A Forensically Relevant Unsupervised Learning Approach That Accurately Clusters Single-Cell Electropherograms

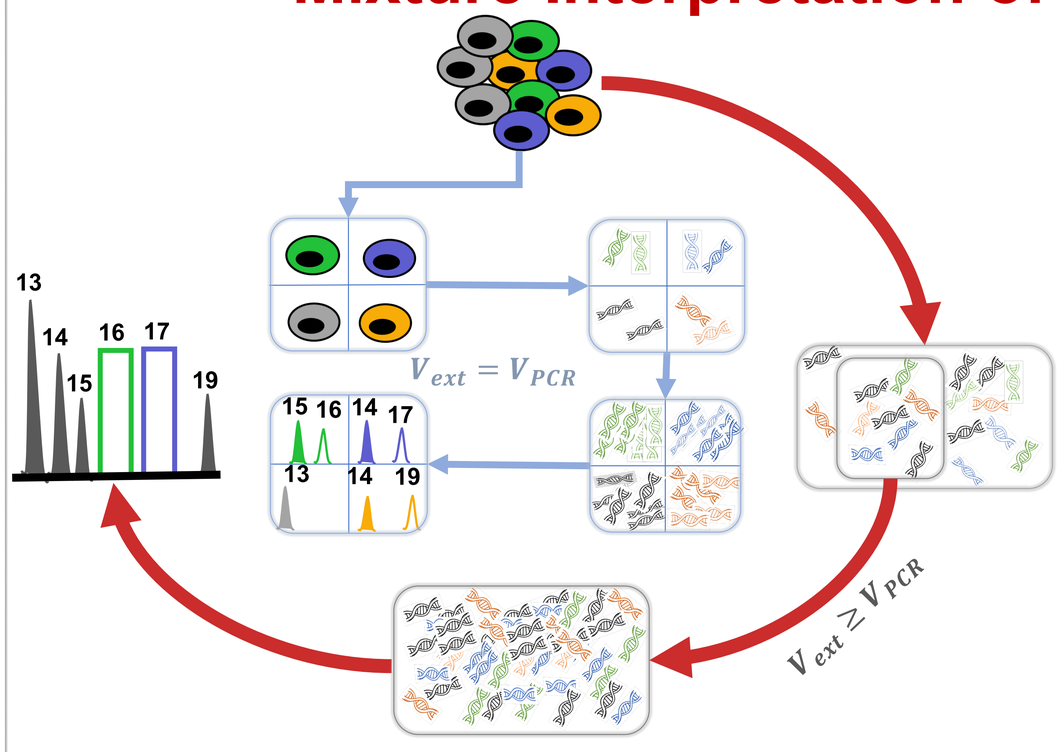
LFTDI

Nidhi Sheth [a,*], Leah O'Donnell [b], Madison Mulcahy [a], Ken R. Duffy Ph.D. [b], and Catherine M. Grgicak, Ph.D. [a]

[a]Rutgers University, Camden, US; [b]Maynooth University, Ireland

RUTGERS UNIVERSITY | CAMDEN

Mixture Interpretation of Bulk Samples vs. Single Cell Samples



Schematic of the single-cell (inner loop) and traditional bulk mixture (outer loop) pipelines.

In bulk mixture systems the cells are lysed, and the DNA is extracted into a final volume, V_{ext}. A portion of V_{ext} is transferred to a PCR tube, V_{PCR}, to which PCR reagents are added.

In contrast, single-cell systems isolate cells into distinct vessels prior to extraction, which is performed using direct-to-PCR reagents. The addition of PCR reagents to the entire V_{ext} follows, therein circumnavigating fractionation effects that lead to complete allele drop-out. Fragment analysis ensues, resulting in an EPG for each cell. Each EPG, therefore, is composed of only a single contributor's DNA.

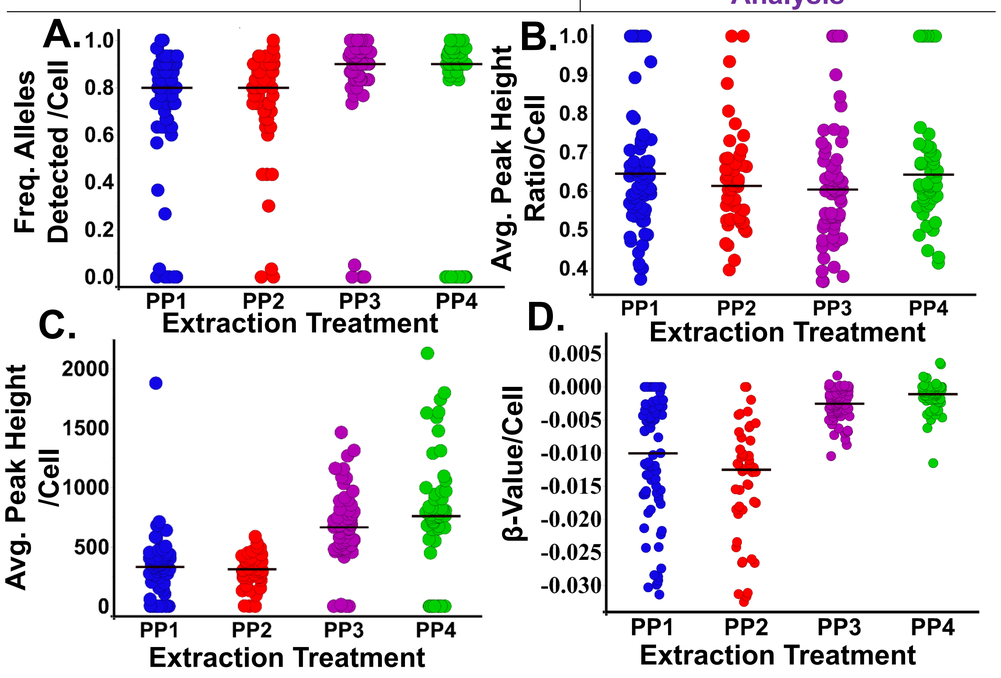
Treatment Comparison

DEParray™ benchtop instrument combines microfluidics and microelectronics into an image-based system that isolates nucleated single-cells.

Extraction Treatment Variations	
PP1	Higher Pro K and PBS (1X)
PP2	Lower Pro K and PBS (1X)
PP3	Lower Pro K and PBS (0.5X)
PP4	Lower Pro K and PBS (0.25X)

Experimental Conditions

Sample Set	241 Single-cell EPGs(scEPGs)
Amplification, Separation, Analysis	Global Filer™ (29 Cycles), 25 sec injection on 3500 Genetic analyzer, and Osiris (10.3.1) Analysis



Extraction Treatment	(A)	(B)	(C)	(D)
PP1	0.8	0.645	333	-0.01
PP2	0.8	0.614	314	-0.012
PP3	0.9	0.604	667	-0.002
PP4	0.9	0.643	761	-0.001

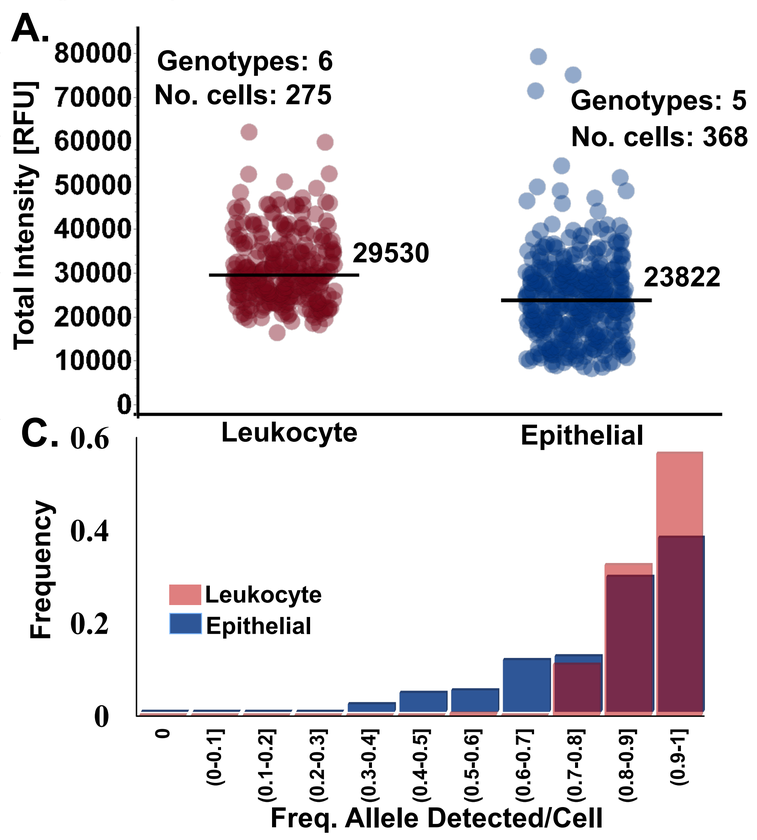
(A) The frequency of detection for heterozygous alleles separated by extraction method; (B) Average peak height ratio per cell; (C) Average peak height per cell; (D) β-value – i.e., EPG sloping – per cell. The table shows the medians of each.

Single-Cell Exploratory Analysis

Exploratory analysis on single leukocytes and epithelial cells collected via two collection techniques: i) semi-automated DEParray™ NxT and ii) manual pico-pipetting technique.

Experimental Conditions

Sample Set	Leucocyte : 275 scEPGs Epithelial : 368 scEPGs
Extraction	PicoPure Extraction [Manufacture's recommendation for Epithelial cell and PP3 when collecting with DEPArray™]
Amplification, Separation, Analysis	Global Filer™ (29 Cycles), 25 sec injection on 3500 Genetic analyzer, and Osiris(10.3.1) Analysis

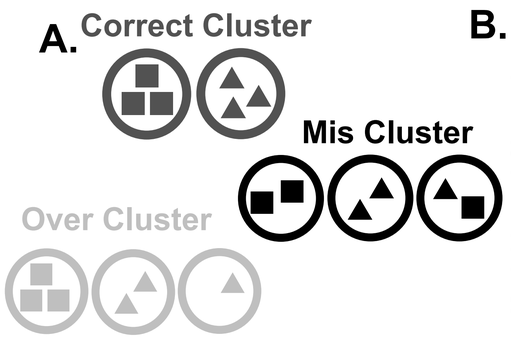


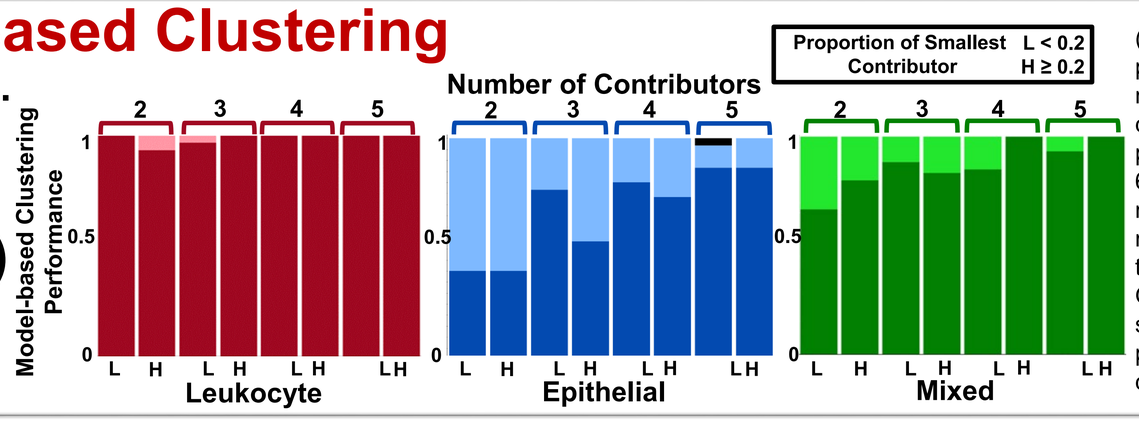
(A) Scatter plot of the Total Intensity in Relative Fluorescence Unit [RFU] per scEPG; (B) Degree of sloping (i.e. β-value) against Total Intensity [RFU]; (C) Frequency of observing heterozygous alleles for each cell type; (D) Stacked plot of heterozygous allele for different cell type across all scEPGs ordered by color and locus.

Model-Based Clustering

We developed a method to cluster/group profiles by scEPG similarity, wherein the scEPGs of each group are then used to determine the Likelihood Ratio for a cluster of cells. Various methods were evaluated, with the most favorable being the model-based clustering as implemented in the R package *mclust* [1]. There are two types of clustering errors that might occur: i) overclustering, and ii) misclustering.

Figure (A) is a schematic showing correct clustering, over clustering (known genotype is clustered in more than one group), and misclustering (where at least one group contains more than one genotype).





(B) The stacked plot depicts model-based clustering performance for 630 different mixture samples ranging from 2- to 5- persons Only one sample presented a mis clustered result.

Conclusion

1. PicoPure™ DNA extraction treatment coupled with a lower concentration of PBS and Pro K led to higher peak heights, larger allele detection rates, and lower sloping values.
2. MBC, as implemented in *mclust*, showed promising results as it correctly clustered samples consisting of 2- to 5- person mixtures with only one exception.

Reference

1. L. Scrucca, M. Fop, T. B. Murphy, and A. E. Raftery, "Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models," *The R journal,* vol. 8, no. 1,pp. 289-317, 2016.

Funding and Collaborations

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.

RUTGERS Maynooth University National University of Ireland Maynooth