PREDICTIVE ASSOCIATIONS FROM GENOMIC DATA: FACIAL MORPHOLOGY AND AGE OF DONOR

NANI GRIMMER
RESEARCH SCIENTIST

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior Interior Business Center (DoI/ICB) contract number D15PC0002. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.
The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official polices or endorsements, either expressed or implied, of IARPA, DoI/IBC, or the US Government.
There is growing interest in the forensic science and intelligence communities to determine observable physical characteristics (phenotype) based on genomic information.

A predictive phenotypic tool would benefit from incorporating multiple characteristics including:

- Biogeographic Ancestry
- Physical Traits (i.e. hair and eye color)
- Potential craniofacial morphology
- Age of an Individual
FACIAL PREDICTION APPROACH

• Deeper investigation into the morphology of one discrete area of the face vs whole face interrogation
• Use of one population cohort to limit potential variation due to biogeographic ancestry
• Use of 3D imagery
TECHNICAL CONCEPT

• Discrete Facial Feature
  • Nasal Region

• Population cohort
  • European-Caucasian biogeographic ancestry
  • No history of nasal reconstructive surgery or breakage
  • Adults (18-60+ years old)

• 3-D Imagery using the 3dMD® Face System
• Obtaining Genomic Data
  • Whole genome sequencing
  • Human genotyping microarrays
  • Targeted re-sequencing (Targeted DNA Capture)
    • Gene regions identified by Genome Wide Association studies (GWAs)
• Targeted DNA capture
• Illumina® Miseq sequencer
  • Sequencing-by-synthesis (SBS)
  • Detection of single bases as they are incorporated
  • Base-by-base data for region of interest

http://core-genomics.blogspot.com/2015/03/a-better-way-to-sequence-exomes.html
DNA PROCESSING

• Sequencing of 18 genes
  • Known to be significantly correlated with nasal development

• Sequenced area includes areas outside of the known gene region to include potential regulatory regions
  • Identification of potential causal variants (SNPs) not yet discovered for observed nasal phenotypes
TARGETING A GENE REGION

- **Gene Region:** RDH1
- **2000 bp upstream and downstream from gene region**

[UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly](https://genome.ucsc.edu/)
FACIAL PRINCIPAL COMPONENTS

- Manually annotated facial landmarks
  - Topographic distances
  - Curvature
  - Angles
FACIAL PRINCIPAL COMPONENTS

- Facial PCs generated by algorithms developed by Oak Ridge National Laboratory (ORNL)
DATA ANALYSIS

- Principal Component Analysis (PCA)
- Predictor Variables
  - Genotype of detected variants (SNPs)
  - Sex
  - Subjective attributed characteristics
  - Age
THE IMPORTANCE OF AGE DETERMINATION

PREDICTING AGE OF DONOR

• Age approximation of an unknown individual is typically performed on skeletal remains.
• Relies on specific skeletal structures (e.g. pelvic bones, femur, teeth)

http://www.sfu.museum/forensics/eng/pg_media-media_pg/anthropologie-anthropology/
A number of studies have identified specific DNA regions where the degree of methylation is significantly correlated with the age of a donor.

http://jap.physiology.org/content/109/1/243
DNA METHYLATION

- Method of gene silencing without altering the nucleotide sequence
- C-G dinucleotides are affected
  - Most non-methylated C-G dinucleotides are correlated with housekeeping, developmental, and tissue-specific genes
  - Clusters of non-methylated CG dinucleotides = CpG islands
- Methylation status is designated by the $\beta$-value (ratio of methylated vs. non-methylated cytosines)

http://www.precisionnutrition.com/epigenetics-feast-famine-and-fatness
Problem:
- Which CpG islands are targeted?
- How many are needed?

Recommendations differ across publications
- Tissue type
- Method of methylation status determination

Example:
- As few as three CpGs for saliva samples
- Between 3 and 8 CpGs for blood samples
- Over 300 to compare across tissue types
• Evaluate age-predicting CpG loci
  • Targeting over 300 sites

• Investigate the accuracy and discriminatory power of DNA methylation patterns
  • Blood sample

• Associations between methylation age, attributed age, and chronological age
  • Subjective analysis of age
DNA PROCESSING: AGE

- Bisulfite-modification of DNA
  - Non-methylated Cytosines (C) to Uracils (U)
- Custom padlock probes
  - Targets sites of interest
- Bisulfite-sequencing on the Illumina MiSeq
- $\beta$-values will be determined for each locus in each sample
FUTURE GOALS

• **Predicting facial morphology**
  • Continue data collection
    • 3D imagery with DNA collection
  • Identify the most efficient DNA processing method or approach
  • Comparative investigations between populations
    • Ethnic populations
    • Twins
  • Identify the most effective facial phenotyping method
  • Investigate other epigenetic factors on facial morphology

• **Predicting Age of Donor**
  • Optimization of assay for different tissues
  • Validate method to establish performance metrics
Questions?
Nani.Grimmer@Bodetech.com