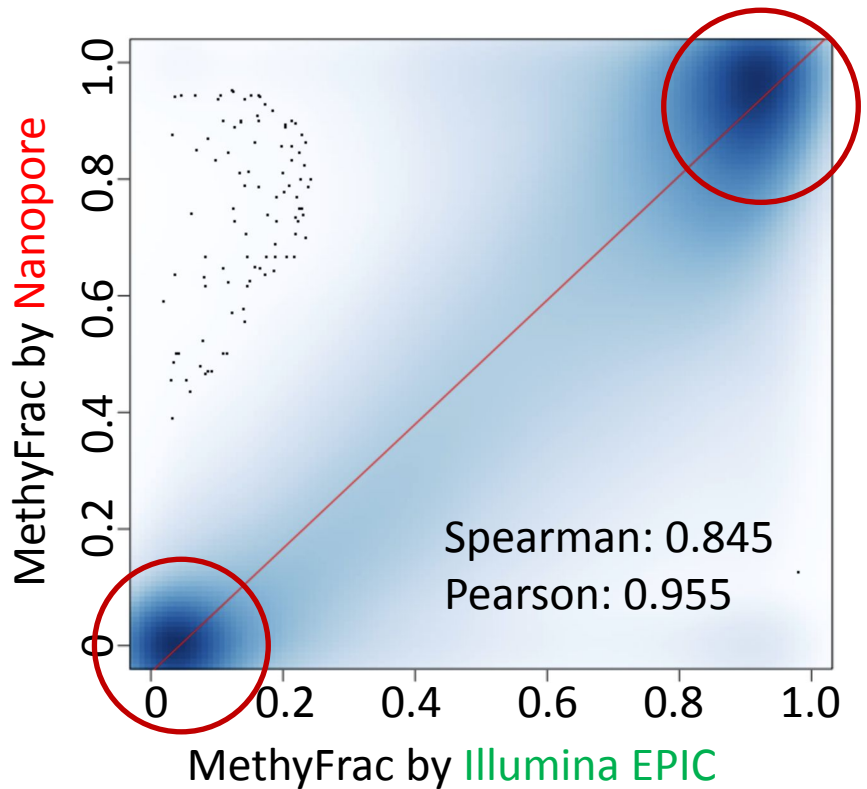
COVERAGE OF HRS SAMPLES



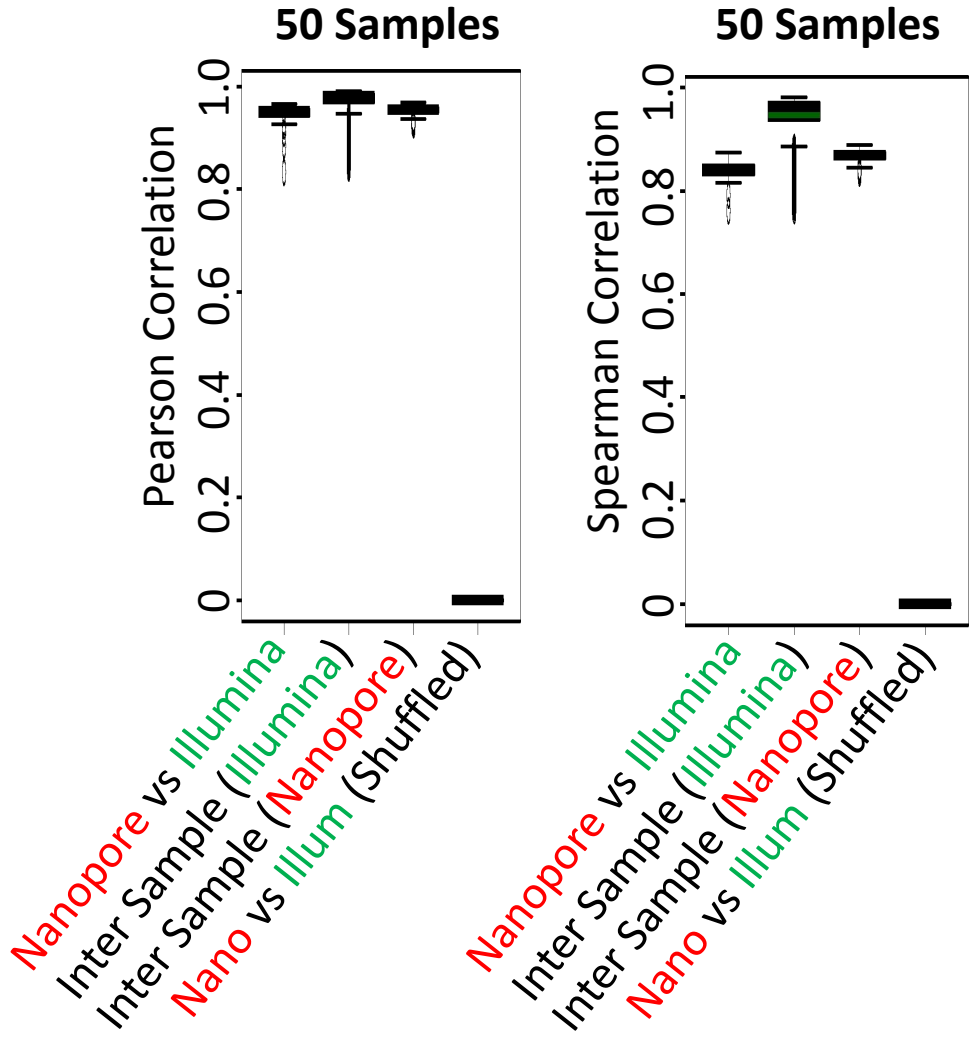
HRS RESULTS: ACCURACY AND READ COVERAGE



HRS: PLATFORM COMPARISON

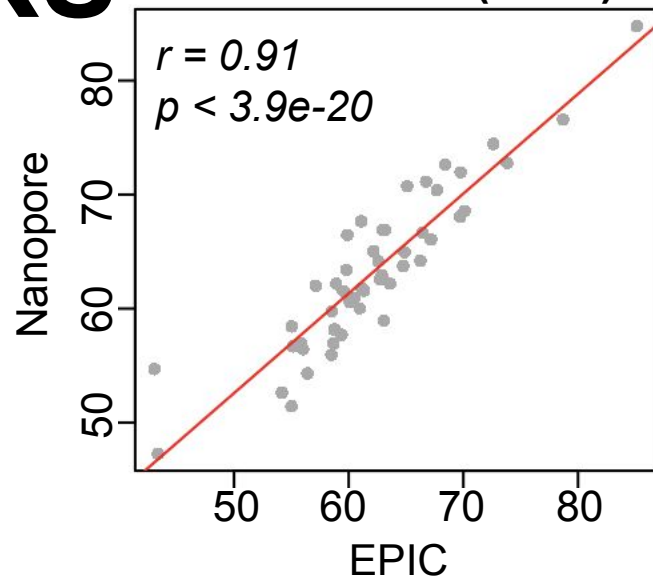


- **Illumina EPIC**: 836,461 loci
- **Nanopore**: 75,968,092 loci
- **Overlap**: 831,875

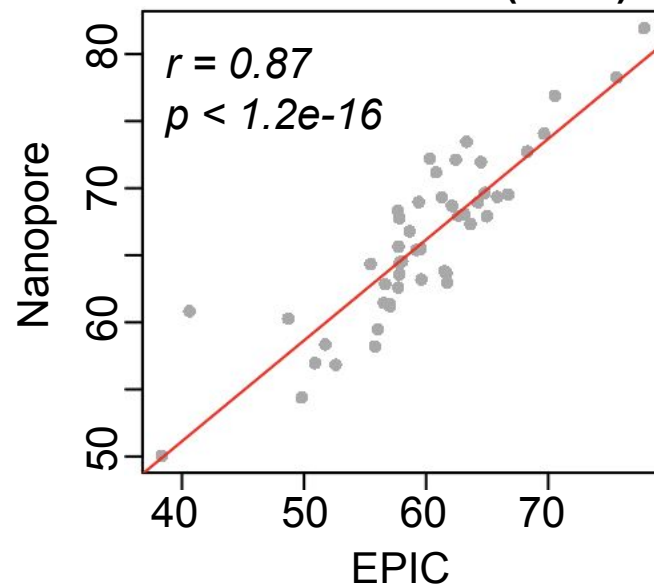


HRS

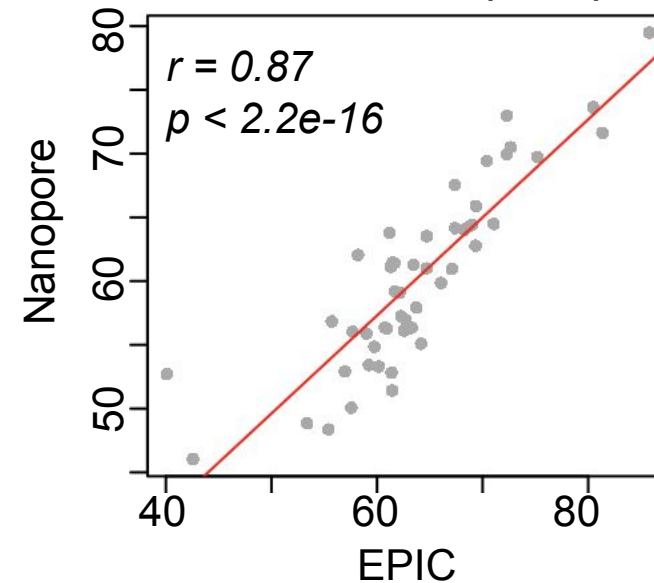
PCHorvath1 (n=50)



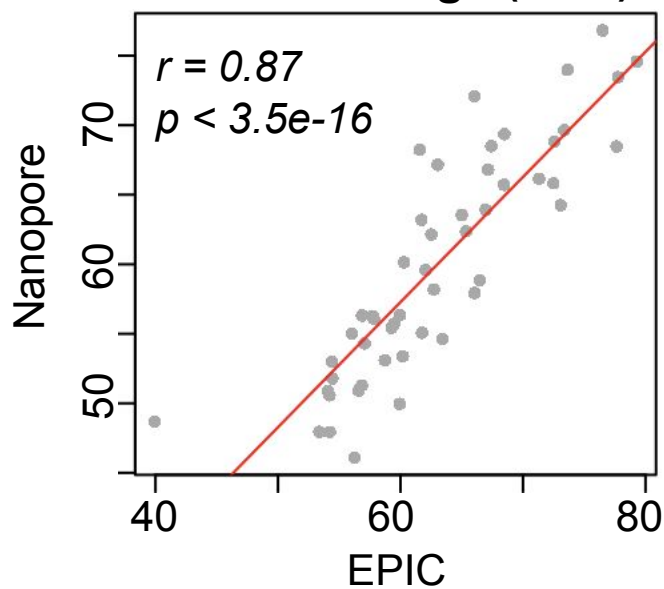
PCHorvath2 (n=50)



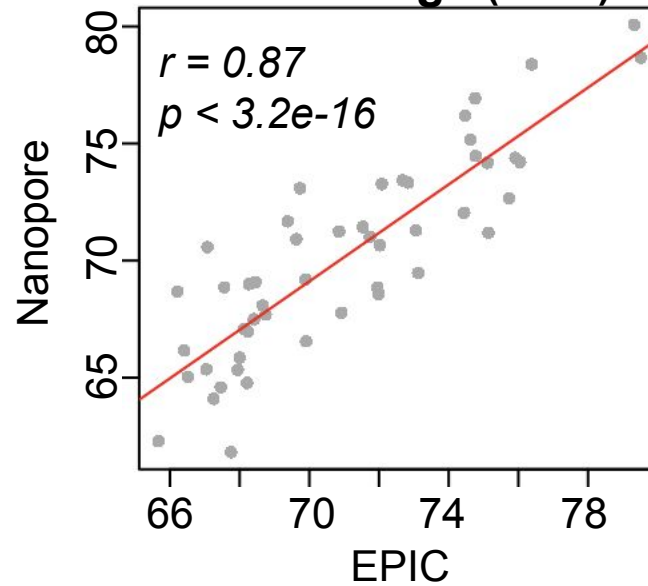
PCHannum (n=50)



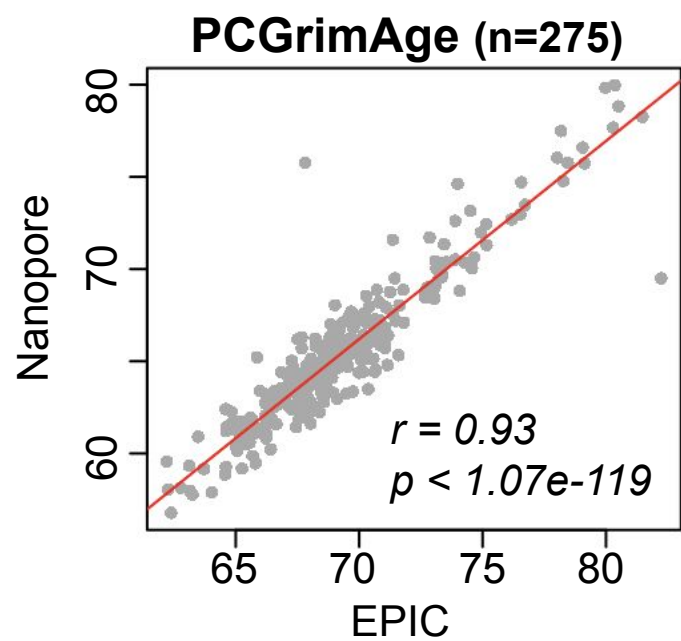
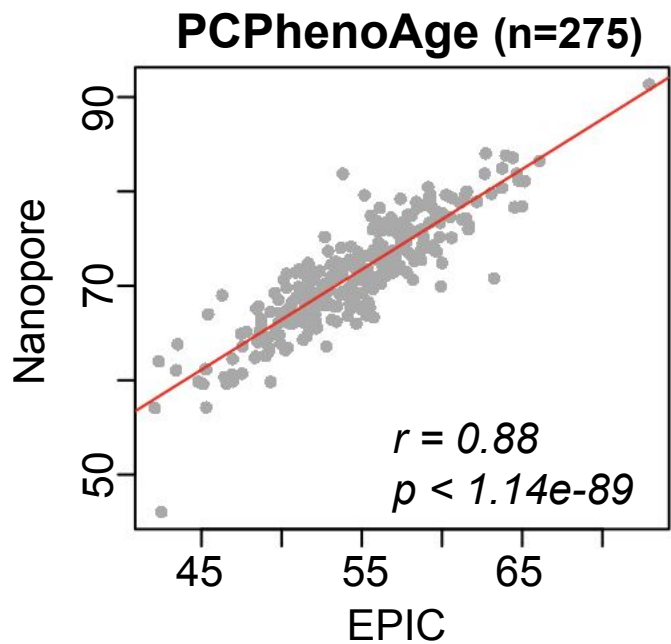
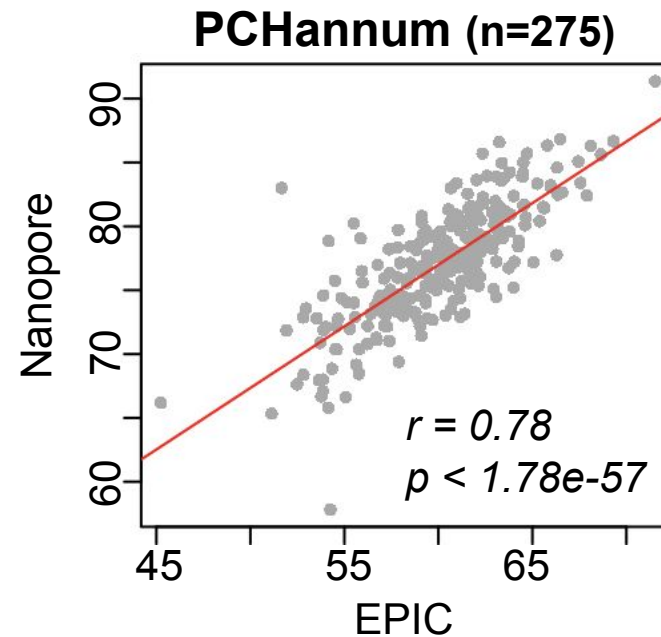
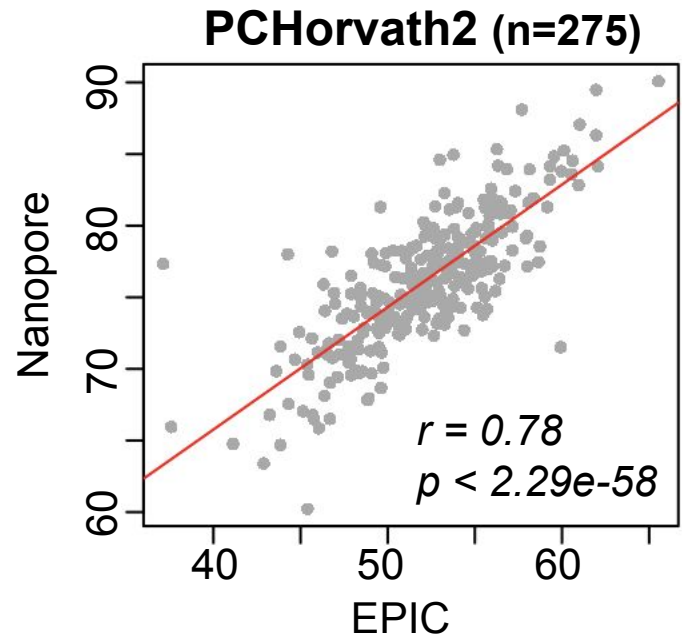
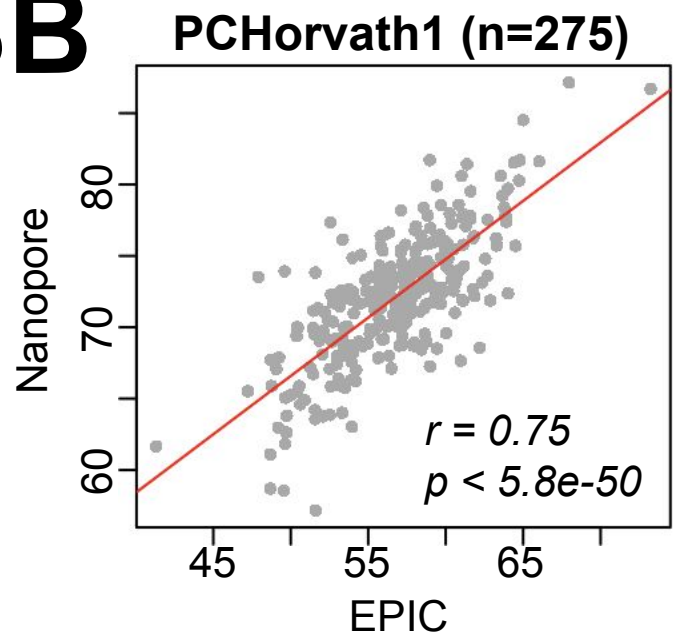
PCPhenoAge (n=50)



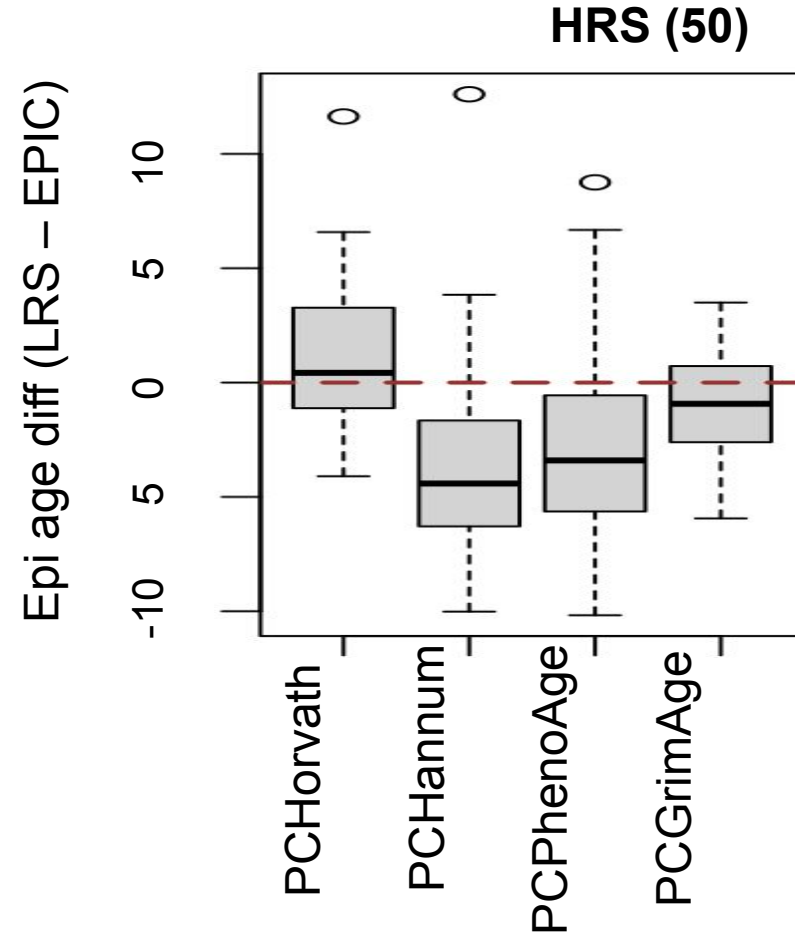
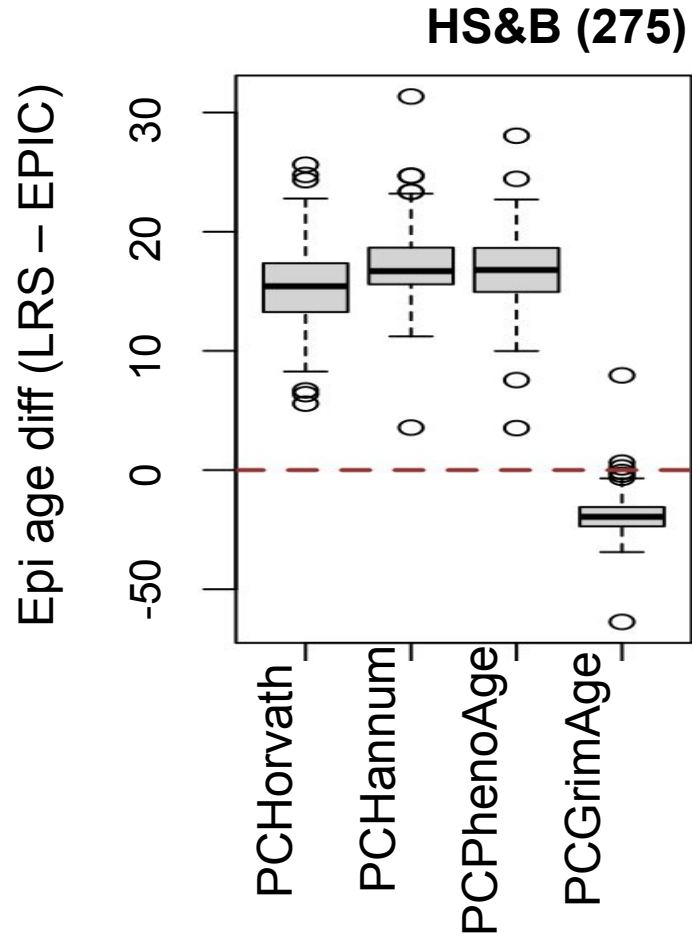
PCGrimAge (n=50)



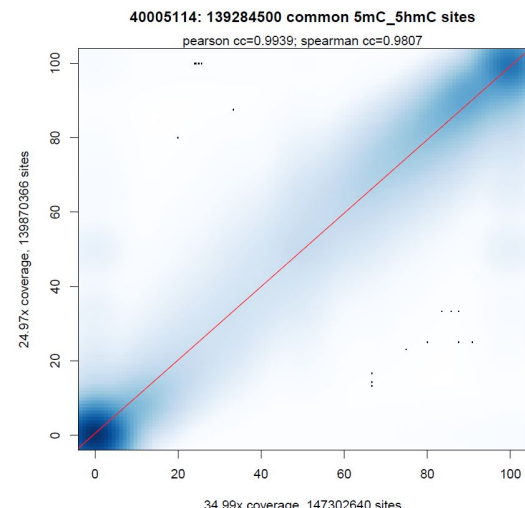
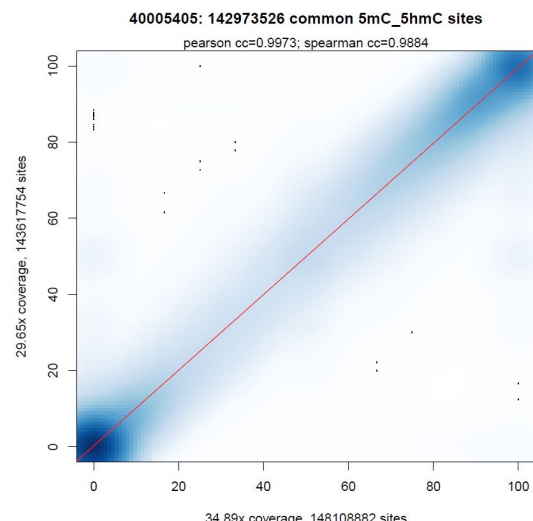
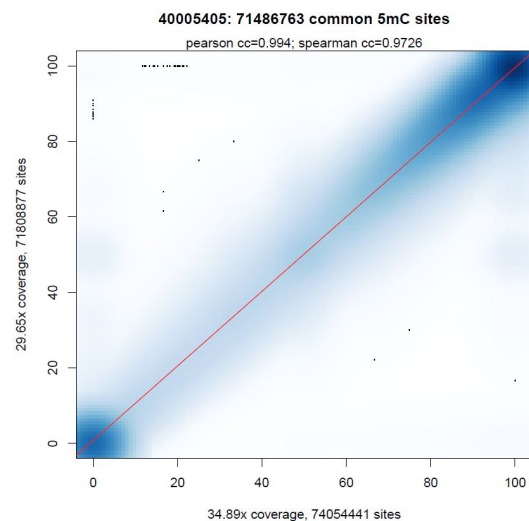
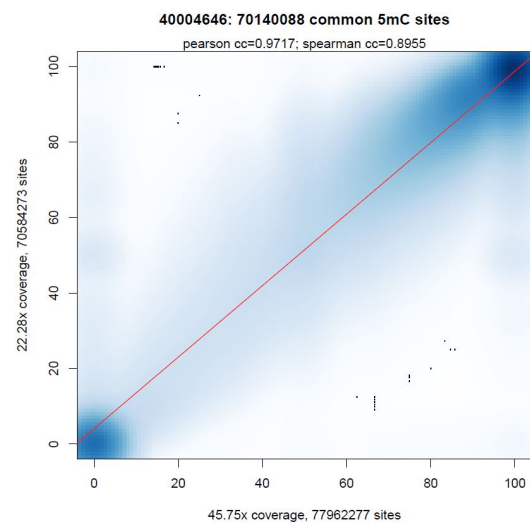
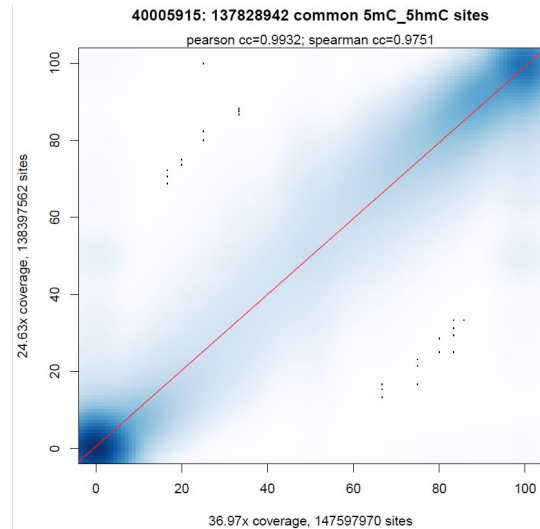
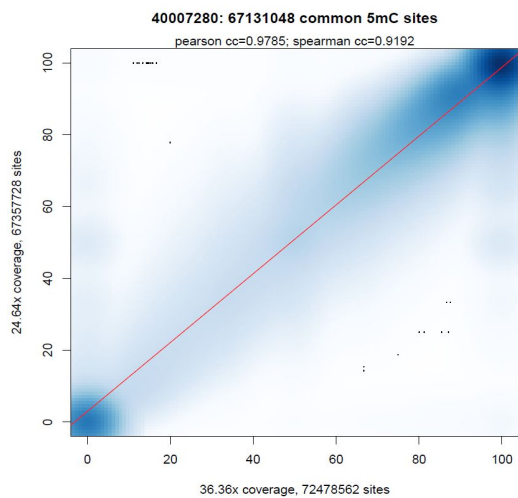
HSB



ABSOLUTE DIFFERENCE IN EPIGENETIC AGE: LRS VS. EPIC



REPRODUCIBILITY: DUPLICATE ANALYSIS



- Six samples sequenced in duplicate
- High correlation for 60M-70M CpG sites ($r > 0.97$)

RESULTS AND DISCUSSION

- Overall, a good correlation at the level of individual CpG sites between LRS and EPIC arrays
 - Large differences in Epigenetic Age
 - Magnitude of difference appears to be cohort and clock specific
- LRS shows good reproducibility of DNA methylation results
- Opportunities to obtain consistent coverage of DNA methylation across the genome
 - New clocks
 - New opportunities to integrate DNA methylation with genomic variation to get cell type specific genomic alterations

ACKNOWLEDGEMENTS

- **Health and Retirement Study**: National Institute on Aging (NIA U01AG009740) and the Social Security Administration.
- **High School and Beyond**: National Institute on Aging (U01AG058719) and the Alzheimer's Association (SG-20717567).
- Department of Laboratory Medicine and Pathology, University of Minnesota