Green Mountain
July 22, 2024
Burlington, VT

LFTDI

LABORATORY FOR FORENSIC TECHNOLOGY DEVELOPMENT & INTEGRATION

THE SALIENCE, LEGITIMACY AND CREDIBILITY OF SINGLE CELL GENETICS TO FORENSICS

Catherine M. Grgicak

RUTGERS UNIVERSITY | CAMDEN

BOSTON UNIVERSITY

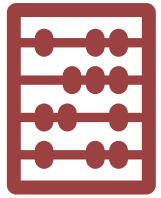
Maynooth University
National University of Ireland Maynooth

Netherlands Forensic Institute
Ministry of Justice and Security

R CAMDEN

Northeastern University

In this presentation we review the logic supporting the use of single cell (sc)data in forensics and the findings buttressing that position



Salience. Refers to the relevance of information affecting a stakeholder or specific domain

Legitimate. Refers to whether an actor perceives the process/technology as unbiased and meeting standards of fairness

Credible. Refers to whether an actor perceives information as meeting standards of scientific plausibility and exceeding current technical adequacy

David Cash, William C. Clark, Frank Alcock, Nancy M. Dickson, Noelle Eckley, Jill Jäger. Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making. KSG Working Papers Series, 2003. https://dash.harvard.edu/handle/1/32067415

Single cell treatments are defined by extracting R/DNA one cell at a time and using direct amplification

Two features common to all single-cell experiments:

that intact cells or nuclei are isolated before the cell is lysed; and

that the extraction and amplification (or library preparation) occurs in the same vessel to which the cell was added

Salience. Relevance of information affecting a stakeholder or domain

Can we sample enough cells from a minor contributor?

Can scDNA be used to fulfill investigative aims in the absence of a suspect?

Can scDNA be used to fulfill evaluative aims when there is a named suspect, and be used with compound hypotheses?

Not all DNA is found in cells. What about cell-free (cf)DNA?

Information limit is defined by the number of cells isolated, rather than detector saturation

Since we sample without replacement, we can determine the probability that we isolate at least one cell from a total of t cells, where t_d is the number of cells from d, and when m cells are isolated by,

$$\Pr(r \geq 1) = 1 - \frac{\binom{t-t_d}{m}}{\binom{t}{m}}$$

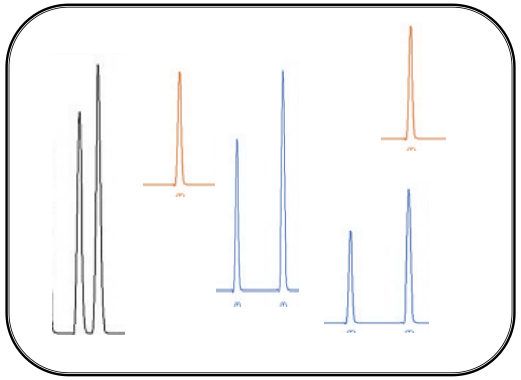
e.g.1, t=100; t_d=5 (1 in 20 mixture); m=40 cells, this evaluates to 92%. By isolating m=80 cells the probability increases to 99.8%

Supports the position to accelerate research into high throughput single-cell forensics

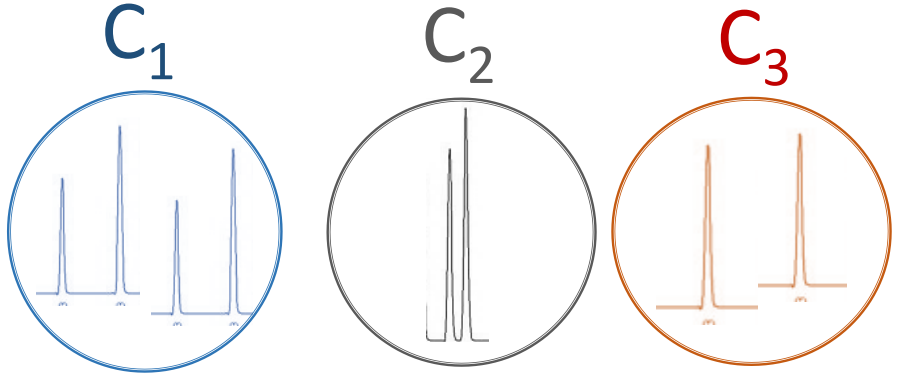
With faithful suspect-agnostic clustering EESCIt™ addresses investigative & evaluative aims

- 5 scEPGs
- One locus
- Colors=different donors

ADMIXTURE, *A*



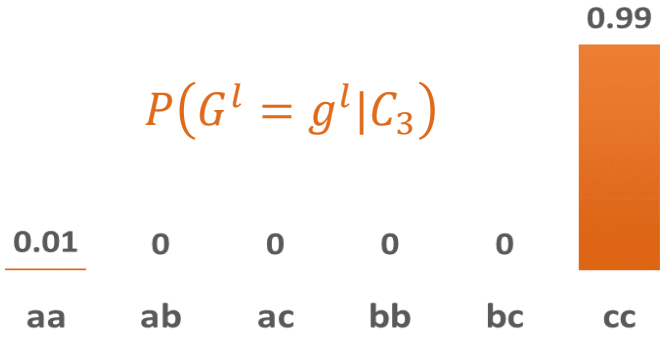
Suspect-agnostic clustering



INVESTIGATIVE (NO SUSPECT)

$$P\left(G^l = g^l \middle| C^l\right) = \frac{\{\prod_{i=1}^{v} P\left(E_i^l \middle| G^l = g^l\right)\} P\left(G^l = g^l\right)}{\sum_{g^l}\{\prod_{i=1}^{v} P\left(E_i^l \middle| G^l = g^l\right)\} P\left(G^l = g^l\right)}$$

$P\left(G^l = g^l | C_3\right)$



e.g. At *P*>0.998, credible set for this locus in this cluster is {cc and aa}

EVALUATIVE (SUSPECT)

Sub-sub-source evaluation, i.e., cluster evaluation

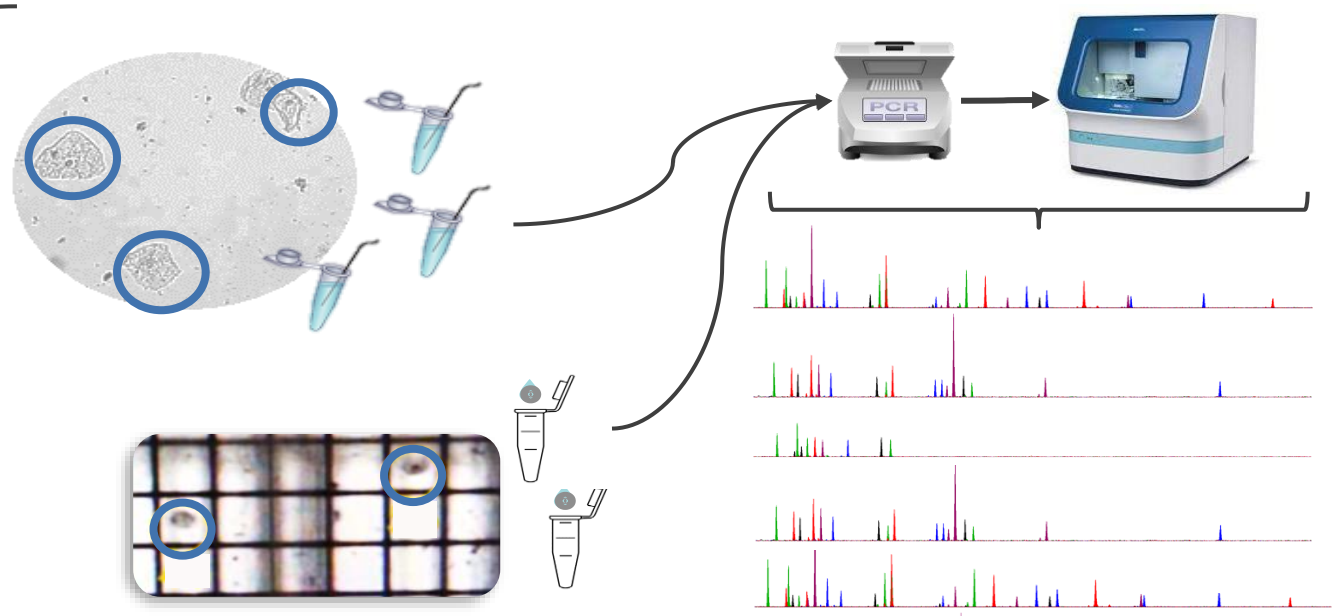
$$\mathrm{LR(C, s)} = \frac{\prod_{l=1}^{L}\prod_{i=1}^{v} P\left(E_i^l \middle| G^l = s^l\right)}{\prod_{l=1}^{L}\sum_{g^l}\prod_{i=1}^{v} P\left(E_i^l \middle| G^l = g^l\right) P\left(G^l = g^l\right)}$$

Sub-source evaluation, i.e., for the entire admixture, *A*, of cells continues by averaging the LR across clusters

$$\mathrm{LR(A, s)} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{LR}(C_i, s)$$

e.g. For suspect, *s*, $\mathrm{LR}(A, s) = \frac{1}{3}[10^{-40} + 10^{-40} + 10^{30}] = 10^{29}$

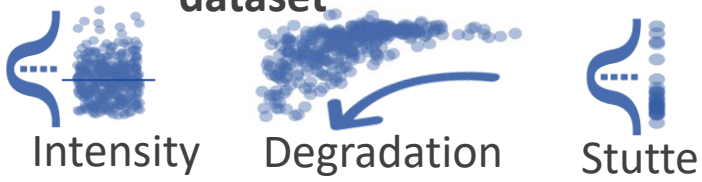
630 test mixtures probabilistically clustered and evaluated with EESCIt™



Isolate epithelial or blood cells from single source samples by way of manual or fluidic treatments. Follow with direct-to-PCR extraction, STR amplification, electrophoresis and fragment analysis. Each scEPG is, therefore, of known genotype allowing performance evaluations

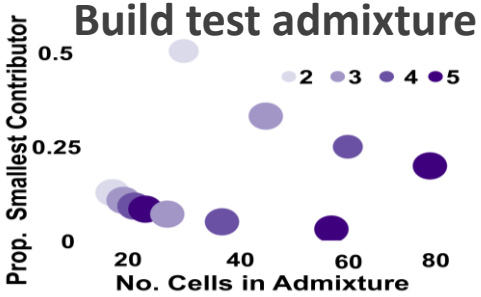
Fit EESCIt™ models to scEPG calibration dataset

Calibration **N=1420**



Build test admixture dataset

Test **N=643**



630 mixtures
2 to 5 donors
17 to 75 cells
3.5 to 50% minor

Performance assessment

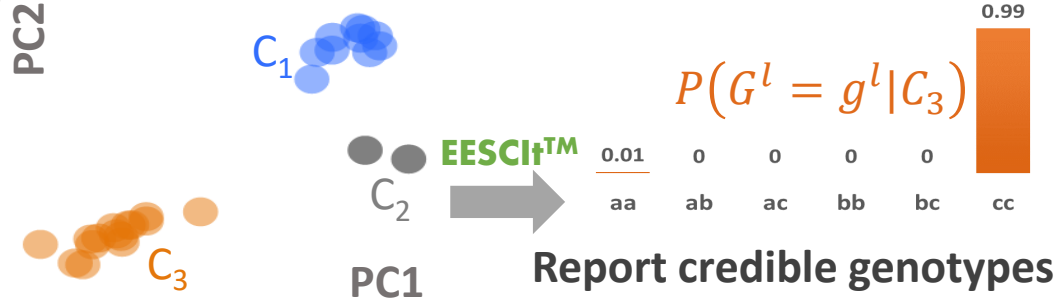
Sensitivity

Donor genotype
-Included
-Excluded

Robustness

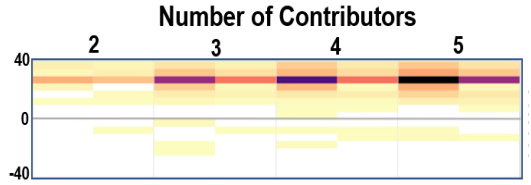
Specificity

Speed



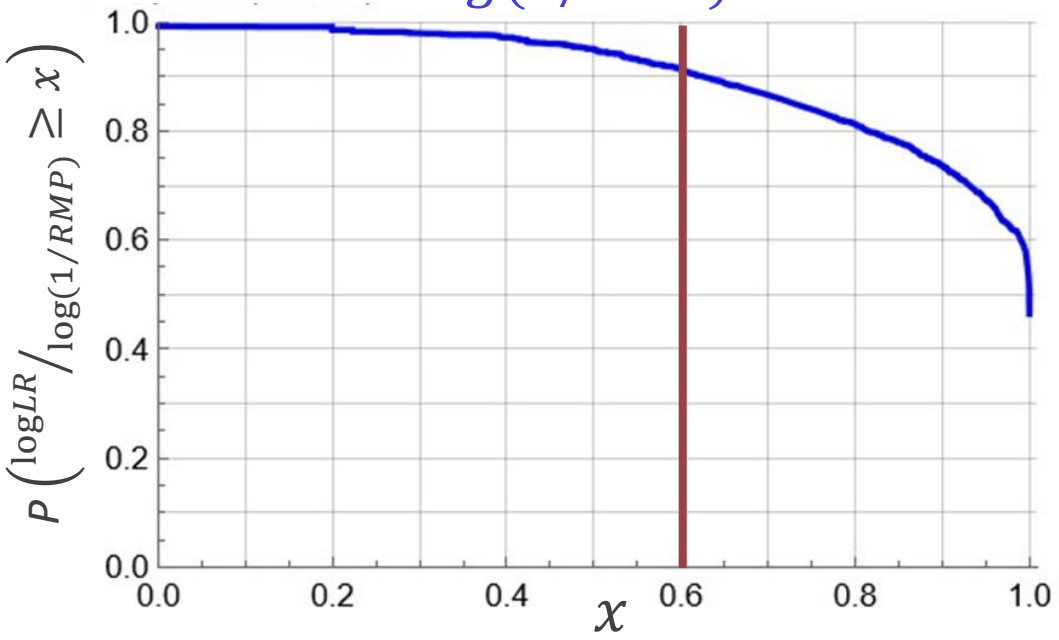
Cluster scEPGs in each test sample by similarity using mclust, <u>without reference to a suspect</u>

Report credible genotypes or ***sub*-source LRs**



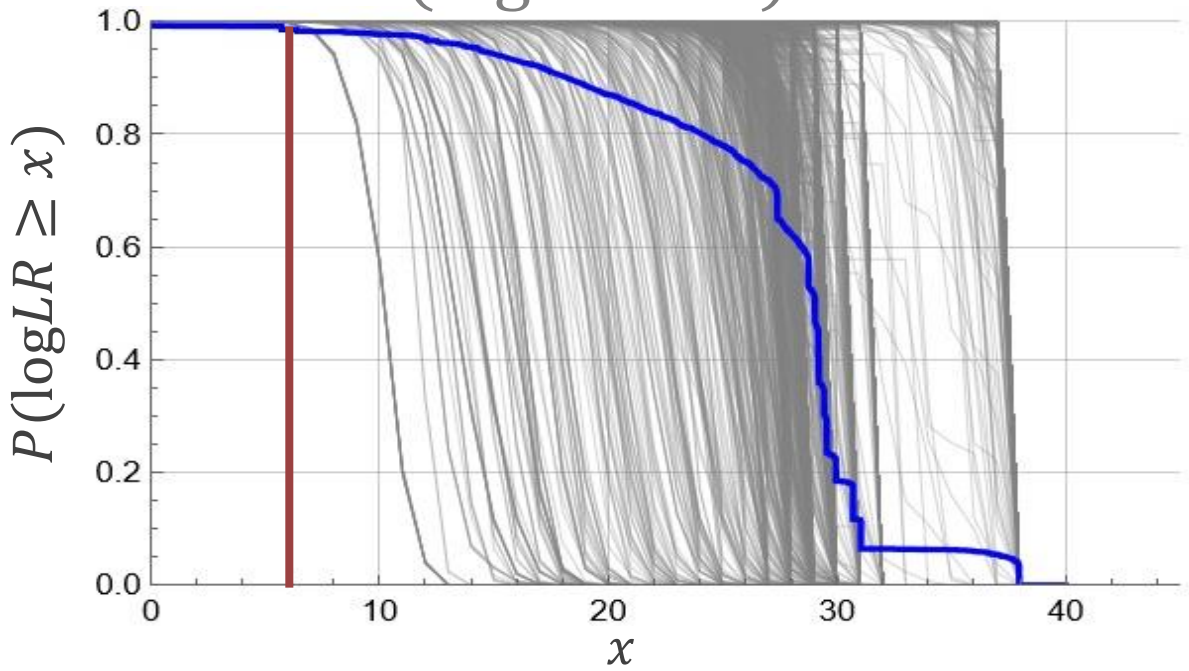
Investigative single cell genetics: All components met stringent searchability criteria

Proportion of 2,519 clusters for which $\frac{logLR}{log(1/RMP)} \geq x$



91% of the clusters give at least 60% of the maximal amount of information that could have been returned, which corresponds to LR ca. 10^{18}

For each cluster 10,000 LRs were sampled from same source distribution to get $P(\log LR \geq x)$



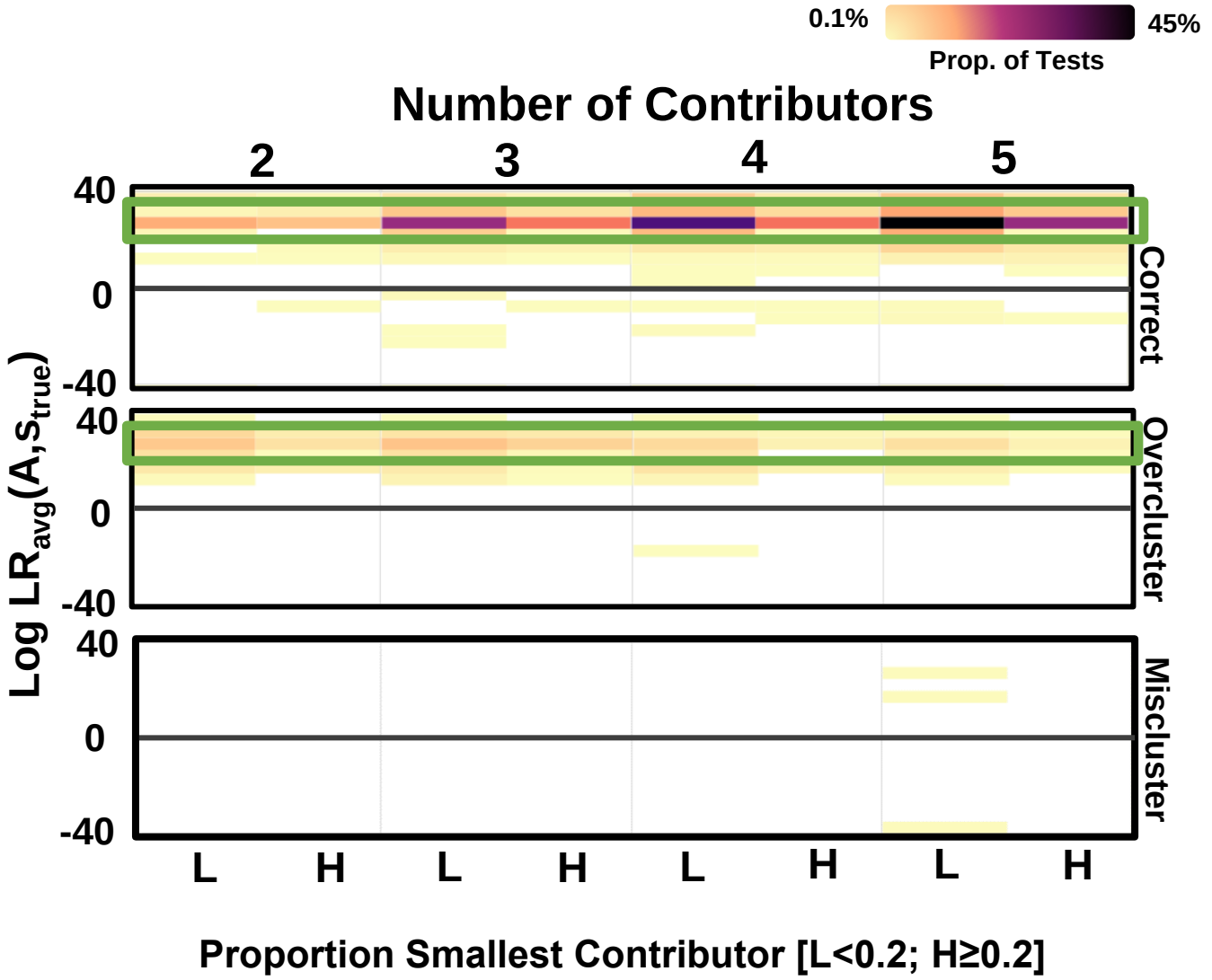
$P(\log LR > 6) \cong 1$ for all clusters, meaning **every cluster was of a searchable state**

Sub-source (suspect) evaluations are well resolved across mixture complexities

Out of 2,310 suspect-mixture comparisons all but 21 gave LR>1

High density of log_{avg}LRs at [25-30) across TrueNOC shows robustness across all complexities

WoE are **not** dependent on the mixture's qualities, making it the first **fully robust forensic data type**



A unifying framework jointly evaluating cf- and sc-DNA data in EESCIt™ – an example

3-person mixture, with the makeup as follows:

Donor ID	scEPGs	cfEPG
1	15	0
2	5	1
3	0	4

WoE results:

PoI	Combined	Single-cell	Extracellular
1 (in scEPGs only)	**37**	38	-25
2 (in both)	**36**	31	9
3 (in cfEPG only)	**19**	-40	20
4 (in neither – H_d is true)	**-40**	-40	-17

Salience. Relevance of information affecting a stakeholder or domain

Can we sample enough cells from a minor contributor? Yes.

Can scDNA be used to fulfill investigative aims in the absence of a suspect? Yes.

Can scDNA be used to fulfill evaluative aims when there is a named suspect, and be used with compound hypotheses? Yes.

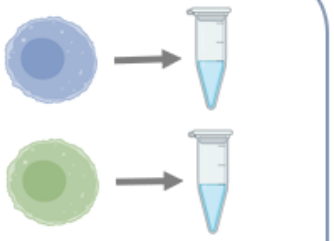
Not all DNA is found in cells. What about cell-free (cf)DNA? Use it.

Legitimate. Is scDNA, and its interpretation (perceived as) fair?

Are scWoE calibrated? Do they over- or under-state the evidence?

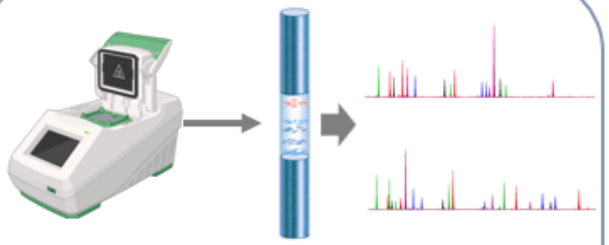
Are scWoE calibrated across different model architectures? Are they impervious to different model architectures?

996 singlet scEPGs evaluated using 3 model architectures



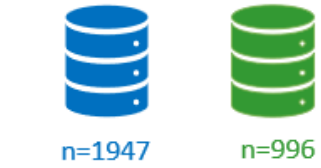
Isolate 2,943 epithelial or blood cells from single source samples by way of Bulldog Bio™ picopipetting or DEPArray™ NxT

STAGE 1



Extract with PicoPure™, amplify at ½ reaction volume with GlobalFiler™ for 30 cycles, inject for 25s on 3500 Genetic Analyzer, and analyze each scEPG with OSIRIS v 2.10.3

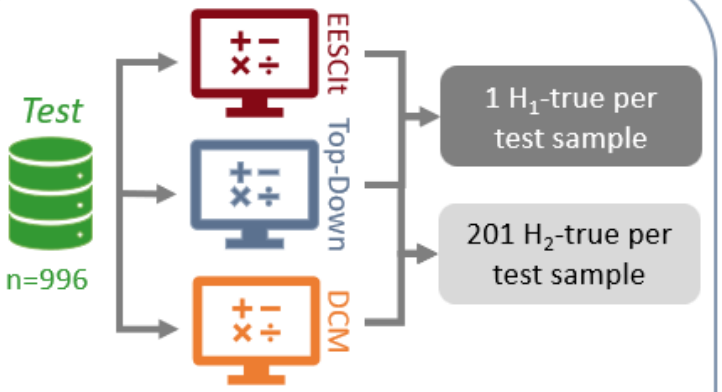
STAGE 2



Split scEPGs into a set to parameterize EESCIt and mechanistic models and a test set

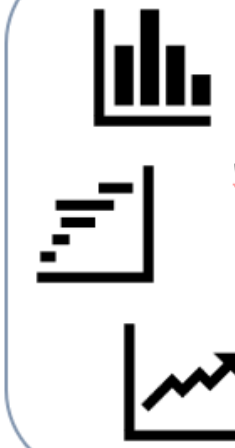
STAGE 3

STAGE 4



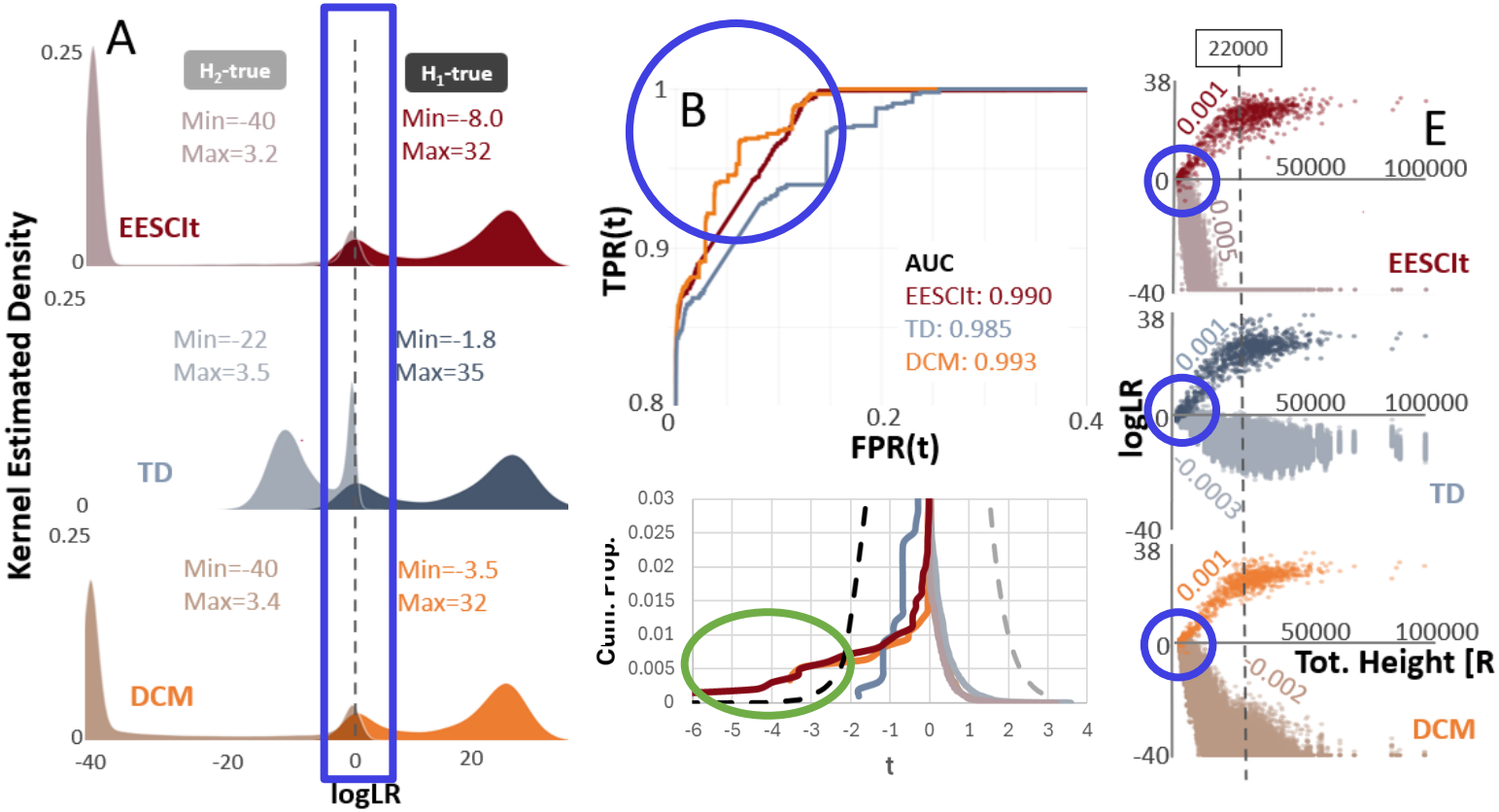
Perform 1 H_1-true and 201 H_2-true tests per scEPG, for a total of 201,192 outcomes per model

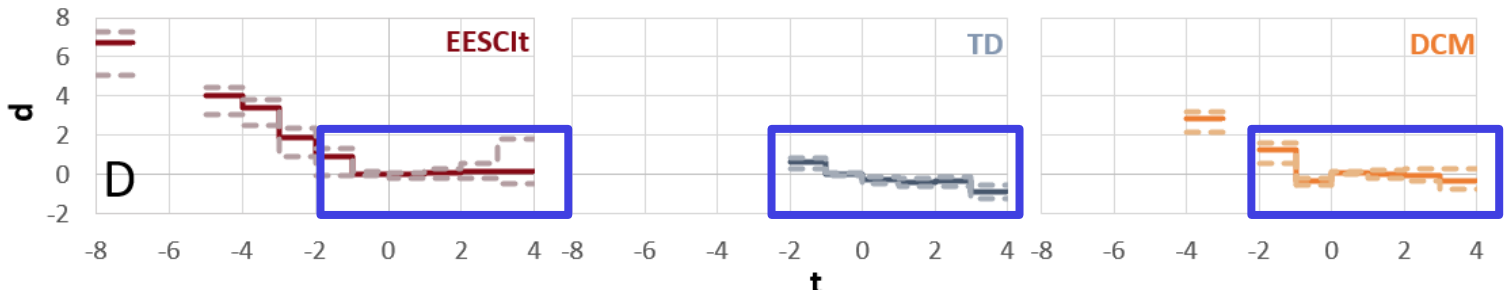
STAGE 5



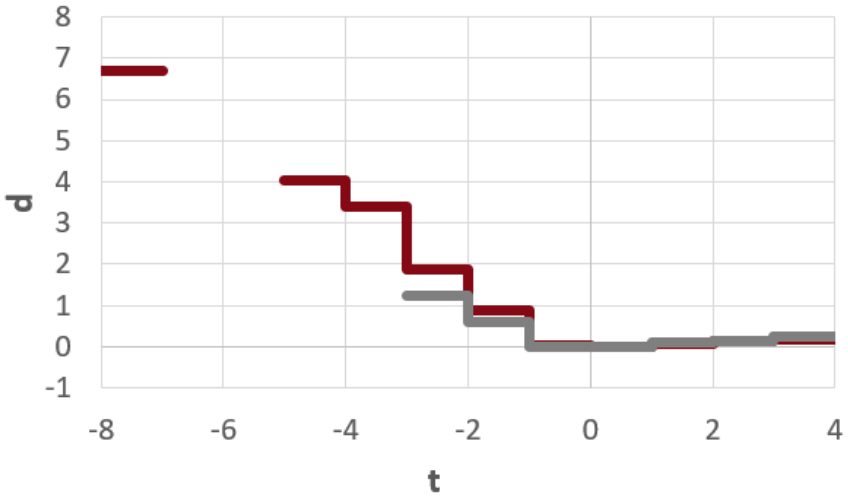
Data Analysis:
WoE distributions, calibration evaluation, WoE comparisons, and ROC/Tippet analysis for each model

scWOE are calibrated after acknowledging rare (hard to measure), though impactful, events





Applying a **prior probability the data are of another distribution re-calibrates scWoE** and is a way to address rare, impactful, hard to estimate probabilities



Legitimate. Is scDNA, and its interpretation, fair?

Are scWoE calibrated? Yes.

Are scWoE similarly calibrated and equivalent across substantively different model architectures? Yes.

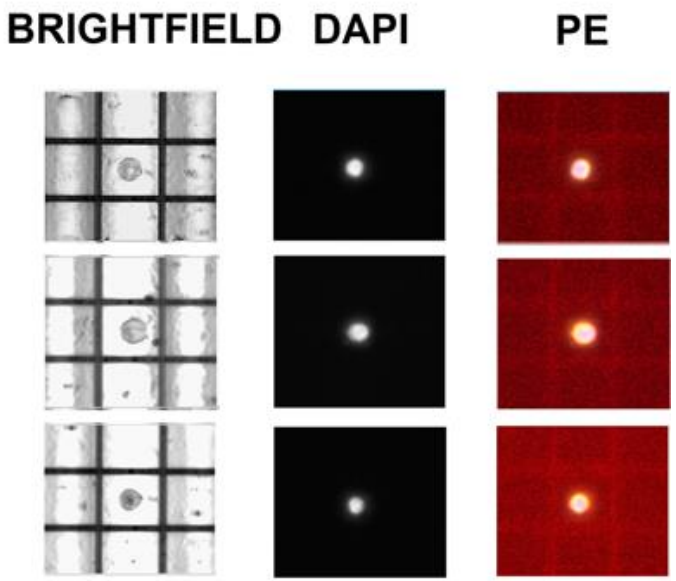
Credible. Is scDNA technically adequate and can it 'outperform' mixedDNA?

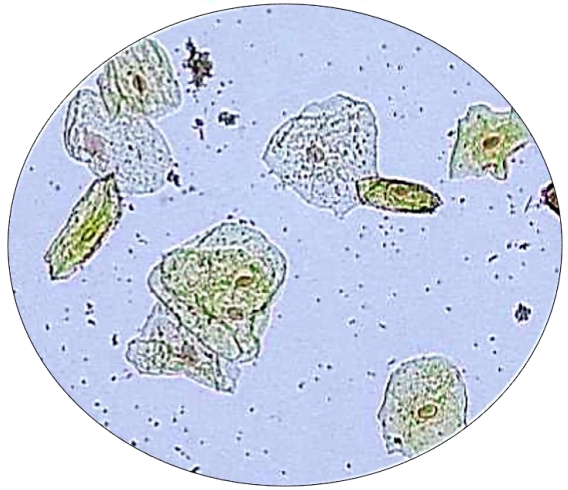
Can scDNA provide information beyond sub-sub-source and sub-source evaluations?

Where do the limits of scDNA interpretation lie?

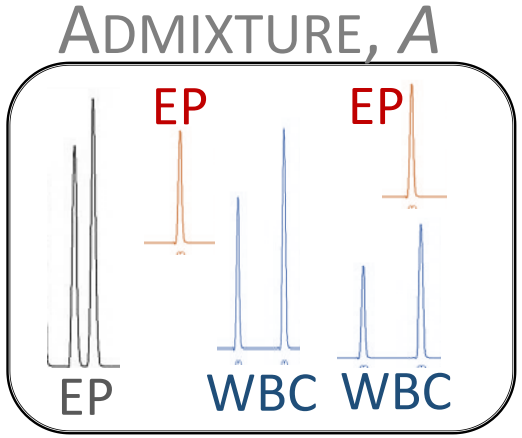
What is the computational burden?

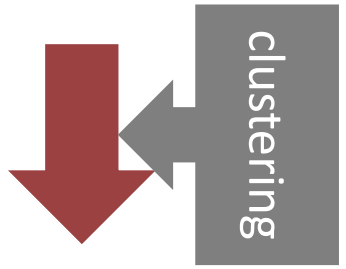
Coupled with imaging, scDNA provides source information – i.e., cell-type

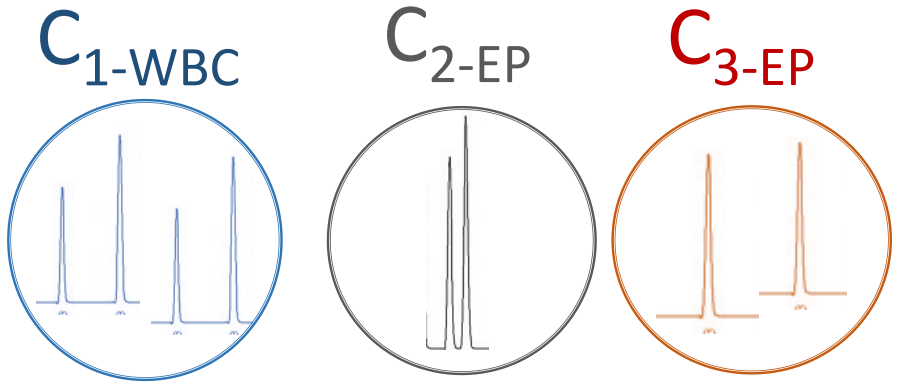




- 5 scEPGs
- One locus
- Colors=different donors
- w/ cell-type labels



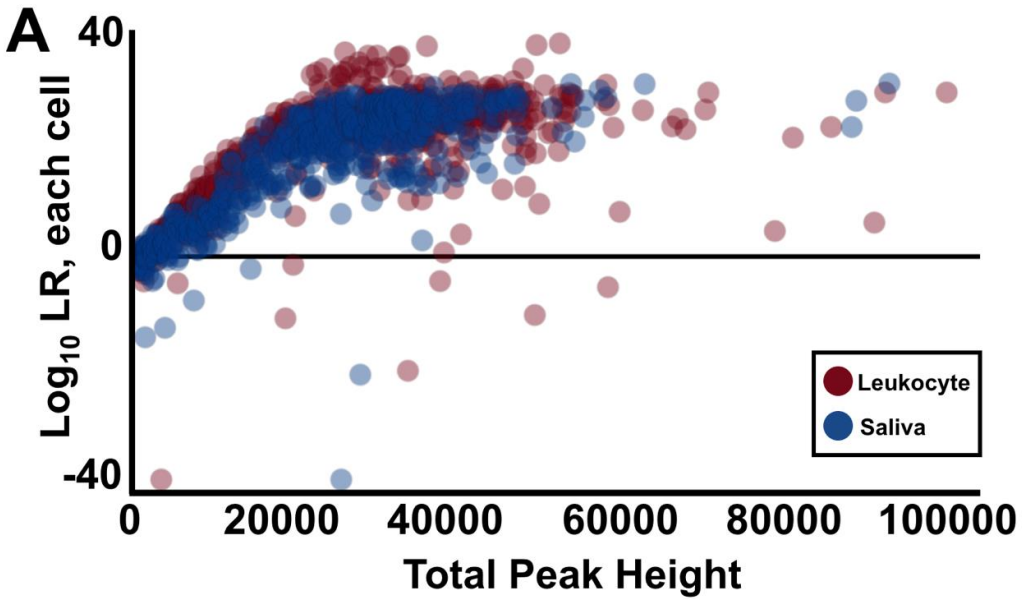




scDNA interpretation depends only on isolation throughput, and clustering is more efficient than deconvolving

The logLR of one scEPG can be just as informative as a single-source high- template EPG

Slope in linear region (0.001 [$\frac{logLR}{RFU}$]), shows that for every 1000 RFU – ca. 2 alleles – logLR will, on average, increase by 1



scWoEs of a **twelve person**, 643 cell mixture with all WoE-g_{true} near log(1/RMP), and taking **2 hours** on a laptop

Person	$\log(1/\mathrm{RMP})$ based on known genotype	Single cell log LR
1	30.59	29.88
2	29.09	28.39
3	29.58	28.69
4	29.55	28.79
5	29.41	26.59
6	31.04	30.29
7	29.00	28.29
8	29.11	28.39
9	27.37	26.69
10	28.70	27.99
11	29.78	29.09
12	38.44	37.19

Credible. Does scDNA interpretation ‘outperform’ that of mixedDNA?

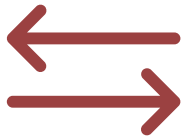
Can scDNA provide information beyond sub-sub-source and sub-source evaluations? Yes.

Where does the limit of scDNA interpretation lie? In the amount of data, not the qualities of the mixture.

What is the computational burden? Very low, expanding the remit of samples that can be faithfully interpreted.

scDNA is salient, legitimate and credible since:

it supports efficient database searching across all contributors in all mixtures, making it a fully robust data-type for investigations



is better able to discriminate hypotheses, and WoE are calibrated

– e.g., H_p and H_d across all mixtures

relies on clustering rather than deconvolving, reducing computational limitations

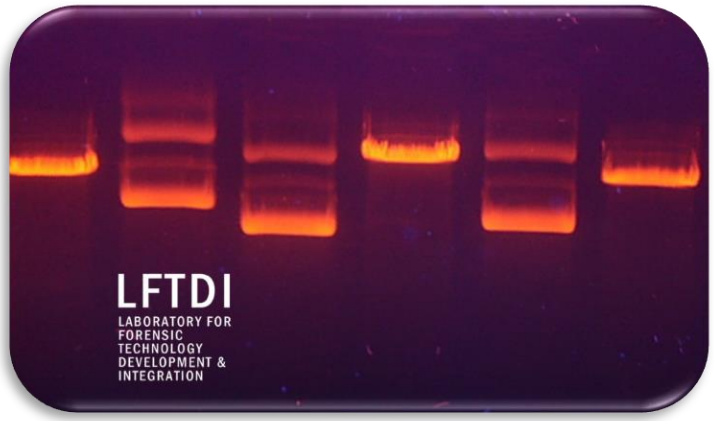
it addresses questions related to cell type

Future work will address efficiency in laboratory treatments and EESCIt™ (interpretive) expansions

Funders and collaborators

This work was supported by **NIJ2014-DN-BX-K026, NIJ2018-DU-BX-0185** and **NIJ2020-R2-CX-0032** awarded by NIJ, OJP, U.S. DOJ. 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. Also supported by **Henry Rutgers Chair Endowment**, Rutgers University

lftdi.com



RUTGERS CAMDEN

Desmond S. Lun
Ken R. Duffy
Klaas Slooten
Robert Cowell

PROVEDIt contributor:
Mike Marciano

Harish Swaminathan
Amanda Gonzalez
Leah O'Donnell
Nidhi Sheth
Qhawe Bhembe
Madison Mulcahy

Maynooth University
National University of Ireland Maynooth

Netherlands Forensic Institute
Ministry of Justice and Security

BOSTON UNIVERSITY

CAMDEN

Northeastern University