FINAL SUMMARY OVERVIEW
2014-DN-BX-K031

Title: Improving the prediction of human quantitative pigmentation traits such as eye, hair and skin color using a worldwide representation panel of US and European individuals.

PI: Susan Walsh, Indiana University Purdue University Indianapolis

Purpose of the project

The overall purpose of the project as per the accepted proposal in 2014 was "to assess eye color, hair color and skin color on a quantitative (continuous) scale for new and known associated SNPs/genes". The proposal touched on areas where improvement and fundamental research was required for more accurate prediction of pigmentation from DNA in both categorical and quantitative areas. In terms of eye color this was trying to understand green or intermediate eye color and in terms of hair color, understanding changes in ones lifetime with age, or defined as age-related hair color changes, lastly skin color required investigating both categorical and quantitative prediction systems as there was not a system in place at that time.

Project design and methods

Part One: sample collection and GWAS

There were three parts to this proposal, Part one proposed a new and highly specific genome-wide association study on quantitative color for eye, hair and skin association using a highly variable European set of 1000 individuals. We are in the midst of performing GWAS currently and are currently writing up several publications on these results.

Part Two: assessment of the most predictive SNPs for quantitative eye, hair and skin color.

Due to the nature of genome-wide association studies, many view associated pigment markers as being quite specific to the population searched (i.e. European) and in turn

have suggested that tools such as IrisPlex and HIrisPlex that have been developed on such European sets are then more accurate in the prediction of European individuals and lack accuracy in admixed individuals that display variable traits but are not European, such as individuals from a US population. To circumvent this, we assessed already-known associated hits in > 2000 US individuals (genotyping of HIrisPlex) to test epistatic interactions in addition to generating genotype frequencies of known eye, hair and skin color markers. We are still generating and assessing this data currently for their contribution to quantitative color in addition to categorical color; this data has been added to the database for further analyses, which is ongoing.

Part Three: produce an easy to use computer-software that enables users to input genotype information to yield a highly specific quantitative color result for eye, hair and skin color

The laboratory is currently in talks with a computer scientist to make a software that will take the values associated with the quantitative color space (i.e. HSV) (a reverse engineered program that was made from part one of this proposal) to simulate a color print-out for the eye, hair and skin color quantitative prediction. This work is currently in progress and shall be used to train the computer software's prediction algorithm. This work has taken longer than expected due to the time to collect, generate and input the genotypic data, which was at the core of this proposal. This has also been addressed in the most recent progress reports, however this tool is expected to be developed once all the computer program has been written.

Timeline deliverables Year 1-3

1. >2000 samples shall be collected from the US site and >1500 samples shall be collected from both the Irish and Greek site (enriched for green eye color and age related hair color changes). 200 shall be sent for SNP Chip array genotyping. **Achieved - >3600 individuals with genotype and phenotype information**
2. Data Imputation performed from SNP array data. **Achieved (even with an admixed cohort)**
3. Quantitative program to phenotype eye, hair and skin color using a new algorithm. Results shall be published in peer-reviewed journals. **A categorical program was achieved for all 3 traits; a quantitative program is in progress**
4. >3000 samples from both the US site and European sites shall be completely quantitatively phenotyped for all three pigment traits by Dr. Wollstein. **Achieved**
5. The same >3000 samples from both the US site and European sites shall be completely genotyped for all new (or near genome-wide significance) hits and already known pigment associated (either categorical or quantitative hits from previous publications) shall be genotyped at the PI's laboratory in the US. – **In Progress**
6. An assessment of the SNP genotype/phenotype database for all three quantitative traits shall be performed yielding the most predictive SNPs for all three traits individually. **- In Progress**

7. A prediction model for each trait shall be created based on this broad sample set of 3000 individuals consisting of European and US individuals with high degrees of variation in eye, hair and skin color – **In progress**, however have indicated in previous progress reports, IF novel hits were found during GWAS, it would not be appropriate to use the same set for prediction modeling, this depends on the trait and the hits that may be found (a newly submitted NIJ proposal in 2018 also addresses this item and tries to tie both grants together for this reason – more details on this can be provided if required).
8. Create a worldwide database of genotype/phenotype information to further test and improve worldwide quantitative prediction models for individualized eye, hair and skin color prediction. – **In progress**, rather than a database, a site will be made to show current models and their performance, this has been addressed in the last 2 progress reports.
9. Development of a program to generate visual quantitative output of the iris, hair and skin from genotypic data – including model training and error range. – **In progress**, due to delays in data collection and GWAS, this part of the proposal has been met with delay and it is unlikely the final deliverable shall be made prior to the completion of the grant, however this work will continue past the completion of the grant and this work will be finished and published (indicating NIJ support from this 2014 grant).

Project subjects (Year 1-3)

In order to perform this research, it was necessary to work with a worldwide representation panel of European and US individuals that cover the highest levels of variation. Therefore a large portion of this supported grant was a subject collection as it was required to perform fundamental research to find new genes related to categorical and quantitative pigment and to create an individualized system capable of a prediction print out for eye, hair and skin color. As notified in progress report 2, the Greek site was removed from this collection due to their economy. The US site collection was extremely successful and surpassed its goal with >3000 individuals collected. >600 samples were collected from the other site in Ireland, just shy of the 750 samples expected. In addition a Lebanese cohort was also collected consisting of 212 individuals to replace the Greek site. In total >3600 individuals were collected and also SNP genotyped due to changing to a different provider (with approval from NIJ and in Progress Report 2) and optimizing the quote to include genotyping for the additional individuals at the same quoted price. Full IRB approval was obtained prior to collection at all 3 sites, US, Ireland and Lebanon, these details can be included if needed.

Data Analyses (Year 1-3)

Data imputation was also performed on these individuals as per deliverables in addition to quantitative phenotyping performed by the in-house developed quantifier program developed by Dr. Walsh and Dr. Wollstein. Therefore the database now consists of >3600 individuals with imputed genotype data (1.7 million SNP-chip MEGA array, and >37 million SNPs imputed from a combined (1000 genomes and UK Haplotype Reference consortium (HRC)) dataset with phenotypic data: continuous color measure (HSV space

and % Eumelanin, Pheomelanin and no pigment – see publications section) for iris and hair imagery in addition to spectrophotometer readings of skin pigment.

The lab is currently performing genome-wide association studies on the quantitative pigment. All imagery has been quantified for eye and hair color (in addition to spectrophotometer readings for skin color. Sample categorical measures were also classified as part of the questionnaire – type of iris color, hair color and skin color (Fitzpatrick scale) according to the submitted questionnaire. The HIrisPlex-S system was made with an online webtool from data and data analyses using a subset of these samples – it combines all 3 traits for categorical prediction in one system. More information can be found in the published papers on skin color in this document.

Additional work on hair structure (outside the original scope of the proposal)

Additional work that utilized genotypic/phenotypic information derived from the support of this grant was performed with collaborators to understand the genes involved in human hair structure; whether it is curly, wavy or straight. This involved projects that searched for new markers, where this collection was used as replication, to utilizing a separate set of individuals for prediction modeling. This grants support of the sample collection meant it could be used for the descried projects without interfering with the proposed aspects of the proposal. This work will also continue.

Publications and products from this grant

Pośpiech E, Chen Y, Kukla-Bartoszek M, Breslin K EUROFORGEN NoE Consortium members, Branicki W, **Walsh S**, Liu F, Kayser M. Towards broadening Forensic DNA

Phenotyping beyond pigmentation: Improving the prediction of head hair shape from DNA. Forensic Sci. Int. Genet. 2018.

Kukla-Bartoszek M, Pośpiech E, Spólnicka M, Karłowska-Pik J, Strapagiel D, Żądzińska E, Rosset I, Sobalska-Kwapis M, Słomka M, **Walsh S**, Kayser M, Sitek A, Branicki W. Investigating the impact of age-depended hair colour darkening during childhood on DNA-based hair colour prediction with the HIrisPlex system. Forensic Sci. Int. Genet. 2018

Chaitanya L, Breslin K, Zuñiga S, Wirken L, Pospiech E, Kukla-Bartoszek M, Sijen T, de Knijff P, Liu F, Branicki W, Kayser M*, **Walsh S***. The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. Forensic Sci. Int. Genet. 2018

Liu F, Chen Y, Zhu G, Hysi P, Wu S, Adhikari K, Breslin K, Pośpiech E, Hamer MA, Peng F, Muralidharan C, Acuna-Alonzo V, Canizales-Quinteros S, Bedoya G, Gallo C, Poletti G, Rothhammer F, Bortolini MC, Gonzalez-Jose R, Zeng C, Xu S, Jin L, Uitterlinden AG, Ikram MA, van Duijn CM, Nijsten T, **Walsh S**, Branicki W,Wang S, Ruiz-Linares A, Spector T, Martin N, Medland S, Kayser K. Meta-analysis of genome-wide association studies identifies 8 novel loci involved in shape variation of human head hair. Human Molecular Genetics. 2017.

Walsh S*, Chaitanya L, Breslin K, Muralidharan C, Bronikowska A, Pospiech E, Koller

J, Kovatsi L, Wollstein A, Branicki W, Liu F and Kayser M*. Global skin color prediction from DNA. Human Genetics. 2017.

Wollstein A, **Walsh S**, Liu F, Chakravarthy U, Rahu M, Seland H, Soubrane G, Tomazzoli L *et al.* Novel quantitative pigmentation phenotyping enhances genetic association, epistasis, and prediction of human eye colour. Scientific Reports. 2017.

In house JAVA program for quantifying color based on HSV and training data – it generates the amount of Eumelanin, Pheomelanin and non-pigment pixels from an image – see above paper.

Dissemination to communities of interest:

Individuals used in the skin color category portion of research that has been funded by this grant has led to presentations and workshops at ISFG (International society of forensic genetics) for the last 2 sessions on HIrisPlex-S eye, hair and skin color categorical prediction system (with collaborators in NL and Poland) where this grant was quoted. This work has also been presented at ISHI, AAPA, GRC and the NIJ research symposium at AAFS and Webinar for NIJ's Forensic Technology Center of Excellence. Individuals from this database were also used in a hair structure publication in 2018, where additional markers were investigated for straight v curly hair. The PI has given over 15 talks and posters that reference to this NIJ grant. In addition the NIJ is mentioned on the HIrisPlex S webtool site online, which was designed/made/updated, by the PI - https://hirisplex.erasmusmc.nl/. Important to note is that research from this collection will

be coming out for the next several years that reference this grant. The research database is also a resource that other researchers can request access to through adequate collaboration. It is also the intention of the PI to produce a website informing the public of pigmentation research that is currently ongoing and its implications for forensic DNA intelligence work. With the public having a broader understanding of the other capabilities of DNA work (beyond database matching), this could better the change for DNA phenotyping to be a routine procedure used in criminal casework and in turn would change current policy and practice currently observed within the US. This online resource is in progress on the Walsh laboratory website - http://www.iupui.edu/~walshlab/

Implications for policy and practice in the United States

Currently, each state differs when it comes to DNA phenotyping and the prediction of externally visible characteristics. However several states allow this type of method to be used in the collection of intelligence information. The PI believes that with easy to use systems and tools, the practice of these DNA intelligence tools will increase use within the United States and this may even instigate a change in policy within the United States as a whole. With an increase in cold-case investigations that have no success with standard STR profiling, it is these new intelligence techniques that can really make or break a case in terms of new leads. Due to some of the work from this grant (with more to come), categorical eye, hair and skin color prediction is now available for forensic practitioners in a forensically developmentally validated assay and online webtool for prediction located at http://hirisplex.erasmusmc.nl. Forensic practitioners can choose to generate intelligence information on eye, hair and skin color

now on any case they wish using these validated and published tools. Over the course of this grant, the PI has performed work for several agencies on cold cases (n=3 cases) to generate intelligence results (eye, hair and skin color) for police investigation. These cases are ongoing and there have been no reported results from these analyses as of yet. The PI hopes to continue this work to provide a proof of principle to legislators on the usefulness of this technology, in the hopes that it may lead to change in common forensic laboratory practice within the US. This change in current practice may allow more laboratory intelligence methods, such as forensic DNA phenotyping, to be performed (with training) on difficult cases (no suspect or lead) for intelligence purposes only.