OSAC RESEARCH NEEDS ASSESSMENT FORM

Title of research need: Solutions in phenotyping and ancestry analyses

Describe the need:

Forensic DNA phenotyping tends to encapsulate the prediction of externally visible traits and ancestry inference using DNA. Different subsets of markers can be used to generate information for both outcomes, with some single nucleotide markers (SNPs) overlapping. Ancestry inference and phenotype prediction use a variety of statistical calculations from probability to likelihood. There are differing prediction systems utilized, some that use regression and others that use a more Bayesian approach. Ancestry inference using small sets of ancestry informative markers (AIMS) range from using SNPs to microhaplotypes depending on the availability of phased information. Utilizing PCA data reduction techniques and reference populations to compare, continental separation is more achievable, however subpopulation level ancestry inference or admixture inference can be extremely difficult using such techniques. With regards to phenotype prediction, it is possible to predict some traits with higher accuracies than others, i.e., pigment versus height, however this depends on the number of markers used and how the prediction systems were built. Performance metrics and adequate comparisons are vital to improve systems and ensure models are built with a wide variety of variation (multiple populations to individuals with unique admixture proportions). In addition, there is a severe lack of fundamental genetic knowledge to accurately predict some characteristics, yet there are claims in the field that they are possible i.e., facial prediction. Overall, this area is very new and therefore there are many areas of improvement needed before forensic DNA phenotyping (both phenotype and ancestry) may be routine for casework implementation. Areas of improvement for both include 1. solutions for mixture deconvolution and/or low input samples using phenotype/ancestry markers - both hardware and software (i.e., MPS). 2. Development of tools and training programs for phenotyping and/or ancestry inference. 3. A comparison of the applications of phenotyping using DNA/epigenetics/proteomics. 4. Identification and characterization of traits currently available/used for prediction and detailed procedures/kits/models of emerging traits beyond pigment that show promise but should not be attempted yet.

Keyword(s): SNPs, Massive Parallel Sequencing, pipelines, software, prediction models and systems

Submitting subcommittee(s): Human Biology **Date Approved:** 10/05/2021

(If SAC review identifies additional subcommittees, add them to the box above.)

Background Information:

1. Does this research need address a gap(s) in a current or planned standard? (ex.: Field identification system for on scene opioid detection and confirmation)

Yes, in the use of Forensic DNA phenotyping (appearance and ancestry) for casework.

2. Are you aware of any ongoing research that may address this research need that has not yet been published (e.g., research presented in conference proceedings, studies that you or a colleague have participated in but have yet to be published)?

Yes.

3. Key bibliographic references relating to this research need:

1) Manfred Kayser. Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Sci Int Genet*. 2015;18:33-48,
2) Schneider PM, Prainsack B, Kayser M. The Use of Forensic DNA Phenotyping in Predicting Appearance and Biogeographic Ancestry. *Dtsch Arztebl Int*. 2019;51-52(51-52):873-880.
3) McNevin, D. Forensic inference of biogeographical ancestry from genotype: The Genetic Ancestry Lab. *WIREs Forensic Sci*. 2020; 2:e1356.
4) Kenneth K. Kidd, William C. Speed, Andrew J. Pakstis, Manohar R. Furtado, Rixun Fang, Abeer Madbouly, Martin Maiers, Mridu Middha, Françoise R. Friedlaender, Judith R. Kidd. Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Sci Int Genet*. 2014;10:23-32,
5) Christopher Phillips, Maria de la Puente; The analysis of ancestry with small-scale forensic panels of genetic markers. *Emerg Top Life Sci* 2021; ETLS20200327.
6) de la Puente M, Ruiz-Ramírez J, Ambroa-Conde A, et al. Broadening the Applicability of a Custom Multi-Platform Panel of Microhaplotypes: Bio-Geographical Ancestry Inference and Expanded Reference Data. *Front Genet*. 2020;11:581041.

4. Review the annual operational/research needs published by the National Institute of Justice (NIJ) at https://nij.ojp.gov/topics/articles/forensic-science-research-and-development-technology-working-group-operational#latest? Is your research need identified by NIJ?

No.

5. In what ways would the research results improve current laboratory capabilities?

Must compare and contrast commercial and non-commercial methods/software/lab techniques used in generating data for forensic DNA phenotyping in addition to performance metrics of prediction models and a comparison of the data used to build these models. Training of what is currently available and possible, including caveats and explanation of errors is needed. Full disclosure on prediction models, how they are built, utilized and their output must be mandatory for any approach and direct comparisons of the same samples (a control set) should ideally be implemented.

6. In what ways would the research results improve understanding of the scientific basis for the subcommittee(s)?

Better exploration of the caveats of bad samples/data and minimum requirements for accurate predictions, from the very first step in data generation, to the final stated prediction. Expansion into visual outcomes may benefit the field, especially the performances so that laboratories may decide when it may be time to implement this intelligence method in their cases.

7. In what ways would the research results improve services to the criminal justice system?

It will provide an overview of the current field in terms of commercial and non-commercial kits/systems, what is possible and what is not. It can show that successful applications of FDP can aid many cases through intelligence driven answers (alongside genetic genealogy for limiting potential relationships to explore, limiting suspect

pools for expensive STR typing, intelligence information generated through rapid DNA work that may be useful on the ground in time-sensitive cases). Training is vital for laboratory implementation of these methods and for them to move from the research realm to the practitioner realm.

8. Status assessment (I, II, III, or IV): II

	Major gap in current knowledge	Minor gap in current knowledge
No or limited current research is being conducted	I	III
Existing current research is being conducted	II	IV

This research need has been identified by one or more subcommittees of OSAC and is being provided as an informational resource to the community.