# INTERPRETATION AND VALIDATION OF SINGLE AND MIXED SPERM CELL ELECTROPHEROGRAMS WITH EESCIt™

LFTDI

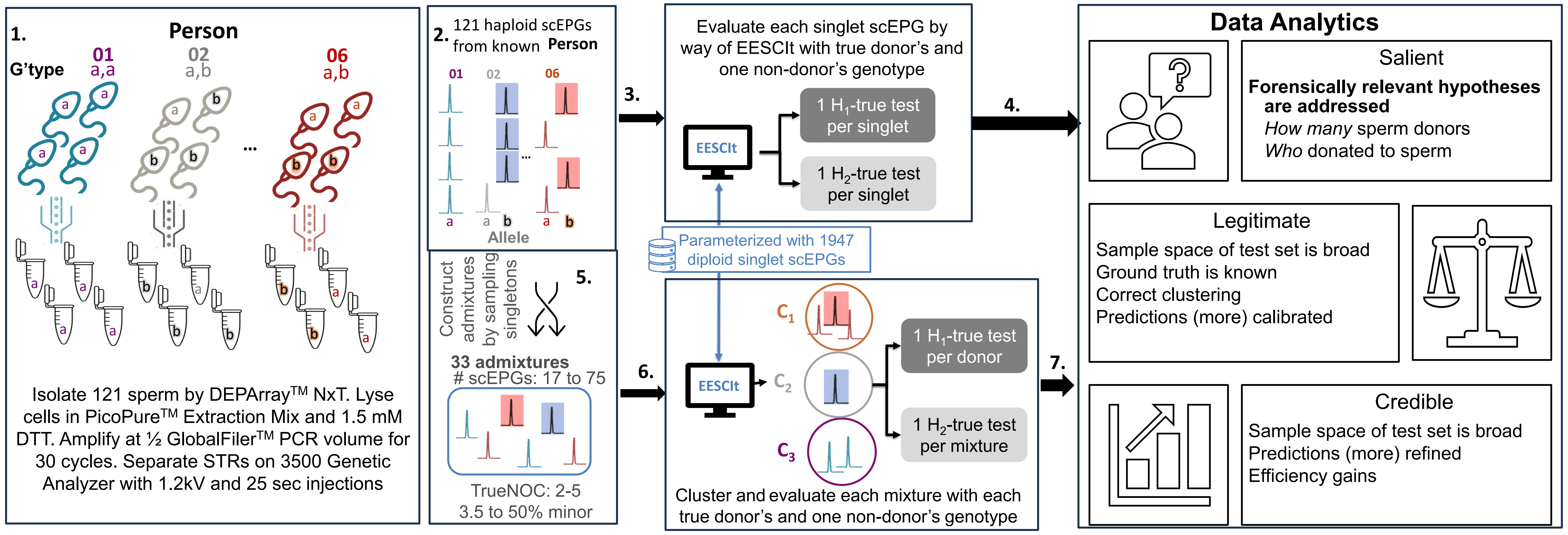
**Anu J Khandelwal M.S.; Qhawe A. Bhembe M.S.; Desmond S. Lun Ph.D.; Catherine M. Grgicak Ph.D.**

RUTGERS
UNIVERSITY | CAMDEN

## HIGHLIGHTS

- **In this study, we evaluated the ability of EESCIt™ — a single-cell probabilistic assessor — to evaluate electropherograms of sperm cells given (non-) contribution and a clustering.**
- **We evaluate the question regarding *who* donated to an admixture of sperm scEPGs**
- **The test set included the following, with a key feature being that ground truth donor(s) were known :**
  - **121 singlets (i.e., individual haploid sperm cells) from six donors**
  - **33 admixtures made from the 121 single-cell electropherograms (scEPGs), with varying numbers of contributors (2-5), varying numbers of scEPGs (17-75), and multifarious minor contributor ratios (3.5 to 50 %)**
- **Evaluations exhibited unequivocal discrimination of the $H_1$- and $H_2$-true test classes despite haploid scEPGs containing only half of the information compared to diploid scEPGs.**

## ABSTRACT



1\. Isolate 121 sperm by DEPArray™ NxT. Lyse cells in PicoPure™ Extraction Mix and 1.5 mM DTT. Amplify at ½ GlobalFiler™ PCR volume for 30 cycles. Separate STRs on 3500 Genetic Analyzer with 1.2kV and 25 sec injections

2\. 121 haploid scEPGs from known **Person**

3\. Evaluate each singlet scEPG by way of EESCIt with true donor's and one non-donor's genotype

- 1 $H_1$-true test per singlet
- 1 $H_2$-true test per singlet

Parameterized with 1947 diploid singlet scEPGs

5\. Construct admixtures by sampling singletons

**33 admixtures**
# scEPGs: 17 to 75
TrueNOC: 2-5
3.5 to 50% minor

6\. Cluster and evaluate each mixture with each true donor's and one non-donor's genotype

- 1 $H_1$-true test per donor
- 1 $H_2$-true test per mixture

4\. / 7. **Data Analytics**

Salient
**Forensically relevant hypotheses are addressed**
*How many* sperm donors
*Who* donated to sperm

Legitimate
Sample space of test set is broad
Ground truth is known
Correct clustering
Predictions (more) calibrated

Credible
Sample space of test set is broad
Predictions (more) refined
Efficiency gains

## SINGLETS; s=121

### SAMPLE DATASET

Summary of the amount of genetic information contained in each singlet scEPG

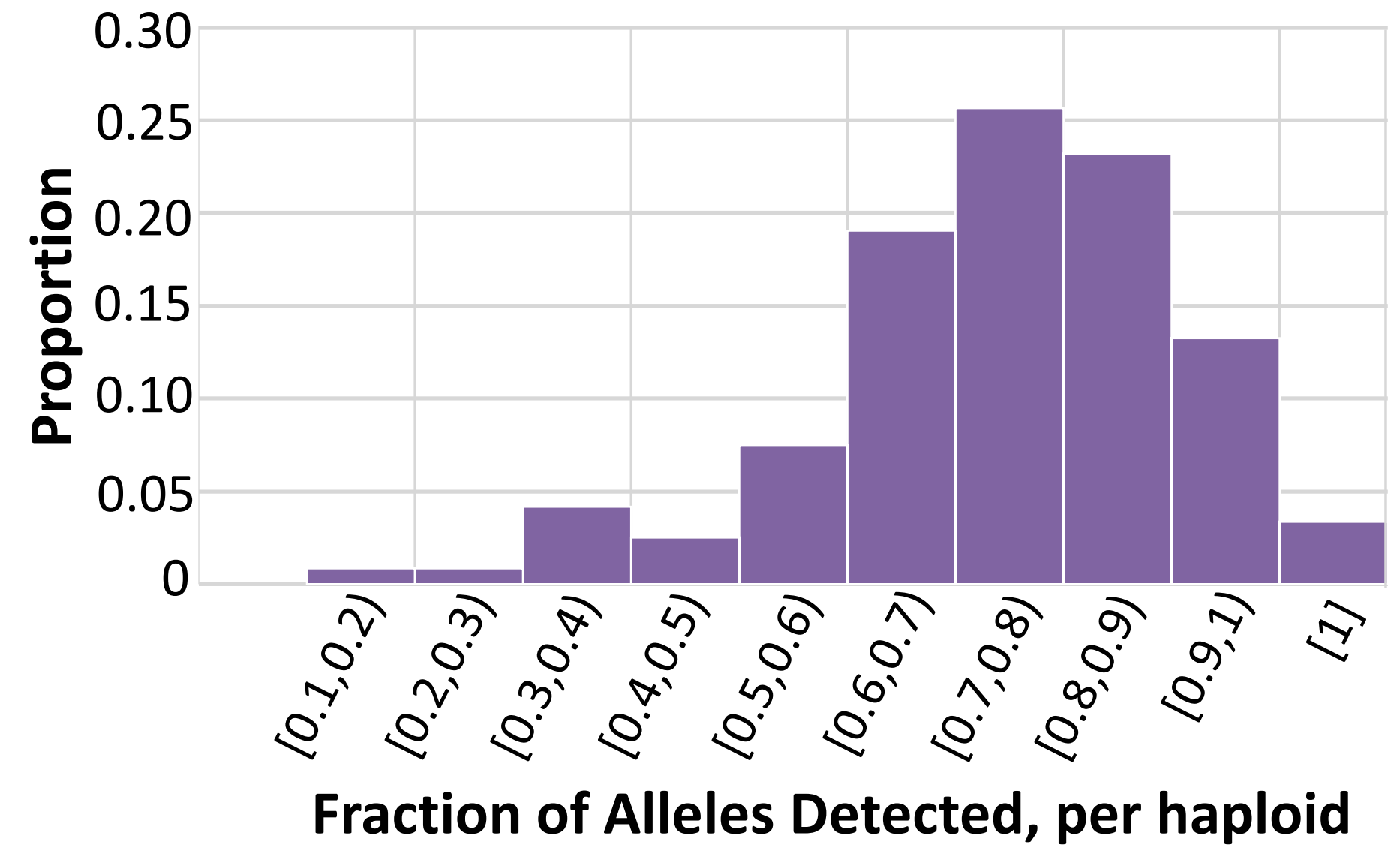
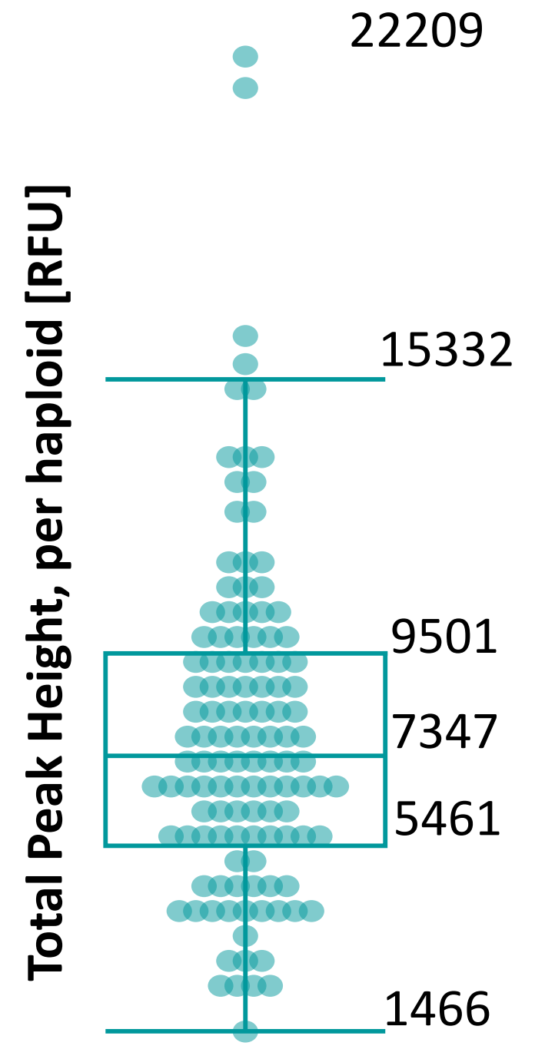


**Figure A:** Histogram reporting the allele detection rate, per scEPG. The mode of [0.7,0.8) shows that most scEPGs carry signal in 70-80% of their allele positions.



Maximum:22209
3rd Quartile:9501
Median:7347
1st Quartile: 5461
Minimum:1466

**Figure B:** Boxplot showing the total intensity carried by each scEPGs in RFU. Y-axis lists for Minimum, 2nd, 3rd, 4th, maximum values.

### EVALUATION GIVEN THE CLUSTERING

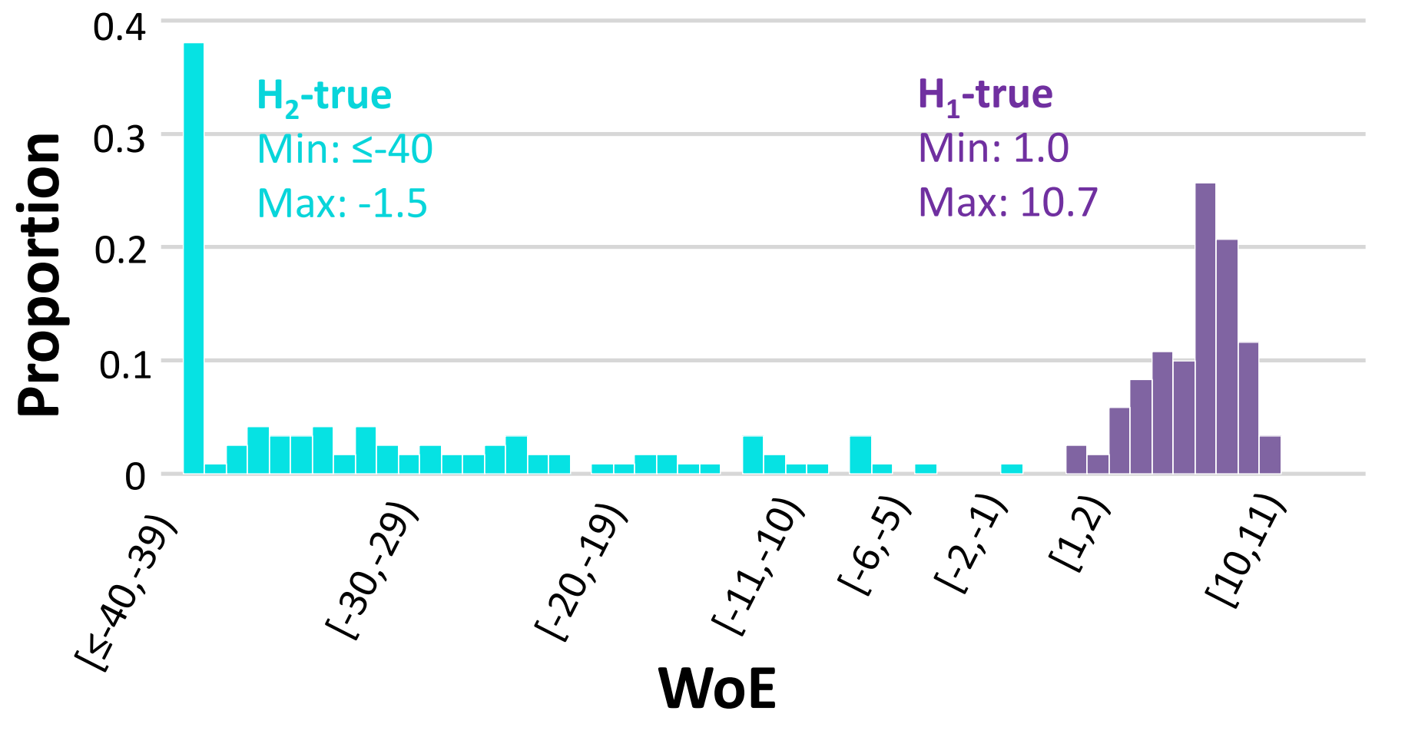


**Figure C**: Histogram reporting WoEs for $H_1$-true and $H_2$-true test class for singlets, with no overlap between the two test classes, $H_1$-true and $H_2$-true.

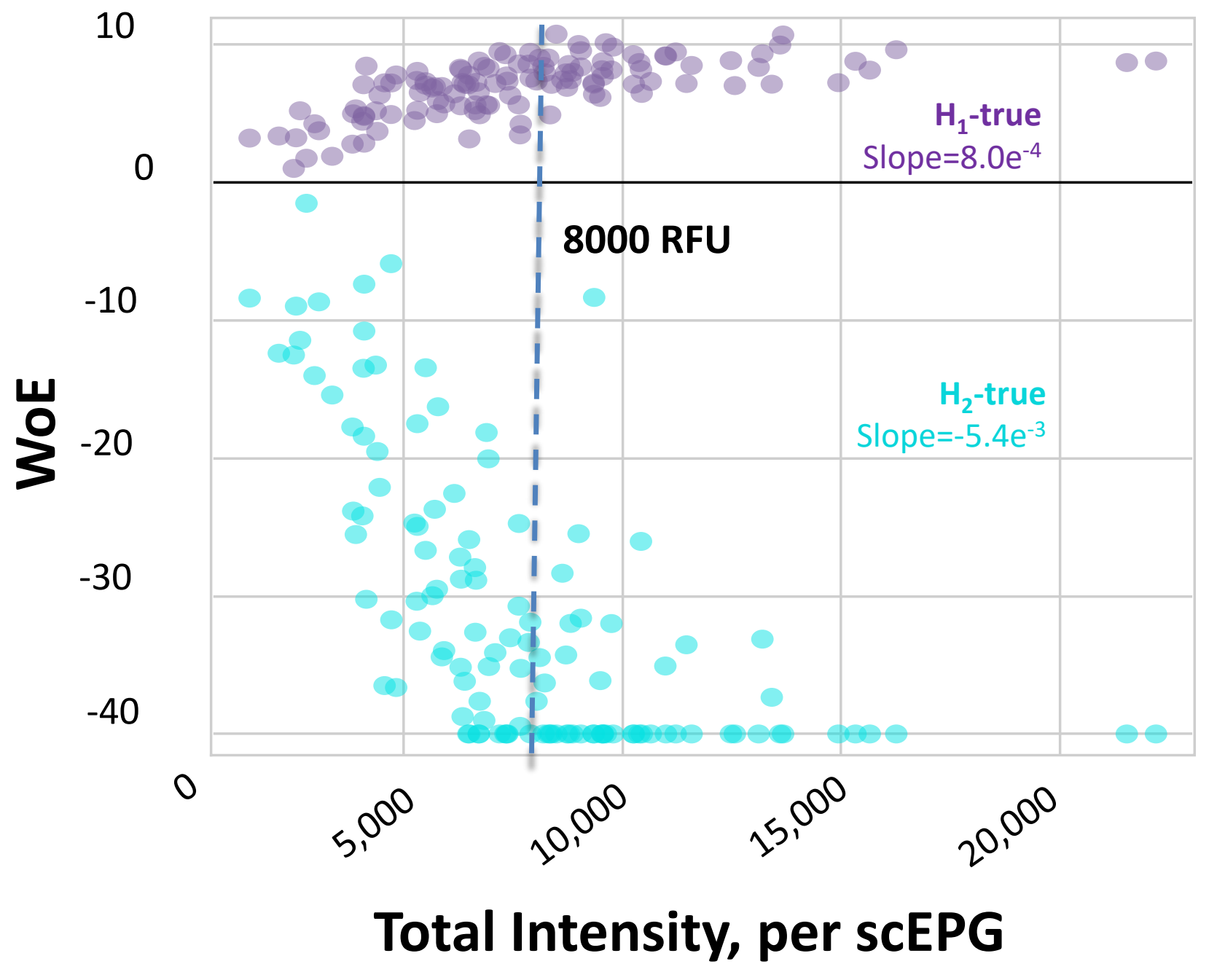


**Figure D:** Scatter plot showing the relationship between $\log_{10}LR$ and the total intensity [RFU] carried by each scEPG. The slope shows that we retain 1 WoE for every ca. 1200 RFU carried by the singlet when $H_1$-true, and plateaus at 8000 RFU.

## MIXTURES; s=33

### SAMPLE DATASET

Summary of admixture construction and scEPG qualities in the admixtures

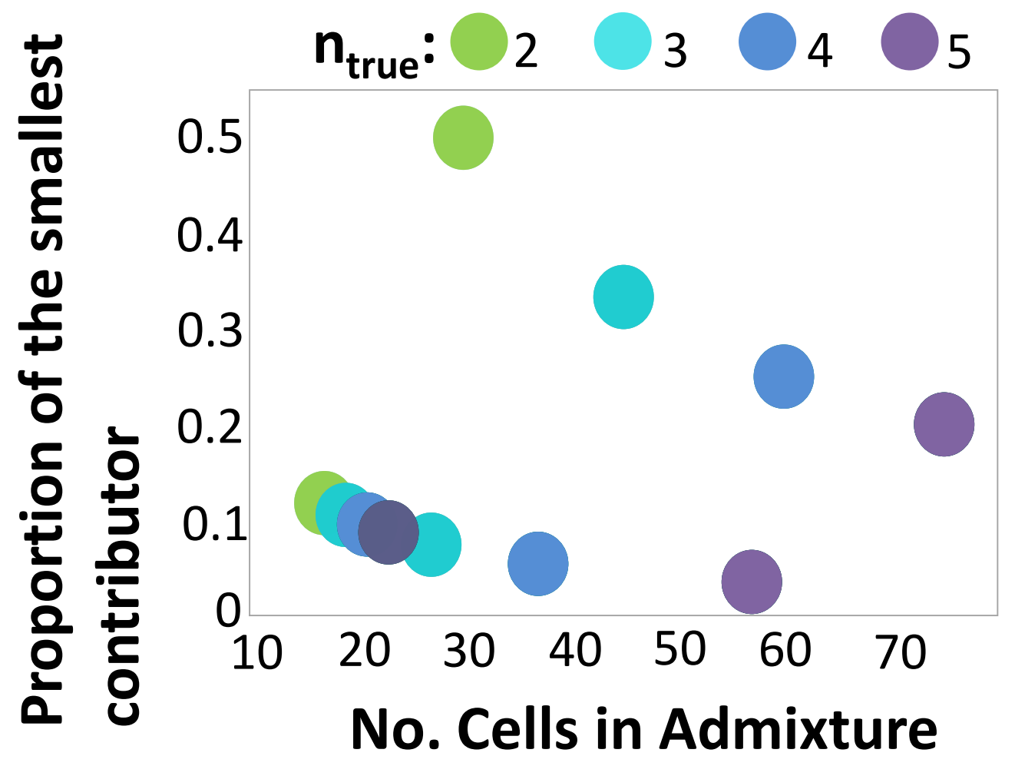


**Figure E:** Dot plot reporting broadness of admixture factor space, consisting of 2- to 5- donors, 17-75 scEPGs, and minor contributors with as low as 3.5% (or 2-cells).

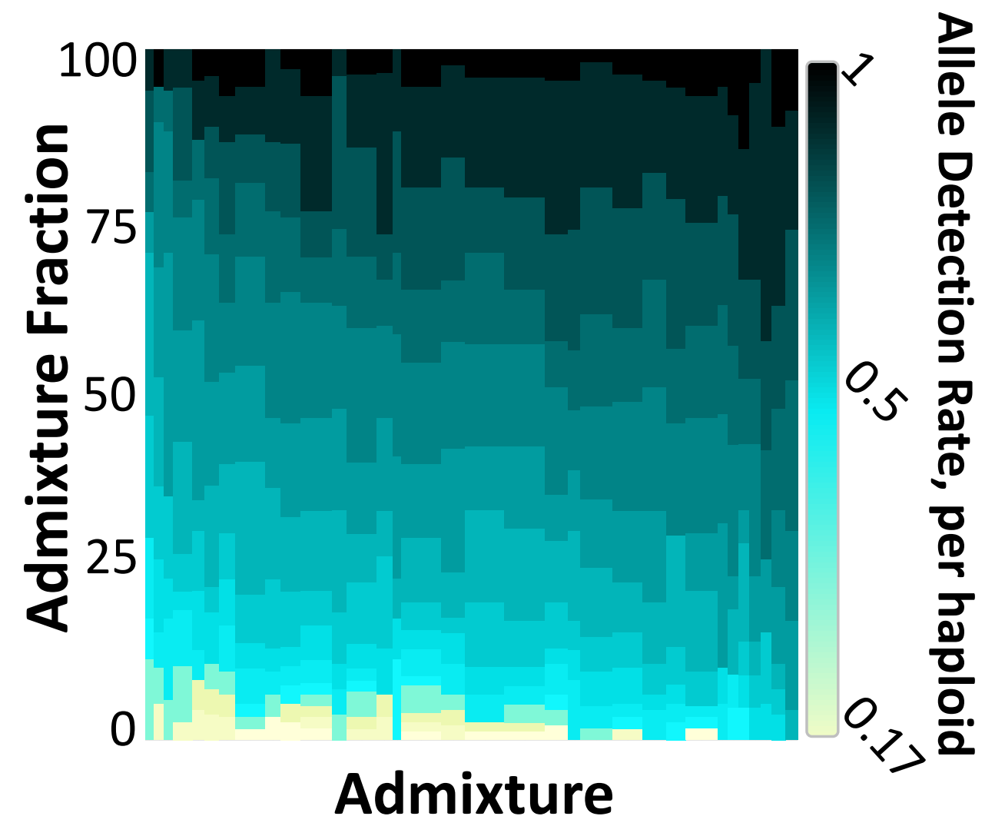


**Figure F**: Stacked plot depicting the fraction of scEPGs within an admixture whose allele detection was 17 to 100%.

### EVALUATION GIVEN THE CLUSTERING

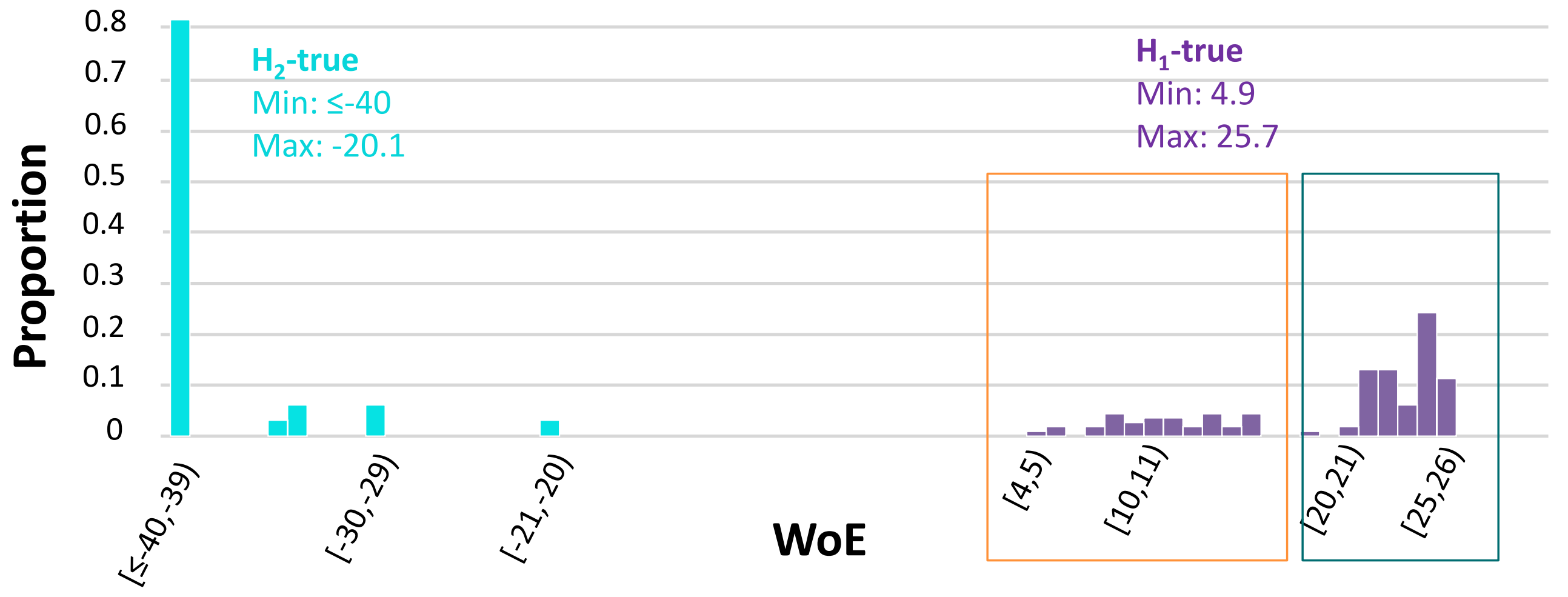


**Figure G:** Histogram reporting the WoE for the $H_1$-true and $H_2$-true test for admixtures.

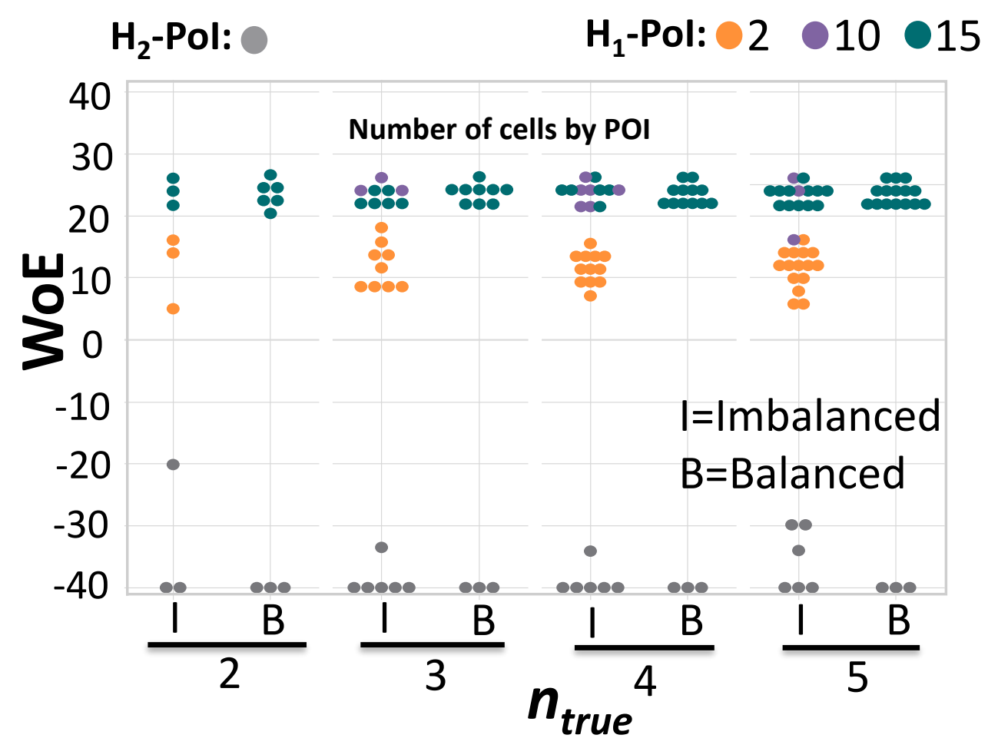


**Figure H:** Scatter plot of the admixture's WoE for each PoI, separated by $n_{true}$, and I = at least one donor giving less than 20% of scEPGs in the admixture. (●) = $H_2$-POI, and (colored) = $H_1$-POI; i.e. (number of scEPGs donated by each POI) in the admixture.

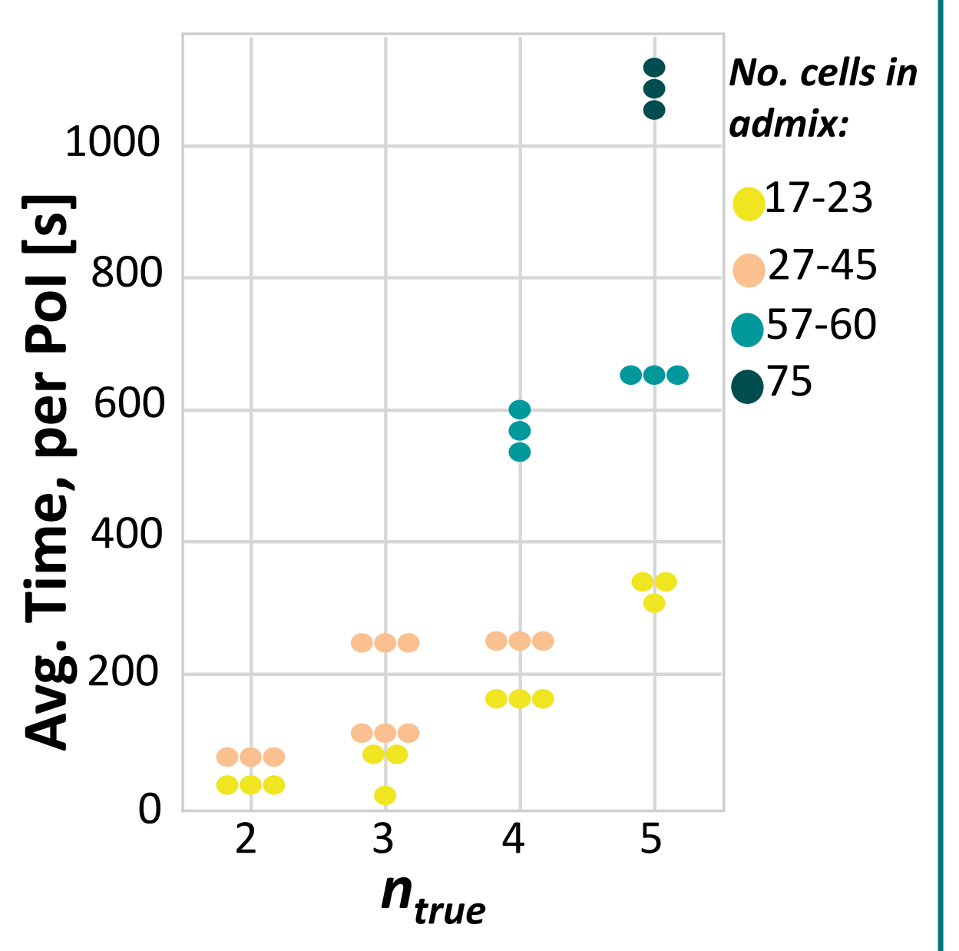


**Figure I:** Average computational time, in seconds, per PoI comparison, separated by total number of cells in the admixture and $n_{true}$.

## CONCLUSION

- EESCIt™ 's predictions are calibrated, and no WoE pointed in the wrong direction – i.e., in no instance did haploids support the counter hypothesis.
- EESCIt™ is still relatively assertive, even with individuated haploid scEPGs that carry half of the donor's genetic information.

## NEXT STEPS

We will confirm that EESCIt™ 's WoE calibration and discrimination hold at source-level – i.e., at the level of cell-type – by evaluating across a series of admixtures containing sperm blood, and epithelial cells.

## ACKNOWLEDGEMENT

This work was partially supported by NIJ2020-R2-CX-0032 and NIJ2018-DU-BX-K0185 awarded by the National Institute of Justice, Office of Justice Programs. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not reflect those of the Department of Justice.