Proteins observed in scalp hair from preschool children

Authors: Kratika Singhal[1], Ryan D. Leib[1], Cynthia R. Rovnaghi[2], Maria Xenochristou[3], Nima Aghaeepour[3], Allis S. Chien[1], Grace K-Y. Tam[2], Monica O. Ruiz[2,5], Deendayal Dinakarpandian[4], Kanwaljeet J. S. Anand[2,5]*

Affiliations:

[1]Stanford University Mass Spectrometry (SUMS) Lab, Stanford University, Stanford, CA.

[2]Pain/Stress Neurobiology Lab, Maternal & Child Health Research Institute, Stanford University School of Medicine, Stanford, CA.

[3]Departments of Anesthesiology (Research), Biomedical Data Science & Pediatrics (Neonatology), Stanford University School of Medicine, Stanford, CA.

[4]Department of Medicine (Biomedical Informatics Research), Stanford University School of Medicine, Stanford, CA.

[5]Departments of Pediatrics (Critical Care Medicine) and Anesthesiology (by courtesy), Stanford University School of Medicine, Stanford, CA.

To whom correspondence should be addressed:

Dr. KJS Anand, 770 Welch Road, Suite 435, Stanford, CA 94304 MC: 5876.
Email: anandam@stanford.edu , Phone: (650) 498-6313 , Fax: (650) 736-9186.

One Sentence Summary: The non-structural proteins observed in scalp hair from preschool children show evidence for heritability, reflect biological functions such as brain development, or immune function and regulation of stress responses, and exhibit age- and sex-related differences across periods of early childhood development.

ABSTRACT

Early childhood experiences have long-lasting effects on subsequent mental and physical health, education, and employment. Measurement of these effects relies on insensitive behavioral signs, subjective assessments by adult observers, neuroimaging and neurophysiological studies, or remote epidemiologic outcomes. Despite intensive search, no biomarkers for developmental changes in the brain have been identified. We analyzed scalp hair from healthy children and their mothers using an unbiased proteomics platform to reveal 1368 hair proteins commonly observed in children, 1438 proteins commonly observed in mothers, and 1288 proteins observed sporadically in individual subjects. Mothers showed higher numbers of peptide spectral matches and hair proteins compared to children, with important age-related differences between mothers and children. Age-related differences were also observed in children, with differential protein expression patterns between younger (2 years and below) and older children (3-5 years). Boolean analyses showed greater conservation of hair protein patterns between mothers and their biological children as compared to mothers and unrelated children. The top 5% proteins driving population variability represent biological pathways associated with brain development, immune signaling, and stress response regulation. Non-structural proteins observed in scalp hair may include promising biomarkers to investigate the developmental changes associated with early childhood experiences.

(196 words)

INTRODUCTION

Early human development remains exquisitely sensitive to parental, environmental, and societal influences that multiplex the history of each individual (via genetic and epigenetic factors) with their daily experiences. Variations in these factors, such as the social determinants of health, can singly or collectively introduce differences in their developmental outcomes(*1, 2*). Such differences are then magnified in the higher-order cognitive and behavioral capacities of humans, which are built on a series of sequential or staggered developmental epochs that can enable or constrain their future role(s) in society, as well as their mental and physical health(*3-5*).

Although the outcomes of early childhood are easily assessed, but *objectively* assessing their social, emotional, or other environmental inputs across multiple timescales is challenging. These challenges result from most subjects being pre-verbal, coming from unknown environments, or accompanied by unreliable historians(*1, 6*). Developmental timescales can also range from milliseconds to minutes (e.g., affecting acute neuromodulatory tone, neuronal oscillations, neuroendocrine changes), days to weeks (e.g., affecting circadian rhythms, metabolic functions, memory and learning), or months to years (e.g., affecting brain growth and brain plasticity, or emerging cognitive, behavioral, or social capacities)(*7*). Neurophysiological, neuroimaging, and observational studies have attempted to describe and quantify these early developmental changes, but there remains a need for non-invasive, objective biomarkers that can be measured serially across the months and years required for childhood development(*8-10*).

Human scalp hair, derived from the neuroectoderm and mesoderm, grows constantly at about 1 cm/month and evolves via prenatal lanugo, postnatal vellus, intermediate medullary, and terminal hair stages. Hair contains 65-85% proteins, 15-35% water, 1-9% lipids, and 0.1-5% pigments like melanin and trace elements(*11*). Constantly growing scalp hair incorporates both

endogenous and exogenous proteins in a time-averaged chronological manner(*12*), unlike any other biospecimens(*13*). Therefore it is used for monitoring drug exposures, heavy metals, or other environmental toxins(*14*). Developmentally regulated hair proteins could offer biomarker candidates for brain development in early childhood.

However, all published data on hair proteins are limited to adult subjects, include relatively small sample sizes, and focus mainly on structural hair proteins. Lee et al. reported 343 hair proteins from three adults, showing evidence for post-translational modifications(*15*). Laatsch et al. analyzed hair from 18 males and 3 females, reporting ethnic differences in keratins and keratin-associated proteins (KAPs)(*16*). Carlson et al. characterized hair proteins from one adult with limited sample availability(*17*), whereas Wu et al. used hierarchical protein clustering to match 10 monozygotic twin pairs and differentiate them from unrelated individuals(*18*). Parker et al. reported quantifiable measures of identity discrimination and racial ancestry by detecting genetically variant peptides in the structural hair proteins for forensic purposes(*19*).

To fill the extant gaps in knowledge, we analyzed non-structural hair proteins using ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) in preschool children and their mothers. Our subjects were not exposed to early life adversity, as evidenced by parental income, household structure, health insurance, and parent education(*4*); all children were developmentally appropriate, healthy, and belonged to stable nuclear families (**Table S1**).

RESULTS

Features of hair proteins

Pooled hair samples from 40 children and 43 mothers were processed initially to identify non-structural hair shaft proteins, using protein purification to remove keratins. We subsequently analyzed individual hair samples from 8 mothers with 16 related children and 16 unrelated

children, to create a master library of hair proteins with 3,124 proteoforms, representing the gene products of 2,278 genes. Expression of protein isoforms, alternative splicing of messenger RNA (mRNA), and post-translation modifications resulted in a higher number of hair proteins than their associated genes(*15, 20*). These data were deposited through the PRIDE repository(*21*) into the ProteomeXchange Consortium(*22, 23*).

Hair proteins observed in individual mothers and children contained 2,269 unique 'proteoforms' or protein isoforms; 1,438 proteins were commonly observed in mothers, 1,368 proteins were commonly observed in children, whereas 1,288 hair proteins showed individual variability among mothers and children. Higher spectral counts (p=0.0004) and higher numbers of proteins (p=0.001) were observed in mothers compared to children (**Fig.1**), perhaps reflecting a wider array of biological functions in adult females related to reproduction(*24-26*), aging(*27, 28*), or disease states(*29*). These age differences are explored further in subsequent analyses.

The Uniprot database (https://www.uniprot.org/) revealed, for example, that 191 proteins observed in hair were regionally enriched in specific brain regions (based on the Allen Brain Atlas (https://human.brain-map.org/static/brainexplorer and the Human Brain Protein Atlas (https://www.proteinatlas.org/search/brain_category). Observed spectral counts for 2 brain proteins were higher in children (Lipocalin-1, Matrix metalloproteinase 7), 47 proteins were higher in mothers, 45 proteins lacked age-related differences, and 97 proteins were observed sporadically in individuals. Cellular functions and biological processes associated with the observed proteins were identified using PANTHER classification systems (**Fig.S1**, www.pantherdb.org).

Hair protein profiles in individuals and families

Peptide spectral matches for each protein were combined to compare protein expression for all individuals and assess Spearman rank correlations. Hair proteins from the mothers were closely correlated with each other, whereas hair proteins in children showed correlations based on age and sex (**Fig.2A**). Euclidean distances were calculated for pairwise comparisons between individuals (**Fig.2B**) and used for hierarchical clustering to identify subjects with similarities in the hair protein patterns (**Fig.2C**). Consistent with the correlation matrix, all mothers were clustered close together, younger children (0-2 years) were mostly located in one cluster, whereas older children were clustered with the mothers (**Fig.2C**). Boolean profiles of the hair proteins for each mother and her two children showed significantly shorter intra-family Manhattan distances (p<0.0002) as compared to 5000 'simulated' families with mismatched mothers and children (**Fig.2D**), revealing a hereditary conservation of hair protein profiles in each family.

Age- and sex-related differences in hair proteins

Principal Component Analysis (PCA)(*30-32*) and t-distributed Stochastic Neighbor Embedding (tSNE)(*33, 34*) were used to reduce the dimensionality of our data and to identify the major contributors of hair protein variability. Principal components 1-5 accounted for 61.6% of hair protein variability for all subjects, 57.5% for all children, 84.0% for all mothers, 60.8% for mothers and related children, and 62.3% for mothers and unrelated children.

Age differences were observed by plotting the first two principal components (PC1, PC2) and tSNE dimensions (**Fig. 3**). We observed clusters of the younger children and mothers, with the older children dispersed across these groups (**Fig. 3A**). Similar clusters were observed from

the remaining principal components. The tSNE projections also showed mothers located separately from the children (**Fig. 3B**). Proteins driving these differences showed higher spectral counts in mothers vs. children for SERPINB4 (serine protease inhibitor), POF1B (actin filament binder), PLEC (cytoskeleton binding protein), A2ML1 (α2-macroglobulin-like proteinase inhibitor), HIST1H3A (histone), UQCRQ (electron transfer from ubiquinol to cytochrome C), and AHCY (adenosylhomocysteine hydrolase). In contrast, mammaglobin-B (SCGB2A1), a heterodimerization protein that binds androgen and other steroids, was observed only in children (**Table S2**). Older children had higher spectral counts for PLEC (plectin), EIF3A (eukaryotic translation initiation factor 3), AHCY (adenosylhomocysteinase), HAL (histidine ammonia-lyase), and TUBA1C (Tubulin alpha 1c), whereas younger children had higher protein spectral counts for SCGB2A1 (secretoglobin 2A member 1) and CSN2 (casein beta) (**Table S3**).

Sex differences showed slightly higher spectral counts in girls vs. boys (p=0.038) but no difference in the number of proteins (**Table S1**). PCA analyses and tSNE projections showed overlapping clusters of boys and girls (**Fig. 3C, 3D**). When comparing individual proteins, higher spectral counts were observed for CSN2 (Casein beta) in boys and ALMS1 (Alström syndrome protein 1) in girls (**Table S4**).

Machine learning algorithms using Random Forest regressions(*35*) were used to predict age from their hair protein profiles. This model predicted age differences in mothers and children (R^2=0.37, **Fig.4A**), but performance improved (R^2=0.45) when predicting age for all children (**Fig.4B**). Random Forest classifier algorithm showed an acceptable mean accuracy for classifying mothers and children based on their predicted vs. observed age (mean area under the ROC curve = 0.93, **Fig.4C**; Wilcoxon test p=0.00011, **Fig.4D**).

A Random Forest classifier to predict sex from hair protein profiles in children could not reliably differentiate boys from girls (mean area under the ROC curve = 0.6, **Fig.4E**; Wilcoxon test p = 0.1703; **Fig.4F**), but predictions improved when classifying all participants including mothers and children (area under the ROC curve = 0.73, **Fig.4G**; Wilcoxon test p = 0.0083, **Fig.4H**). The latter result is likely due to the age-based distinction between mothers and children, although sample size-related effects cannot be ruled out (25 vs. 17 females).

The top contributors to hair protein variability

The top 5% proteins were identified as the most prominent contributors based on their total loading scores (TLS), explaining 64.3% of hair protein variability in all individuals, 89.5% in all mothers, 57.5% in all children, 49.3% in mothers and related children, and 64.6% in mothers and unrelated children (**Fig.5**). Keratins and KAPs are structural components of hair, but usually considered contaminants in most proteomics experiments, due to their high abundance in common lab contaminants. We therefore performed PCA analyses for all individuals with (**Fig.5A**) and without excluding the keratins and KAPs (**Fig.5B**). Structural proteins contributed to hair protein variability but with no differences in their biological significance. Separate PCA analyses performed to characterize the hair proteins observed in subpopulations of mothers (**Fig.5C**), children (**Fig.5D**), mothers and related children (**Fig.5E**), and mothers and unrelated children (**Fig.5F**) showed the same proteins ranked in all individuals and all children. Other than the histones, no other proteins were common between mothers and children. Tubulin alpha 1c (TUBA1C) was the only hair protein ranked within the top 5% in all groups, whereas PLEC, SERPINB4, and UQCRQ were observed in multiple groups.

Biological role(s) of the strongest contributors to hair protein variability

Based on experimentally observed human data in the Ingenuity Knowledge Base, log-fold-change values of the top 5% proteins from our dataset were used to analyze direct and indirect relationships between molecules. Protein networks for the top 5% hair proteins contributing to age-related differences between mothers and children (**Fig.6**) and similar analyses for sex-related differences between girls and boys were examined (**Fig.7**). Using these molecular relationships as input for Ingenuity Pathway Analysis, we found proteins involved in *cellular metabolism* including the protein ubiquitination pathway, Sirtuin signaling pathway, 14-3-3 mediated signaling, Wnt-Ca^{++} pathway, histidine degradation, mitochondrial function, and oxidative phosphorylation (**Fig.8**). Other proteins were associated with *immune responses*, including phagosome maturation, IL-8 signaling, and regulation of macrophages, fibroblasts, and endothelial cells, or involved in the regulation of *stress-related pathways*, including corticotropin releasing hormone signaling, glucocorticoid receptor signaling, prolactin and aldosterone signaling. Finally, hair proteins associated with *brain development* including axonal guidance and gap junction signaling were also identified (**Fig.8**).

DISCUSSION

Growing hair provides a time-averaged molecular record unavailable from any other tissue or biospecimens(*13*). The hair follicle includes most cellular processes, including cell growth, cell death, interactions of histologically different cell types, cell migration and differentiation, and it responds to multiple hormones, microenvironmental and systemic changes(*36*). Structural and functional characteristics of hair follicles also mimic the renal tubules, potentially performing unique, distributed excretory functions(*37*).

The chemical composition of hair(*11, 38-42*) and its structural proteins (keratins, KAPs) are well-studied(*16-19*), but with minimal data on non-structural hair proteins. We found 2,269 non-structural hair proteins with important differences between mothers and children, age- and sex-related differences among preschool children, and conservation of protein profiles within families. Hair proteins driving variability in different populations were found to play vital roles in functions other than those of trichocytes in the hair follicle, including cellular metabolic pathways, brain development, immune signaling, and stress regulation. Thus, unbiased or targeted protein profiles from single or serial hair samples (or sequential hair segments) could be used as probes for child development(*43, 44*) or life-course studies(*29, 45, 46*).

We observed age-related hair protein profiles in children and mothers, with distinct patterns emerging in multiple analyses. Differences between mothers and children were largely driven by increased maternal expression of SERPINB4, PLEC, and UQCRQ. SERPINB4 is a granzyme inhibitor linked with squamous cell carcinomas and chronic liver disease(*47-49*), Plectin mutations were linked with epidermolysis bullosa simplex, but has also emerged as novel susceptibility gene for testicular germ cell tumors(*50-52*), and UQCRQ is a nuclear protein in the mitochondrial respiratory chain complex III essential for normal brain development(*53*). Mammaglobin-B (SCGB2A1), which is linked with familial febrile seizures in preschool children(*54, 55*) and chemoresistant cancers in adults(*56*), was observed only in children's hair.

Our analyses showed minimal sex differences in early childhood, confirmed by Random Forest predictive models. Biological pathways for cellular metabolism and innate immunity appeared more prominent in girls, whereas brain development and regulation of stress responses appeared more prominent in boys. Perhaps sex differences in hair proteins may be accentuated following the onset of puberty(*57*). Although hair protein profiles were conserved in mothers and

their biological children, future studies in mother-child dyads and monozygotic vs. dizygotic twins are required to quantify the heritability of hair proteins(*58*).

Experimental human data from the Ingenuity Pathway Analysis identified hair proteins associated with axon guidance(*59*) and gap junction signaling(*60*). For example, Tubulin alpha 1C was identified from all sub-populations in our dataset. TUBA1C (on chromosome 12) is amongst the most over-expressed brain-specific genes, forming a major component of the microtubules that determine neuronal shape and function, and is typically used as a marker of neural differentiation(*61*). We observed 191 hair proteins regionally enriched in the brain and speculate that hair proteomics could complement neuroimaging and neurophysiological data on early brain development(*8*). For example, specific plasma proteins were associated with higher non-verbal intelligence and pro-inflammatory proteins associated with lower intelligence in children from Nepal(*62*). That study used an FDR of 5%, whereas our FDR threshold was set at 1% or lower.

Future developmental studies with larger sample sizes could correlate hair proteins with cognitive or behavioral outcomes, thus investigating their role in normal and altered brain development(*63*). Clinical trials of novel therapeutics to enhance brain recovery, in stroke or traumatic brain injury for example, would also benefit from brain-region specific biomarkers. *"Despite intensive search, there is no such biomarker in brain imaging or serum"(64)* and therefore, clinical trials of neural repair therapeutics have suffered from reliance on insensitive behavioral outcomes(*64*).

These findings must be interpreted in the light of three limitations. First, our sample size of 32 children was insufficient to examine developmental differences at each age in the preschool period. We selected healthy children from homogenous socioeconomic environments;

they did not experience any adverse conditions and therefore, our data may not represent the full range of hair protein profiles present in the general population. Our sample size, however, is larger than most other studies of hair proteomics in adults and it is the first to include mothers and children. Our study design also allowed us to investigate differences in hair protein profiles between related and unrelated individuals, as well as differences between adults and children. Second, our proteomics platform relied on peptide spectral matches, which presented only semi-quantitative data on the abundance of hair proteins between individuals. Since this was the first-ever study investigating non-structural proteins from hair in humans, we chose a 'shot-gun' proteomics approach rather than targeted and more quantitative approaches. Now that we have established large libraries of hair proteins for mothers and children, future studies can be designed for the quantitation of specific protein targets or protein groups. Lastly, we did not correlate hair proteins with the child's developmental milestones or their cognitive and behavioral data. We feel that the sample size limitations at each age would preclude any generalizable conclusions from such analyses.

Despite these limitations, our initial findings reveal the potential role of non-structural hair proteins as biomarkers for brain development, metabolic, immune, or stress-regulatory pathways, providing a rich source of chronologically-ordered information for life-course studies and early childhood development.

MATERIALS AND METHODS

After IRB approval and parental consent, mothers and children aged 1-6 years were enrolled from local preschool facilities. We excluded children with tinea capitis, alopecia areata, eczema, or other scalp conditions; those receiving any prescription or over-the-counter drugs; or steroid therapy in the past 3 months; or those with chronic medical conditions, developmental delay, or chemical exposures to hair prior to study entry. Hair samples from the posterior vertex area were trimmed at 0.1mm from the scalp, weighed, and stored in Ziploc® bags at 4ºC.

Hair extraction: Finely chopped hair (10 mg) was incubated in 500µl Tri-Reagent® (Molecular Research Center, Inc; Fisher Scientific) at room temperature (RT) for 10min, followed by 100µl chloroform (shaken vigorously for 15sec, stored for 10min). Samples were centrifuged at 12000g, 4ºC for 10 min to separate the aqueous RNA fraction. Samples were incubated with 1.5ml acetone at –30ºC for 1 hour, followed by centrifugation at 8000g for 5 min to sediment protein. The supernatant was removed, the sediment suspended in 1ml, 0.3 M guanidine hydrochloride/95% ethanol/2.5% glycerol, stored for 10min at RT, followed by centrifugation at 8000g for 5 min. The supernatant was decanted and the wash repeated twice. The protein pellet was washed twice in 95% ethanol (without glycerol), centrifuged for 5min, 8000g at 4ºC and the pellet stored frozen under fresh 95% ethanol at –20ºC.

Proteomics methods: Samples were brought to RT, centrifuged at 8000g for 5 min, the ethanol decanted and dried under vacuum. Protein pellets were resuspended in 50mM ammonium bicarbonate in the presence of 0.0015% ProteaseMAX (Promega) and total protein amount was estimated with Pierce BCA assays (Thermo Fisher Scientific). Proteins were digested with

0.25µg of Trypsin/LysC (Promega) at a 1:100 enzyme/substrate ratio overnight at 37°C. Proteolytic digestion was quenched with 1% formic acid; peptides were dried by speed vac before dissolving in 30µl of reconstitution buffer (2% acetonitrile + 0.1% Formic acid) to a concentration of 1 µg/µl; 2µl of this solution was injected into the MS instrument.

Experiments were performed on the Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) coupled with ACQUITY M-Class ultra-performance liquid chromatography (UPLC, Waters Corporation). For a typical LCMS experiment (Liquid Chromatography/Mass Spectrometry), a flow rate of 450 nL/min was used, where mobile phase A is 0.2% formic acid in water and mobile phase B is 0.2% formic acid in acetonitrile. Analytical columns were pulled using fused silica (I.D. 100 microns) and packed with Magic 1.8-micron 120Å UChrom C18 stationary phase (nanoLCMS Solutions) to a length of ~25 cm. Peptides were directly injected onto the analytical column using a gradient (2-45% B, followed by a high-B wash) of 80 minutes. The MS was operated in data-dependent fashion using CID (collision induced dissociation) for generating MS/MS spectra, collected in the ion trap with collisional energy set at 35.

The *.RAW data files were processed using Byonic v3.2.0 (ProteinMetrics) to infer protein isoforms using the Uniprot *homo sapiens* database. Proteolysis with Trypsin/LysC was assumed to be semi-specific allowing for N-ragged cleavage with up to 2 missed cleavage sites. Precursor mass accuracies were held within 12 ppm and 0.4 Da for MS/MS fragments. Proteins were held to a false discovery rate (FDR) of 1% or lower, using standard target-decoy approaches(*65*), and only the proteins with >3 spectral counts were selected for further data processing; keratins and KAPs were removed at this stage.

Statistical analysis: Spectral counts were used to calculate Euclidean distances between individuals; Euclidean distances were used for hierarchical clustering. A correlation matrix with Spearman's coefficient was also used for rank-based depiction of similarities between the individual hair proteomes.

Principal Component Analysis (PCA) was used to reduce dimensionality of the dataset. PCA is a widely used technique for data analysis and modeling(*30*) of linear combinations of the original dimensions called principal components(*32*). The largest proportion of data variance is captured by the first principal component, the second largest proportion of variance falls along the second principal component, and so on(*31*). For the first five principal components from each PCA, we multiplied the loading scores for each protein by the percent variance explained by that corresponding principal component; the weighted scores were summed for each protein to give its Total Loading Score (TLS).

*Weighted Score = Loading Score * Proportion of Variance*

$$Total\ Loading\ Score\ (TLS) == \sum_{PC=1}^{5} Weighted\ score$$

Based on their TLS values, top 5% proteins were selected as the main drivers of variability in hair protein expression.

Additionally, we used t-distributed stochastic neighboring embedding (tSNE), a non-linear probabilistic approach(*33, 34*), to visualize proteins with non-linear similarity in high-dimensional space as neighbors in low-dimensional linear depictions. Unlike the reproducible PCA results, the probabilistic nature of tSNE can result in somewhat different results with each computation. To avoid serendipitous results, we ran each computation at least 10 times to ensure reproducibility. For each computation, the maximum number of iterations to converge was set to

1000, and perplexity set to the maximum permitted value. Statistical significance of tSNE clusterings was calculated by how often a given statistic was reproduced in 1000 simulations of permuted versions of the dataset.

Boolean profiles of the hair proteins were also compared between the original dataset (each mother with her own children) and 5000 simulated datasets, created by swapping mothers between families such that no mother was paired with her own children, but the two siblings remained together in all simulated datasets. Observed conservation in pairwise intra-family Manhattan distances from the original dataset could then be attributed to the similarities in hair protein expression between each mother and her children.

For the top 5% proteins in children (n=32), we averaged spectral counts for girls and boys separately, and divided the girls' average by the boys' average. Resulting values were converted to log-base 2. The same process was followed for spectral counts from mothers and children. Log fold-change values of the top 5% proteins were used as input for Ingenuity Pathway Analysis (Qiagen: https://digitalinsights.qiagen.com/products/features/). We analyzed direct and indirect relationships between molecules based on experimentally observed data, restricted to human databases in the Ingenuity Knowledge Base.

We used Random Forest (RF) models for both the classification (boy vs. girl, mother vs. child) and regression (age prediction) tasks, with protein concentrations as model features and individuals as samples(*35*). In classification, the model output was the probability of an individual being female (sex classification), or being a mother (person classification). For regression (age prediction), the model output was the individual's predicted age.

Results were based on a 10-fold cross-validation repeated 100 times. Members of the same family were included in the same set, i.e. either training or test sets, to avoid information

leak due to familial similarities. For the age prediction, we evaluated results using the R^2

coefficient of determination and the linear model p-value fitted on the predicted and observed

data. For the classification tasks, we used area under the ROC curve (AUC) and the Wilcoxon-

Mann-Whitney-test, testing the null-hypothesis that one distribution is not stochastically greater

than the other.

REFERENCES

1. R. H. Bitsko, J. R. Holbrook, L. R. Robinson, J. W. Kaminski, R. Ghandour, R. Smith, G. Peacock, Health Care, Family, and Community Factors Associated with Mental, Behavioral, and Developmental Disorders in Early Childhood — United States, 2011–2012. *MMWR Morb Mortal Wkly Rep* **65**, 221-226 (2016).

2. M. Lopez, M. O. Ruiz, C. R. Rovnaghi, G. K.-Y. Tam, J. Hiscox, I. H. Gotlib, D. A. Barr, V. G. Carrion, K. J. S. Anand, The Social Ecology of Childhood and Early Life Adversity. *Pediatric Research* **(in press)**, (2020).

3. C. A. Nelson, R. D. Scott, Z. A. Bhutta, N. B. Harris, A. Danese, M. Samara, Adversity in childhood is linked to mental and physical health throughout life. *BMJ* **371**, m3048 (2020).

4. K. J. S. Anand, J. Rigdon, C. R. Rovnaghi, F. Qin, S. Tembulkar, N. Bush, K. LeWinn, F. A. Tylavsky, R. Davis, D. A. Barr, I. H. Gotlib, Measuring socioeconomic adversity in early life. *Acta Paediatr* **108**, 1267-1277 (2019).

5. F. B. Palmer, K. J. S. Anand, J. C. Graff, L. E. Murphy, Y. Qu, E. Volgyi, C. R. Rovnaghi, A. Moore, Q. T. Tran, F. A. Tylavsky, Early adversity, social-emotional development, and stress in urban 1-year-old children. *J Pediatr* **163**, 1733-1739 e1731 (2013).

6. K. J. S. Anand, C. R. Rovnaghi, J. Rigdon, F. Qin, S. Tembulkar, L. E. Murphy, D. A. Barr, I. H. Gotlib, F. A. Tylavsky, Demographic and psychosocial factors associated with hair cortisol concentrations in preschool children. *Pediatr Res* **87**, 1119-1127 (2020).

7. R. K. Reh, B. G. Dias, C. A. Nelson, 3rd, D. Kaufer, J. F. Werker, B. Kolb, J. D. Levine, T. K. Hensch, Critical period regulation across multiple timescales. *Proc Natl Acad Sci U S A* **117**, 23242-23251 (2020).

8. T. T. Brown, T. L. Jernigan, Brain development during the preschool years. *Neuropsychol Rev* **22**, 313-333 (2012).

9. T. L. Jernigan, T. T. Brown, D. J. Hagler, Jr., N. Akshoomoff, H. Bartsch, E. Newman, W. K. Thompson, C. S. Bloss, S. S. Murray, N. Schork, D. N. Kennedy, J. M. Kuperman, C. McCabe, Y. Chung, O. Libiger, M. Maddox, B. J. Casey, L. Chang, T. M. Ernst, J. A. Frazier, J. R. Gruen, E. R. Sowell, T. Kenet, W. E. Kaufmann, S. Mostofsky, D. G. Amaral, A. M. Dale, N. Pediatric Imaging, S. Genetics, The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository. *Neuroimage* **124**, 1149-1154 (2016).

10. L. Tozzi, L. Garczarek, D. Janowitz, D. J. Stein, K. Wittfeld, H. Dobrowolny, J. Lagopoulos, S. N. Hatton, I. B. Hickie, A. Carballedo, S. J. Brooks, D. Vuletic, A. Uhlmann, I. M. Veer, H. Walter, R. Bulow, H. Volzke, J. Klinger-Konig, K. Schnell, D. Schoepf, D. Grotegerd, N. Opel, U. Dannlowski, H. Kugel, E. Schramm, C. Konrad, T. Kircher, D. Juksel, I. Nenadic, A. Krug, T. Hahn, O. Steinstrater, R. Redlich, D. Zaremba, B. Zurowski, C. H. Y. Fu, D. Dima, J. Cole, H. J. Grabe, C. G. Connolly, T. T. Yang, T. C. Ho, K. Z. LeWinn, M. Li, N. A. Groenewold, L. E. Salminen, M. Walter, A. N. Simmons, T. G. M. van Erp, N. Jahanshad, B. T. Baune, N. J. A. van der Wee, M. J. van Tol, B. Penninx, D. P. Hibar, P. M. Thompson, D. J. Veltman, L. Schmaal, T. Frodl, E.-M. D. D. C. for the, Interactive impact of childhood maltreatment, depression, and age on cortical brain structure: mega-analytic findings from a large multi-site cohort. *Psychol Med* **50**, 1020-1031 (2020).

11. N. Nicolaides, S. Rothman, Studies on the chemical composition of human hair fat. II. The overall composition with regard to age, sex and race. *J Invest Dermatol* **21**, 9-14 (1953).

12. H. A. Adeola, J. C. Van Wyk, A. Arowolo, R. M. Ngwanya, K. Mkentane, N. P. Khumalo, Emerging Diagnostic and Therapeutic Potentials of Human Hair Proteomics. *Proteomics Clin Appl* **12**, (2018).

13. D. J. Tobin, in *Hair in toxicology: an important bio-monitor,* D. J. Tobin, Ed. (Royal Society of Chemistry (U.K.), Cambridge, 2005), chap. Chapter 2, pp. 34-56.

14. M. Villain, V. Cirimele, P. Kintz, Hair analysis in toxicology. *Clin Chem Lab Med* **42**, 1265-1272 (2004).

15. Y. J. Lee, R. H. Rice, Y. M. Lee, Proteome analysis of human hair shaft: from protein identification to posttranslational modification. *Mol Cell Proteomics* **5**, 789-800 (2006).

16. C. N. Laatsch, B. P. Durbin-Johnson, D. M. Rocke, S. Mukwana, A. B. Newland, M. J. Flagler, M. G. Davis, R. A. Eigenheer, B. S. Phinney, R. H. Rice, Human hair shaft proteomic profiling: individual differences, site specificity and cuticle analysis. *PeerJ* **2**, e506 (2014).

17. T. L. Carlson, M. Moini, B. A. Eckenrode, B. M. Allred, J. Donfack, Protein extraction from human anagen head hairs 1-millimeter or less in total length. *Biotechniques* **64**, 170-176 (2018).

18. P. W. Wu, K. E. Mason, B. P. Durbin-Johnson, M. Salemi, B. S. Phinney, D. M. Rocke, G. J. Parker, R. H. Rice, Proteomic analysis of hair shafts from monozygotic twins: Expression profiles and genetically variant peptides. *Proteomics* **17**, (2017).

19. G. J. Parker, T. Leppert, D. S. Anex, J. K. Hilmer, N. Matsunami, L. Baird, J. Stevens, K. Parsawar, B. P. Durbin-Johnson, D. M. Rocke, C. Nelson, D. J. Fairbanks, A. S. Wilson, R. H. Rice, S. R. Woodward, B. Bothner, B. R. Hart, M. Leppert, Demonstration of Protein-Based Human Identification Using the Hair Shaft Proteome. *PLoS One* **11**, e0160653 (2016).

20. N. R. Barthelemy, A. Bednarczyk, C. Schaeffer-Reiss, D. Jullien, A. Van Dorsselaer, N. Cavusoglu, Proteomic tools for the investigation of human hair structural proteins and evidence of weakness sites on hair keratin coil segments. *Anal Biochem* **421**, 43-55 (2012).

21. Y. Perez-Riverol, A. Csordas, J. Bai, M. Bernal-Llinares, S. Hewapathirana, D. J. Kundu, A. Inuganti, J. Griss, G. Mayer, M. Eisenacher, E. Perez, J. Uszkoreit, J. Pfeuffer, T. Sachsenberg, S. Yilmaz, S. Tiwary, J. Cox, E. Audain, M. Walzer, A. F. Jarnuczak, T. Ternent, A. Brazma, J. A. Vizcaino, The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450 (2019).

22. E. W. Deutsch, A. Csordas, Z. Sun, A. Jarnuczak, Y. Perez-Riverol, T. Ternent, D. S. Campbell, M. Bernal-Llinares, S. Okuda, S. Kawano, R. L. Moritz, J. J. Carver, M. Wang, Y. Ishihama, N. Bandeira, H. Hermjakob, J. A. Vizcaino, The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res* **45**, D1100-D1106 (our dataset identifier: PXD015486 and DOI:015410.016019/PXD015486.) (2017).

23. K. J. S. Anand, R. D. Leib, C. R. Rovnaghi, K. Singhal, S. U. S. o. M. Pain/Stress Neurobiology Lab, Ed. (PRIDE Repository, ProteomeXchange, 2019), chap. December 22, 2019.

24. S. K. Handelman, R. Romero, A. L. Tarca, P. Pacora, B. Ingram, E. Maymon, T. Chaiworapongsa, S. S. Hassan, O. Erez, The plasma metabolome of women in early pregnancy differs from that of non-pregnant women. *PLoS One* **14**, e0224682 (2019).

25. R. Romero, O. Erez, E. Maymon, P. Chaemsaithong, Z. Xu, P. Pacora, T. Chaiworapongsa, B. Done, S. S. Hassan, A. L. Tarca, The maternal plasma proteome changes as a function of gestational age in normal pregnancy: a longitudinal study. *Am J Obstet Gynecol* **217**, 67 e61-67 e21 (2017).

26. J. Gomes, F. Au, A. Basak, S. Cakmak, R. Vincent, P. Kumarathasan, Maternal blood biomarkers and adverse pregnancy outcomes: a systematic review and meta-analysis. *Crit Rev Toxicol* **49**, 461-478 (2019).

27. J. Jylhävä, N. L. Pedersen, S. Hägg, Biological Age Predictors. *EBioMedicine* **21**, 29-36 (2017).

28. J. Lara, R. Cooper, J. Nissan, A. T. Ginty, K. T. Khaw, I. J. Deary, J. M. Lord, D. Kuh, J. C. Mathers, A proposed panel of biomarkers of healthy ageing. *BMC Med* **13**, 222 (2015).

29. J. Deelen, J. Kettunen, K. Fischer, A. van der Spek, S. Trompet, G. Kastenmuller, A. Boyd, J. Zierer, E. B. van den Akker, M. Ala-Korpela, N. Amin, A. Demirkan, M. Ghanbari, D. van Heemst, M. A. Ikram, J. B. van Klinken, S. P. Mooijaart, A. Peters, V. Salomaa, N. Sattar, T. D. Spector, H. Tiemeier, A. Verhoeven, M. Waldenberger, P. Wurtz, G. Davey Smith, A. Metspalu, M. Perola, C. Menni, J. M. Geleijnse, F. Drenos, M. Beekman, J. W. Jukema, C. M. van Duijn, P. E. Slagboom, A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat Commun* **10**, 3346 (2019).

30. H. Hotelling, Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* **24**, 417-441 (1933).

31. H. Abdi, L. J. Williams, Principal component analysis. *WIREs Comput Stat* **2**, 433-459 (2010).

32. I. T. Jolliffe, in *Principal Component Analysis*. (Springer, New York, NY, 1986), chap. 2, pp. 8-22.

33. L. van der Maaten, G. Hinton, Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**, 2579-2605 (2008).

34. G. C. Linderman, S. Steinerberger, Clustering with t-SNE, provably. *SIAM J Math Data Sci* **1**, 313-332 (2019).

35. L. Breiman, Random Forests. *Machine Learning* **45**, 5-32 (2001).

36. D. J. Tobin, Biochemistry of human skin--our brain on the outside. *Chem Soc Rev* **35**, 52-67 (2006).

37. J. L. Carre, T. Suzuki, R. Paus, Do hair follicles operate as primitive, multifocal kidney-like excretory (mini-) organs? *Exp Dermatol* **29**, 357-365 (2020).

38. A. Franbourg, P. Hallegot, F. Baltenneck, C. Toutain, F. Leroy, Current research on ethnic hair. *J Am Acad Dermatol* **48**, S115-119 (2003).

39. A. S. Wilson, D. J. Tobin, in *Aging Hair,* D. J. T. R. M. Trueb, Ed. (Springer-Verlag, Berlin, Germany, 2010), chap. Chapter 24, pp. 249-261.

40. C. F. Cruz, M. M. Fernandes, A. C. Gomes, L. Coderch, M. Marti, S. Mendez, L. Gales, N. G. Azoia, U. Shimanovich, A. Cavaco-Paulo, Keratins and lipids in ethnic hair. *Int J Cosmet Sci* **35**, 244-249 (2013).

41. A. L. Horvath, Solubility of structurally complicated materials: 3. Hair. *ScientificWorldJournal* **9**, 255-271 (2009).

42. M. Marti, C. Barba, A. M. Manich, L. Rubio, C. Alonso, L. Coderch, The influence of hair lipids in ethnic hair properties. *Int J Cosmet Sci* **38**, 77-84 (2016).

43. M. S. Breen, S. Ozcan, J. M. Ramsey, Z. Wang, A. Ma'ayan, N. Rustogi, M. G. Gottschalk, M. J. Webster, C. S. Weickert, J. D. Buxbaum, S. Bahn, Temporal proteomic profiling of postnatal human cortical development. *Transl Psychiatry* **8**, 267 (2018).

44. B. C. Carlyle, R. R. Kitchen, J. E. Kanyo, E. Z. Voss, M. Pletikos, A. M. M. Sousa, T. T. Lam, M. B. Gerstein, N. Sestan, A. C. Nairn, A multiregional proteomic survey of the postnatal human brain. *Nat Neurosci* **20**, 1787-1795 (2017).

45. C. R. Pedlar, J. Newell, N. A. Lewis, Blood Biomarker Profiling and Monitoring for High-Performance Physiology and Nutrition: Current Perspectives, Limitations and Recommendations. *Sports Med* **49**, 185-198 (2019).

46. A. A. Johnson, M. N. Shokhirev, T. Wyss-Coray, B. Lehallier, Systematic review and analysis of human proteomics aging studies unveils a novel proteomic aging clock and identifies key processes that change with age. *Ageing Res Rev* **60**, 101070 (2020).

47. S. Markovina, S. Wang, L. E. Henke, C. J. Luke, S. C. Pak, T. DeWees, J. D. Pfeifer, J. K. Schwarz, W. Liu, S. Chen, D. Mutch, X. Wang, M. A. Powell, B. A. Siegel, F. Dehdashti, G. A. Silverman, P. W. Grigsby, Serum squamous cell carcinoma antigen as an early indicator of response during therapy of cervical cancer. *Br J Cancer* **118**, 72-78 (2018).

48. P. M. van Kempen, R. Noorlag, J. E. Swartz, N. Bovenschen, W. W. Braunius, J. F. Vermeulen, E. M. Van Cann, W. Grolman, S. M. Willems, Oropharyngeal squamous cell carcinomas differentially express granzyme inhibitors. *Cancer Immunol Immunother* **65**, 575-585 (2016).

49. A. Biasiolo, N. Tono, M. Ruvoletto, S. Quarta, C. Turato, G. Villano, L. Beneduce, G. Fassina, C. Merkel, A. Gatta, P. Pontisso, IgM-linked SerpinB3 and SerpinB4 in sera of patients with chronic liver disease. *PLoS One* **7**, e40658 (2012).

50. A. Charlesworth, C. Chiaverini, J. Chevrant-Breton, M. DelRio, A. Diociaiuti, R. P. Dupuis, M. El Hachem, B. Le Fiblec, A. M. Sankari-Ho, A. Valhquist, E. Wierzbicka, J. P. Lacour, G. Meneguzzi, Epidermolysis bullosa simplex with PLEC mutations: new phenotypes and new mutations. *Br J Dermatol* **168**, 808-814 (2013).

51. L. Winter, M. Turk, P. N. Harter, M. Mittelbronn, C. Kornblum, F. Norwood, H. Jungbluth, C. T. Thiel, U. Schlotzer-Schrehardt, R. Schroder, Downstream effects of plectin mutations in epidermolysis bullosa simplex with muscular dystrophy. *Acta Neuropathol Commun* **4**, 44 (2016).

52. B. Paumard-Hernandez, O. Calvete, L. Inglada Perez, H. Tejero, F. Al-Shahrour, G. Pita, A. Barroso, J. Carlos Trivino, M. Urioste, C. Valverde, E. Gonzalez Billalabeitia, V. Quiroga, J. Francisco Rodriguez Moreno, A. Fernandez Aramburo, C. Lopez, P. Maroto, J. Sastre, M. Jose Juan Fita, I. Duran, I. Lorenzo-Lorenzo, P. Iranzo, X. Garcia Del Muro, S. Ros, F. Zambrana, A. Maria Autran, J. Benitez, Whole exome sequencing identifies PLEC, EXO5 and DNAH7 as novel susceptibility genes in testicular cancer. *Int J Cancer* **143**, 1954-1962 (2018).

53. O. Barel, Z. Shorer, H. Flusser, R. Ofir, G. Narkis, G. Finer, H. Shalev, A. Nasasra, A. Saada, O. S. Birk, Mitochondrial complex III deficiency associated with a homozygous mutation in UQCRQ. *Am J Hum Genet* **82**, 1211-1216 (2008).

54. L. Lagae, What's new in: "genetics in childhood epilepsy". *Eur J Pediatr* **167**, 715-722 (2008).

55. J. Eckhaus, K. M. Lawrence, I. Helbig, M. Bui, L. Vadlamudi, J. L. Hopper, I. E. Scheffer, S. F. Berkovic, Genetics of febrile seizure subtypes and syndromes: a twin study. *Epilepsy Res* **105**, 103-109 (2013).

56. K. Munakata, M. Uemura, I. Takemasa, M. Ozaki, M. Konno, J. Nishimura, T. Hata, T. Mizushima, N. Haraguchi, S. Noura, M. Ikenaga, S. Okamura, M. Fukunaga, K. Murata, H. Yamamoto, Y. Doki, M. Mori, SCGB2A1 is a novel prognostic marker for colorectal cancer associated with chemoresistance and radioresistance. *Int J Oncol* **44**, 1521-1528 (2014).

57. F. Mauvais-Jarvis, N. Bairey Merz, P. J. Barnes, R. D. Brinton, J. J. Carrero, D. L. DeMeo, G. J. De Vries, C. N. Epperson, R. Govindan, S. L. Klein, A. Lonardo, P. M. Maki, L. D. McCullough, V. Regitz-Zagrosek, J. G. Regensteiner, J. B. Rubin, K. Sandberg, A. Suzuki, Sex and gender: modifiers of health, disease, and medicine. *Lancet* **396**, 565-582 (2020).

58. F. A. Wright, P. F. Sullivan, A. I. Brooks, F. Zou, W. Sun, K. Xia, V. Madar, R. Jansen, W. Chung, Y. H. Zhou, A. Abdellaoui, S. Batista, C. Butler, G. Chen, T. H. Chen, D. D'Ambrosio, P. Gallins, M. J. Ha, J. J. Hottenga, S. Huang, M. Kattenberg, J. Kochar, C. M. Middeldorp, A. Qu, A. Shabalin, J. Tischfield, L. Todd, J. Y. Tzeng, G. van Grootheest, J. M. Vink, Q. Wang, W. Wang, W. Wang, G. Willemsen, J. H. Smit, E. J. de Geus, Z. Yin, B. W. Penninx, D. I. Boomsma, Heritability and genomics of gene expression in peripheral blood. *Nat Genet* **46**, 430-437 (2014).

59. E. T. Stoeckli, Understanding axon guidance: are we nearly there yet? *Development* **145**, (2018).

60. B. Sutor, T. Hagerty, Involvement of gap junctions in the development of the neocortex. *Biochim Biophys Acta* **1719**, 59-68 (2005).

61. M. Caracausi, L. Vitale, M. C. Pelleri, A. Piovesan, S. Bruno, P. Strippoli, A quantitative transcriptome reference map of the normal human brain. *Neurogenetics* **15**, 267-287 (2014).

62. S. E. Lee, K. P. West, Jr., R. N. Cole, K. J. Schulze, L. S. Wu, J. D. Yager, J. Groopman, P. Christian, General intelligence is associated with subclinical inflammation in Nepalese children: A population-based plasma proteomics study. *Brain Behav Immun* **56**, 253-263 (2016).

63. M. D. Lakshmi Priya, A. Geetha, A biochemical study on the level of proteins and their percentage of nitration in the hair and nail of autistic children. *Clin Chim Acta* **412**, 1036-1042 (2011).

64. M. T. Joy, S. T. Carmichael, Encouraging an excitable brain state: mechanisms of brain repair in stroke. *Nat Rev Neurosci* **22**, 38-53 (2021).

65. J. E. Elias, S. P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214 (2007).

Author Contributions: KJSA designed the study and obtained grant funding, GT obtained hair samples, CRR performed the protein extractions, KS and RDL performed the proteomics experiments, KS, RDL, CRR and KJSA wrote initial drafts and edited the manuscript; KS, RDL, DD, MX and NA performed statistical analyses and created the figures; all authors reviewed and made critical revisions of the manuscript and approved the final version to be published.

Competing interests: The authors received no honoraria, grants, or other payments for writing this manuscript and they report no relevant financial relationships and no conflicts of interest.
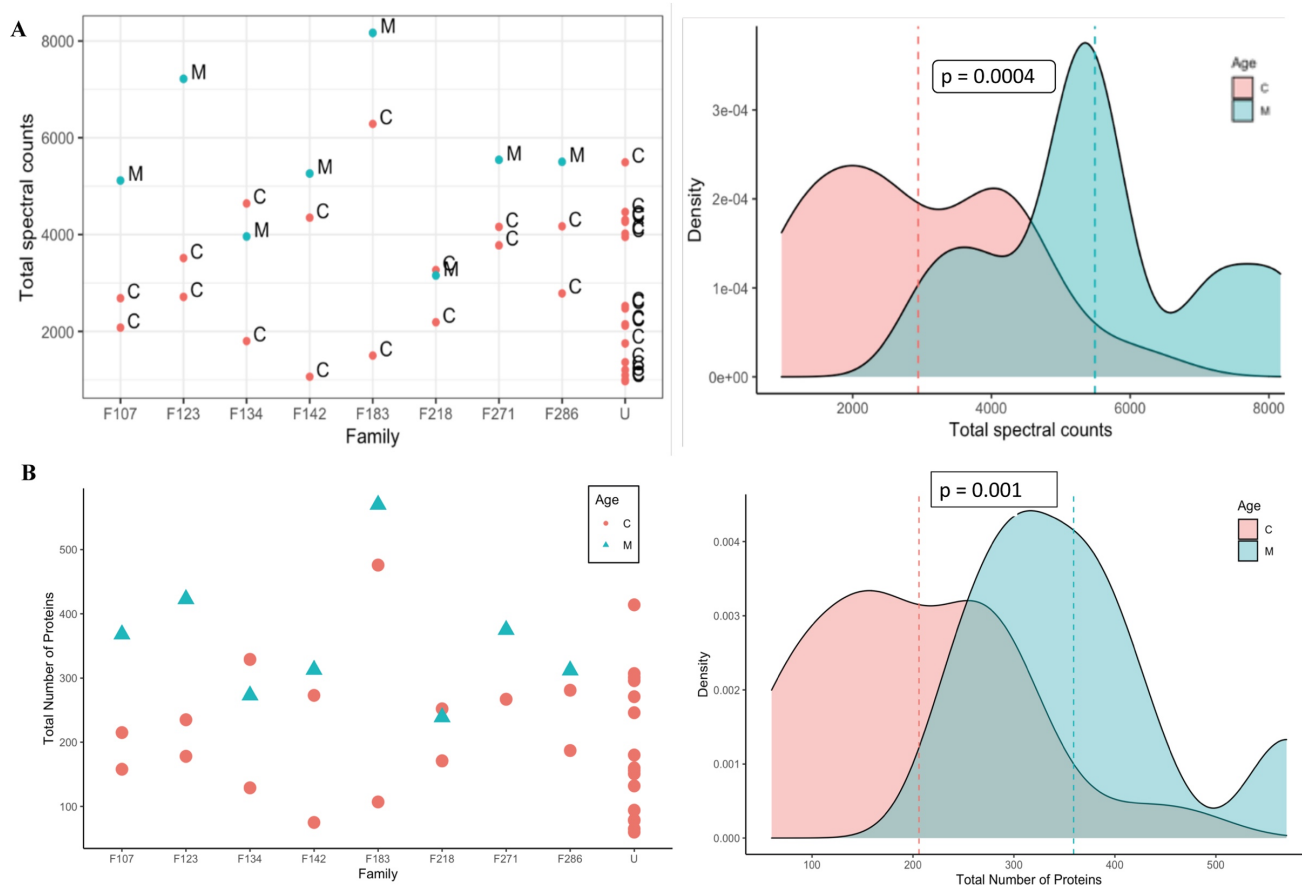
FIGURES

Figure 1: Hair proteins in mothers and children. (A) Protein spectral counts (p=0.0004) and **(B)** the numbers of proteins observed (with spectral counts >3) were consistently higher (p=0.001, Wilcoxon tests) in the mothers (M; Cyan) as compared to children (C; Pink). Mothers and their children (family labels: F107, F123, F134, F142, F183, F218, F271, F288) and Unrelated children (U) are identified on the X-axis: every mother except F134 and F218 had higher spectral counts and more hair proteins than her children.
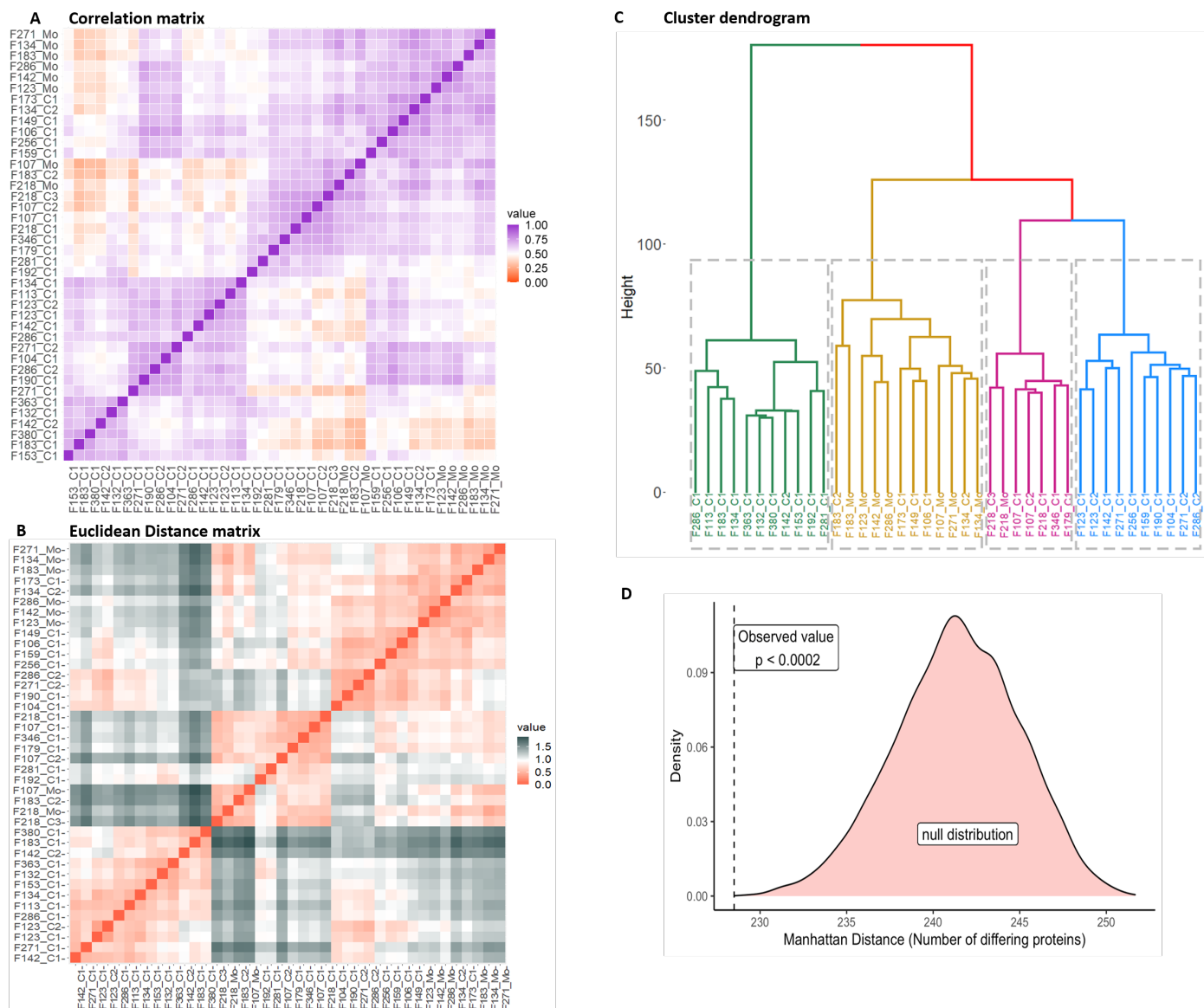
Figure 2: Similarities in hair protein profiles of individuals and families. **(A)** Spearman rank correlation matrix, with high (purple) to low (orange) correlation coefficients*; **(B)** Euclidean distances based on protein spectral counts, showing individuals more closely related (red) or more distant (grey) from each other*; **(C)** Hierarchical cluster dendrogram based on log spectral counts showing 7/8 mothers grouped in one cluster (mustard) with one mother in an adjacent cluster (pink); younger children (0-2 years) in one cluster (green) whereas older children dispersed in the other clusters*; **(D)** Intra-family Manhattan distances from Boolean hair protein profiles were shorter in mothers matched with their own children (p<0.0002) vs. 5000 simulated datasets created with mismatched mothers and children.

*Note: Individuals are listed on the X- and Y-axes with their family identifier, with Mo for mother, C1 for younger child, C2 for older children in each family.
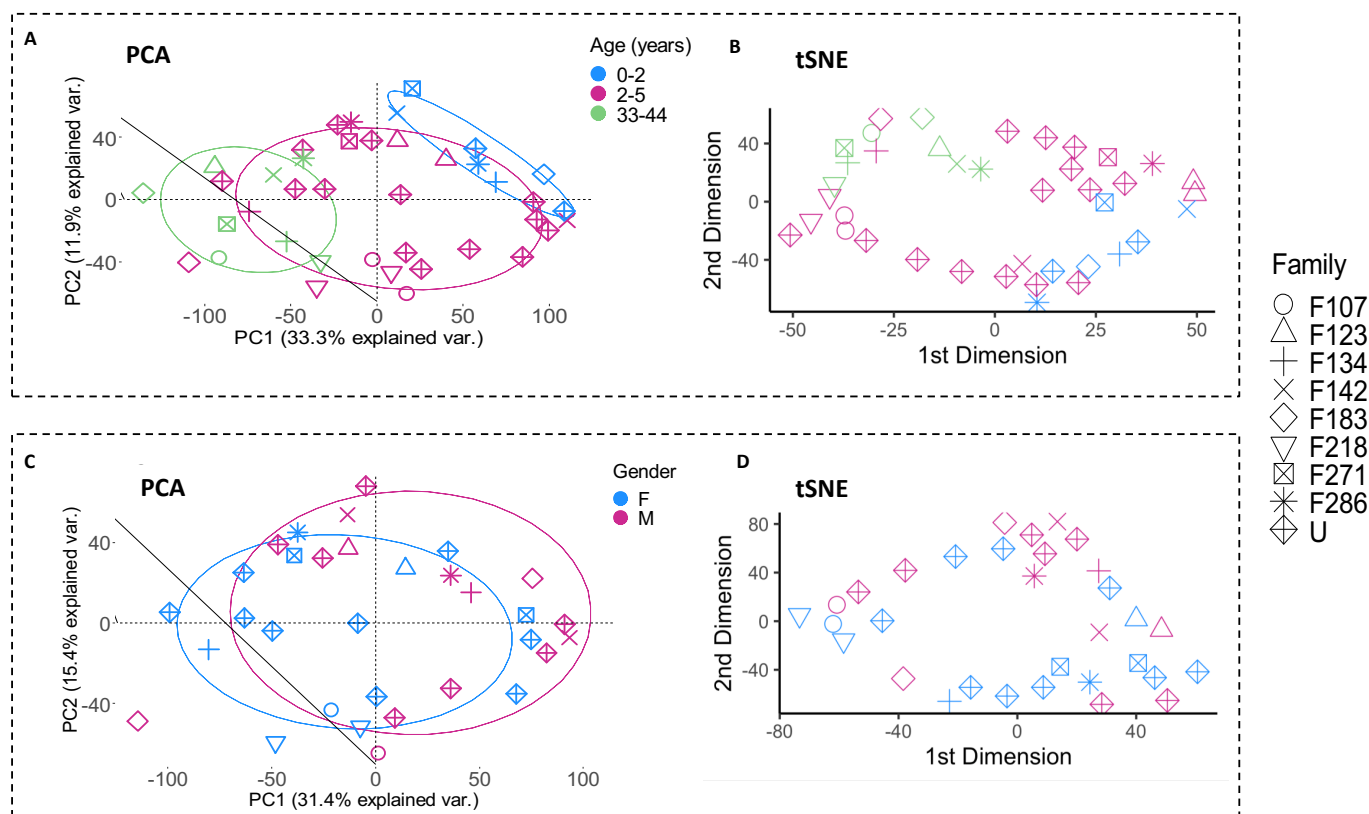
Figure 3: Age and sex-related differences in hair proteins. **(A)** First two principal components showing spatial separations by age, with children above 2 years (pink) located in between the children 0-2 years (blue, upper right) and the mothers (green, lower left). **(B)** The first two tSNE dimensions by age, showing mothers in the left upper quadrant separate from the children. Higher spectral counts for 7/17 hair proteins were observed in mothers (SERPINB4, POF1B, PLEC, A2ML1, HIST1H3A, UQCRQ, AHCY) and one protein (SCGB2A1) in children (Kruskal-Wallis and *post hoc* Benjamini-Hochberg corrections; see Table S2). **(C)** PCA analyses of all children showing overlapping circles for girls (blue) and boys (pink). **(D)** tSNE dimensions by sex, showing overlap between boys and girls. Higher spectral counts were observed only for CSN2 (Casein beta) in boys (p=0.0184) and ALMS1 (Alström syndrome protein 1) in girls (p=0.0214) (see Table S3).
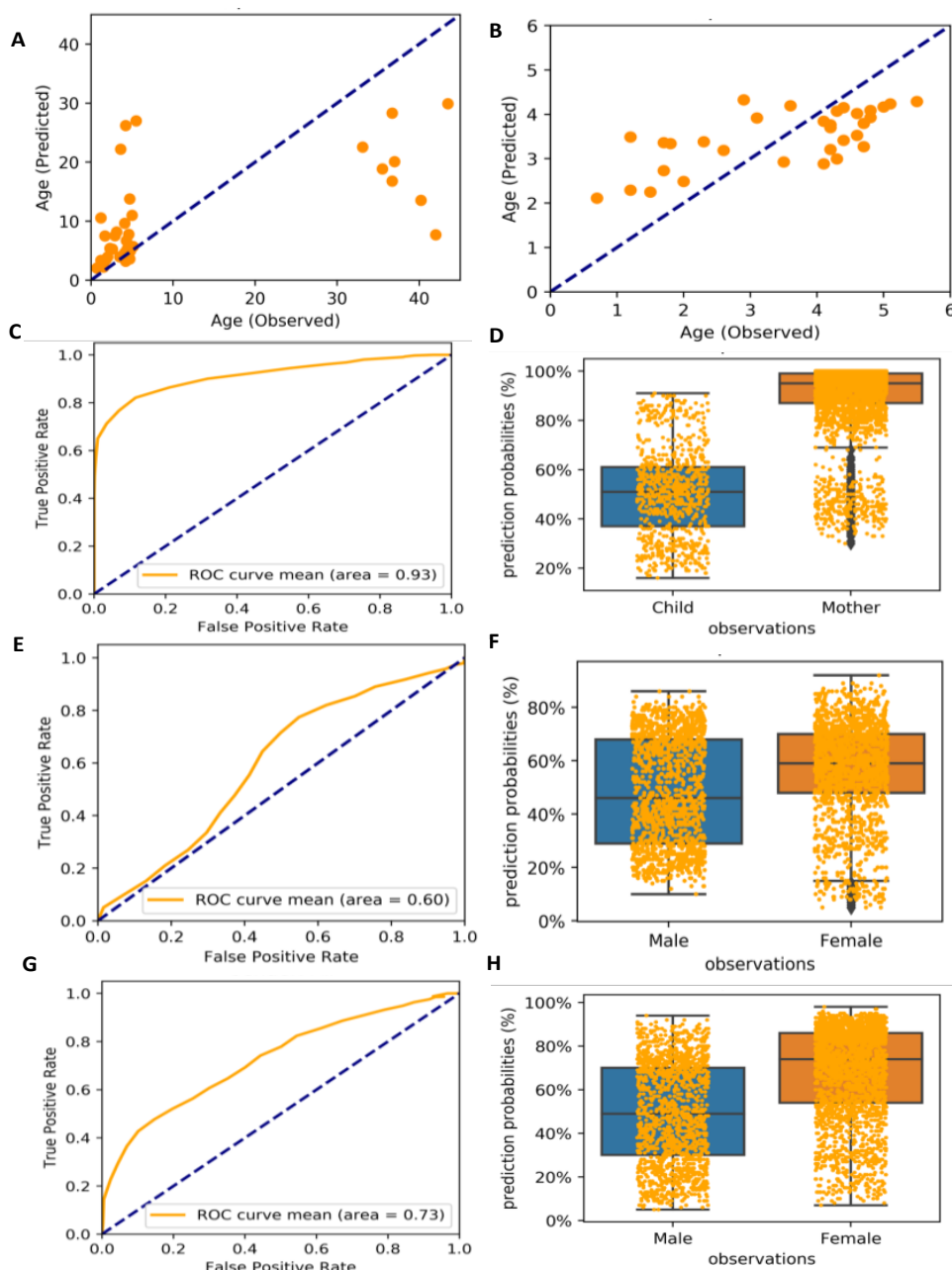
Figure 4: Machine learning algorithms predict age and sex from hair proteins. Mean scatterplot from 100 runs of Random Forest regression showing **(A)** observed vs. predicted age for mothers and children (R^2 0.37, p=0.00005) and **(B)** for all children (R^2 0.45, p=0.00004). **(C, D)** Random Forest plot showing mean accuracy for classifying mothers and children based on hair proteins (mean area under the ROC curve = 0.93, Wilcoxon test p=0.00011). **(E, F)** Random Forest plot showing mean accuracy for classifying by sex based on hair proteins for children (mean area under the ROC curve = 0.60, Wilcoxon test p=0.1703). **(G, H)** Random Forest plot improved when classifying all participants including mothers and children (area under the ROC curve = 0.73, Wilcoxon test p = 0.00831).
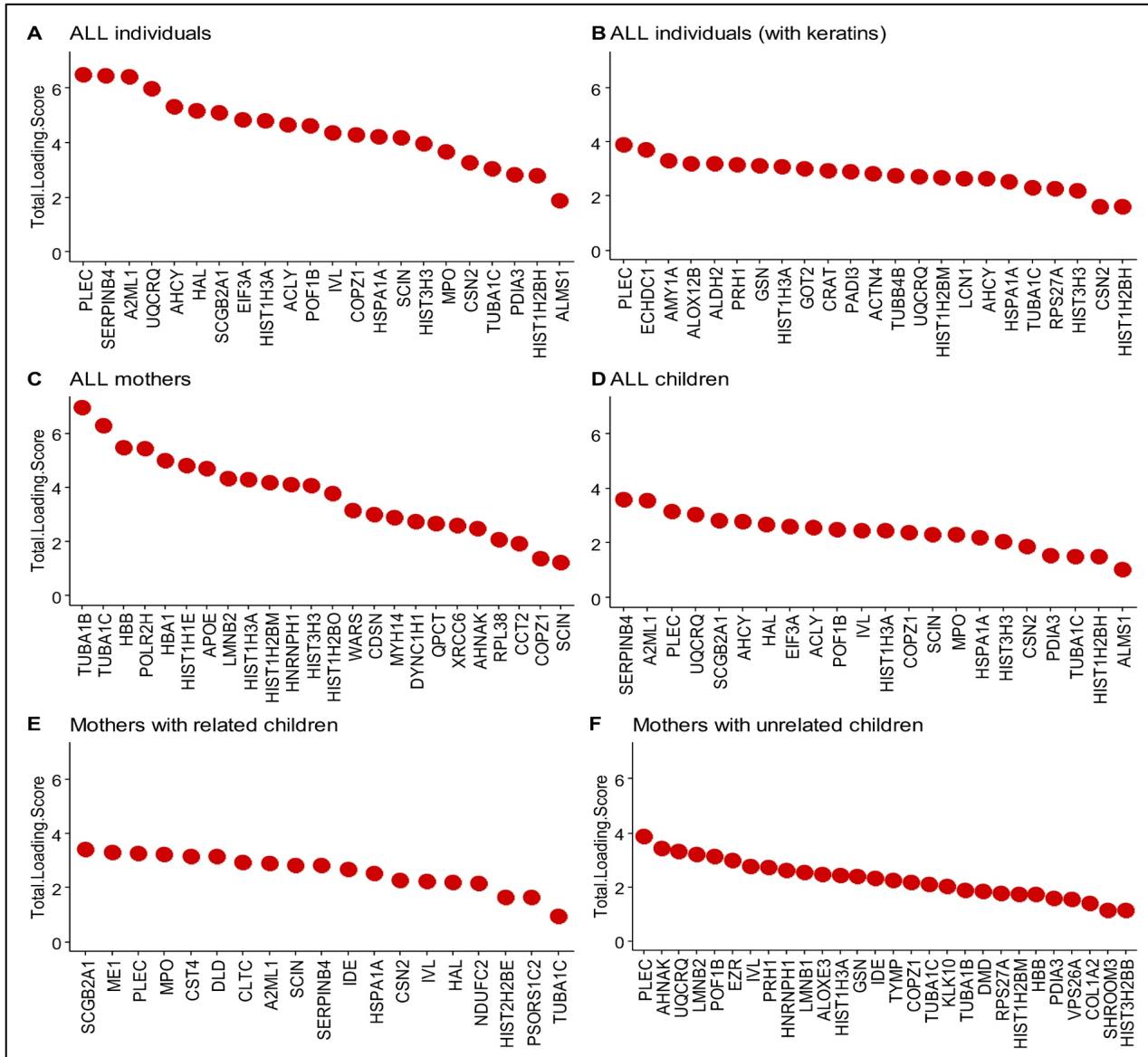
Figure 5: Top 5% proteins contributing to hair protein variability. The loading scores for each protein were weighted by the percent variance explained by each corresponding PC and then summed to give the Total Loading Score (TLS) for each protein. The top 5% proteins based on their TLS were identified as the most prominent contributors in each group. **(A)** All individuals (N=40, 49% of hair protein variability), **(B)** All individuals including keratins and KAPs (N=40, 64.3% variability); **(C)** All mothers (n=8, 89.5% variability); **(D)** All children (n=32, 57.5% variability); **(E)** Mothers (n=8) and their biological children (n=16) (49.3% variability), and **(F)** Mothers (n=8) and unrelated children (n=16) (64.6% variability).

Figure 6: Protein network for the top 5% hair proteins contributing to age-related differences between mothers and children. Some hair proteins had higher spectral counts in children (Orange) and others had higher spectral counts in mothers (Blue); continuous and interrupted lines show direct and indirect relationships, respectively. Mothers show higher protein spectral counts mostly for the 'enzymes' and 'peptidases' involved in cellular and metabolic processes, while proteins with higher spectral counts in children belong to 'other' group involved in growth and biological regulation.
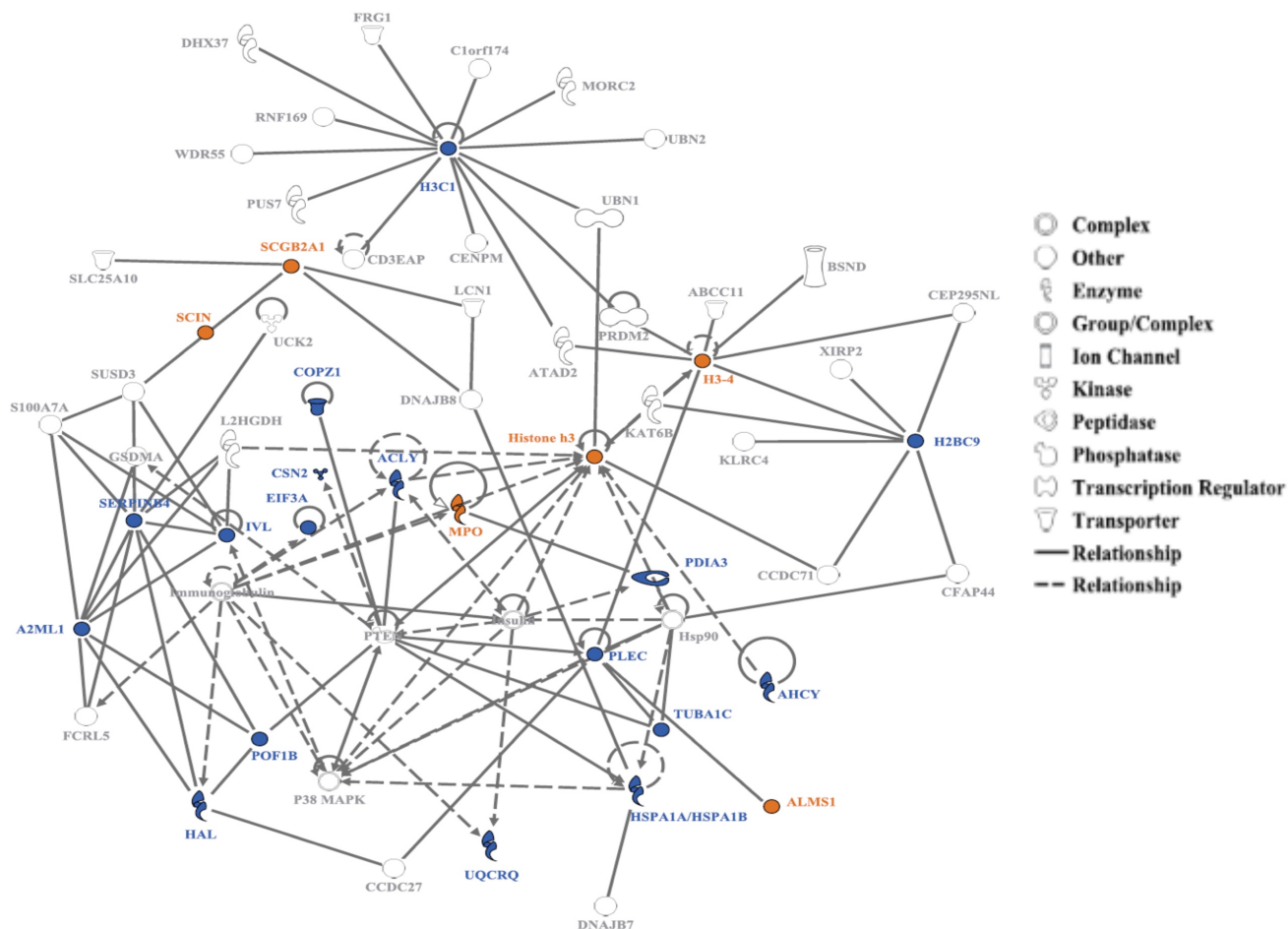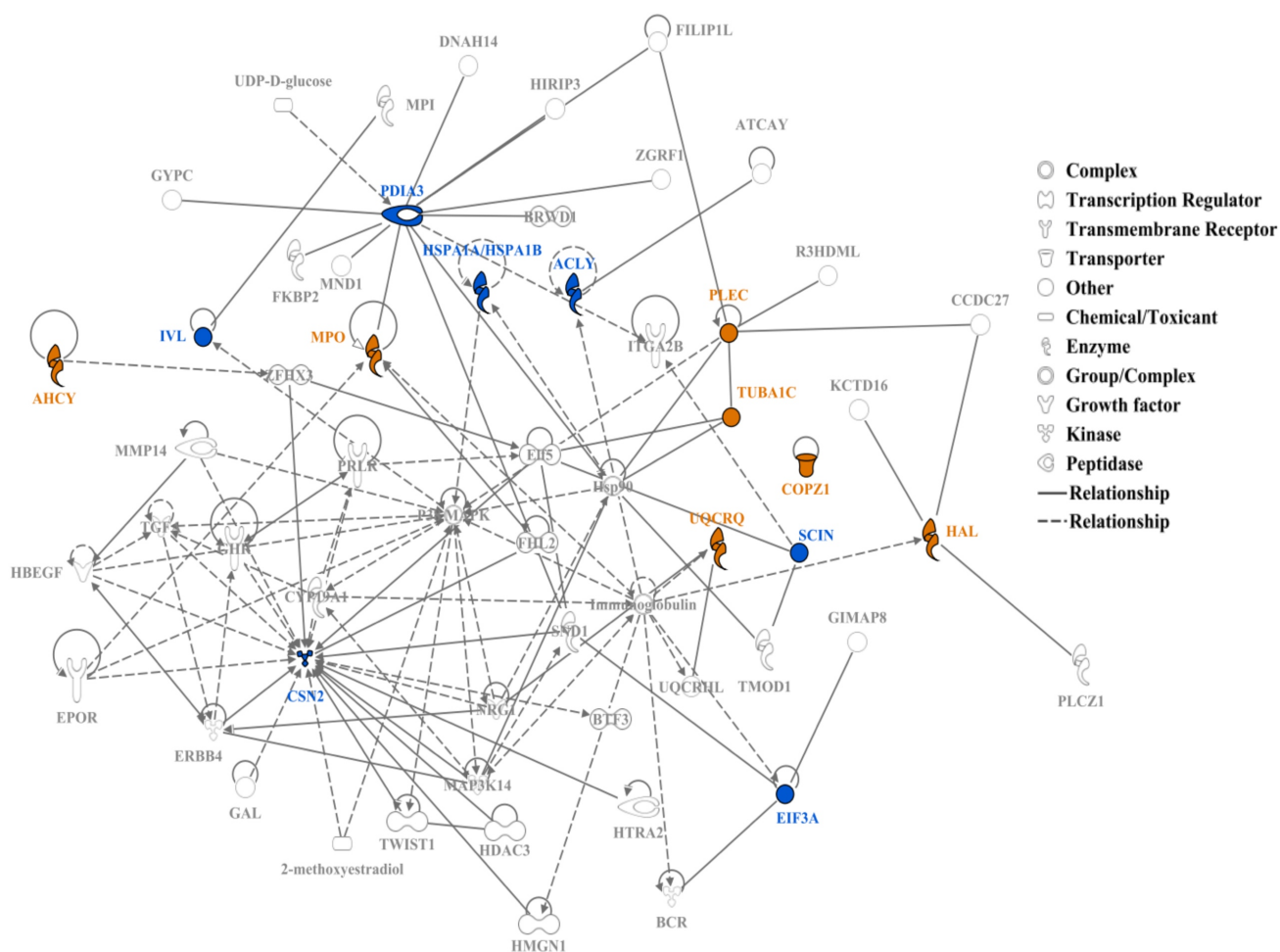
Figure 7: Protein network for top 5% hair proteins contributing to sex differences between boys and girls. Some proteins had higher spectral counts in girls (Orange) and others had higher spectral counts in boys (Blue); continuous lines show direct relationships and interrupted lines denote indirect relationships. Girls show higher protein spectral counts mostly for 'enzymes' or 'transporters' associated with cellular localization and metabolic processes. Proteins with higher spectral count in boys are 'enzymes' like 'kinases' or 'peptidases' associated with cellular and metabolic processes and biological regulation.

Figure 8: Canonical pathways associated with biologically significant proteins from the top 5% variables in all individuals (n = 40) contributing to age- and sex-related differences were identified with Ingenuity Pathway Analysis. Most of these proteins are involved in cellular metabolism, immune responses, brain development, and regulation of stress-related pathways.

SUPPLEMENTARY MATERIALS

Table S1: Demographic Characteristics and Hair Protein Data for Mothers and Children
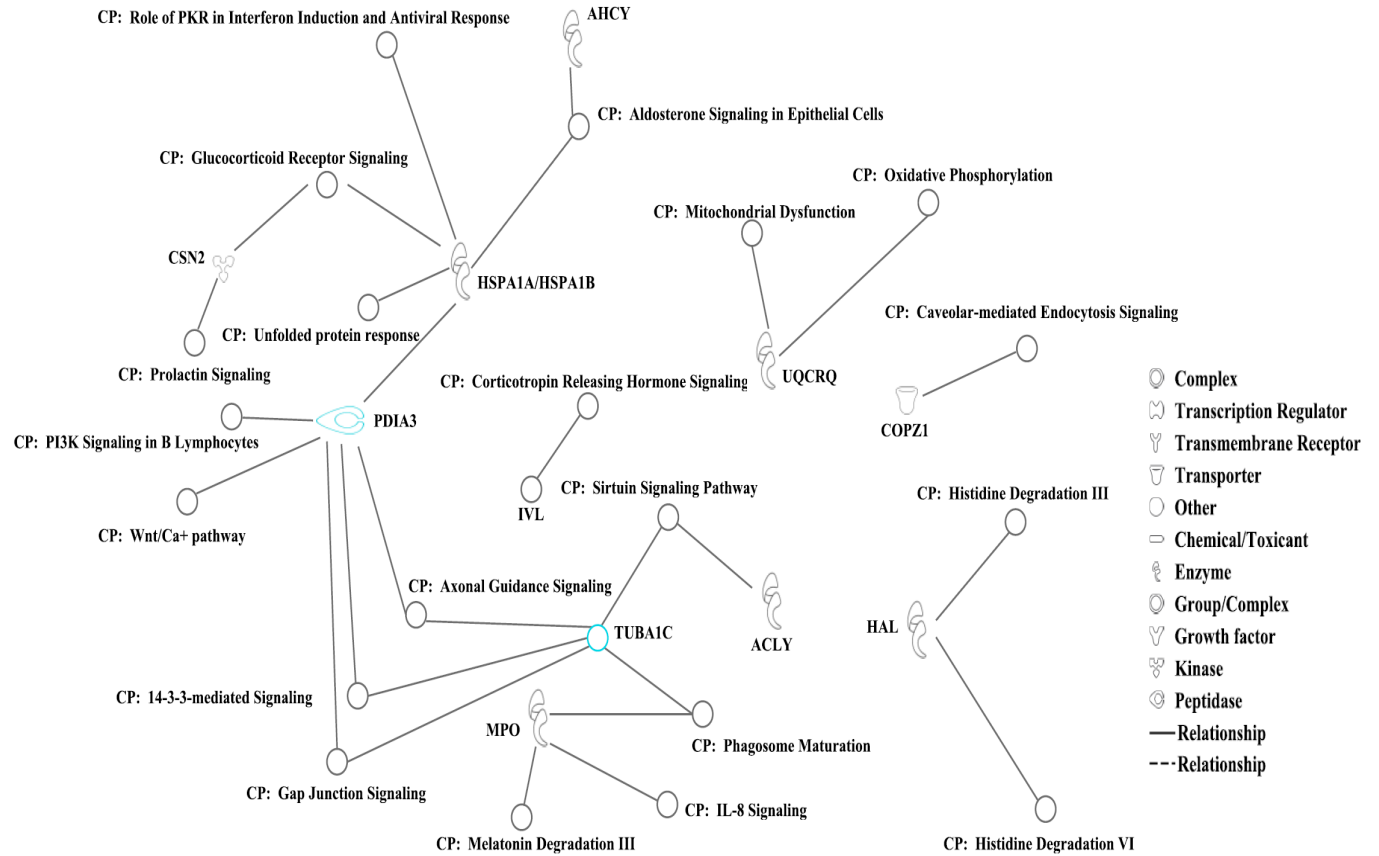
Family Code	Subject	Age (months)	Age (years)	Gender	Race	Ethnicity	# of hair proteins	Peptide Spectral Matches (PSMs)
F107	Mother	450.6	37.6	F	White	Non-Hispanic	819	6533
	Child1	27.6	2.3	F	White	Non-Hispanic	568	3949
	Child2	58.0	4.8	M	White	Other	464	2873
F123	Mother	447.1	37.3	F	White	Non-Hispanic	809	10370
	Child1	24.0	2	F	White	Non-Hispanic	499	3728
	Child2	52.4	4.4	M	White	Non-Hispanic	573	5078
F134	Mother	431.8	35.9	F	White	Missing	684	5445
	Child1	20.9	1.74	M	Mixed	Missing	387	2760
	Child2	67.6	5.6	F	Mixed	Missing	759	6872
F142	Mother	447.3	37.3	F	Asian	Missing	650	8370
	Child1	20.1	1.7	M	Mixed	Missing	581	6208
	Child2	50.6	4.2	M	Mixed	Missing	226	2353
F183	Mother	530.0	44.2	F	Asian	Non-Hispanic	1090	10527
	Child1	8.5	0.7	M	Asian	Non-Hispanic	314	2331
	Child2	44.0	3.7	M	Asian	Non-Hispanic	1010	8065
F218	Mother	504	42	F	White	Hispanic	609	4144
	Child1	35.2	2.9	F	White	Hispanic	631	4475
	Child2	58.5	4.9	F	White	Hispanic	524	3107
F271	Mother	402.6	33.6	F	White	Non-Hispanic	769	7525
	Child1	15.1	1.3	F	White	Non-Hispanic	557	7161
	Child2	42.5	3.5	F	White	Non-Hispanic	600	5615
F286	Mother	489.8	40.8	F	White	Non-Hispanic	616	9209
	Child1	22.0	1.8	M	White	Non-Hispanic	403	4727
	Child2	52.5	4.4	F	White	Non-Hispanic	614	6061
F346	Child	50.3	4.2	M	White	Missing	475	3429
F192	Child	38.1	3.2	F	Other	Hispanic	283	1731
F132	Child	51.4	4.3	F	White	Missing	272	1892
F363	Child	56.6	4.7	M	Mixed	Non-Hispanic	270	1914
F281	Child	53.5	4.5	M	Mixed	Mixed	406	3192
F173	Child	51.3	4.3	F	Other	Other	835	7168
F380	Child	14.8	1.2	M	Asian	Missing	237	1814
F159	Child	62.7	5.2	F	White	Non-Hispanic	485	3830
F179	Child	53.0	4.4	F	Asian	Other	494	2926
F149	Child	61.3	5.1	F	Mixed	Non-Hispanic	698	5733
F106	Child	56.7	4.7	F	Asian	Non-Hispanic	668	6390

F153	Child	57.1	4.8	M	Asian	Missing	275	2549
F256	Child	55.8	4.7	M	Mixed	Mixed	638	7016
F190	Child	31.5	2.6	F	Asian	Other	527	7460
F104	Child	50.2	4.2	M	White	Non-Hispanic	672	8084
F113	Child	17.9	1.5	F	White	Non-Hispanic	441	3640

Demographic data, total number of proteins, and peptide spectral matches observed in all 40 individuals. Mothers' hair (n=8) had significantly higher number of proteins (p=0.001) and protein spectral matches (p=0.0004) compared to children's hair (n=32). Related children (n=16, Cyan) are grouped with their mothers and unrelated children are listed below (n=16, Orange).

Table S2: Hair Proteins mediating differences between Mothers and Children

Entrez Gene Name	Gene Symbol: human	Expr Log Ratio	P-value	Location	Type(s)
Involucrin	IVL	-2.85	0.0576	Cytoplasm	other
Serpin family B member 4	SERPINB4	-2.452	0.0009***	Cytoplasm	other
POF1B actin binding protein	POF1B	-2.097	0.0151*	Plasma Membrane	other
Plectin	PLEC	-1.886	0.0004***	Cytoplasm	other
Alpha-2-macroglobulin like 1	A2ML1	-1.858	0.0042**	Cytoplasm	other
H3 clustered histone 1	HIST1H3A	-1.743	0.0038**	Nucleus	other
Ubiquinol-cytochrome c reductase complex III subunit VII	UQCRQ	-1.716	0.0007***	Cytoplasm	enzyme
Adenosylhomocysteinase	AHCY	-1.472	0.0040**	Cytoplasm	enzyme
Heat shock protein family A (Hsp70) member 1A	HSPA1A	-1.35	0.0569	Cytoplasm	enzyme
H2B clustered histone 9	HIST1H2BH	-1.17	0.5070	Nucleus	other
Histidine ammonia-lyase	HAL	-1.087	0.0851	Cytoplasm	enzyme
COPI coat complex subunit zeta 1	COPZ1	-0.931	0.158	Cytoplasm	transporter
Eukaryotic translation initiation factor 3 subunit A	EIF3A	-0.8	0.0567	Cytoplasm	other
Tubulin alpha 1c	TUBA1C	-0.526	0.262	Cytoplasm	other
Casein beta	CSN2	-0.269	0.491	Extracellular Space	kinase
ATP citrate lyase	ACLY	-0.249	0.0954	Cytoplasm	enzyme
Protein disulfide isomerase family A member 3	PDIA3	-0.051	0.884	Cytoplasm	peptidase
Scinderin	SCIN	0.028	0.221	Cytoplasm	other
Alström syndrome protein 1, centrosome and basal body associated protein	ALMS1	0.18	0.572	Cytoplasm	other
Histone H3.4	HIST3H3	0.64	0.153	Nucleus	other
Myeloperoxidase	MPO	0.925	0.886	Cytoplasm	enzyme
Secretoglobin family 2A member 1	SCGB2A1	5.32	0.0008***	Extracellular Space	other

Note: Mothers showed higher spectral counts than children for 7/17 hair proteins (shaded blue), although children had higher spectral counts for SCGB2A1 (shaded orange). Of these, SCGB2A1 showed the most prominent results, with >5-fold differences from the mothers. Significance was based on Kruskal-Wallis ANOVA with *post hoc* Benjamini Hochberg corrections for multiple comparisons (*p-value ≤ 0.05, **p-value ≤ 0.01, ***p-value ≤ 0.001)

Table S3: Hair proteins mediating differences between Boys and Girls

Entrez Gene Name	Gene Symbol: human	Expr Log Ratio	P-value	Location	Type(s)
Casein beta	CSN2	-3.046	0.0184*	Extracellular Space	kinase
Serpin family B member 4	SERPINB4	-1.303	0.391	Cytoplasm	other
Secretoglobin family 2A member 1	SCGB2A1	-1.036	0.0513	Extracellular Space	other
Protein disulfide isomerase family A member 3	PDIA3	-0.78	0.662	Cytoplasm	peptidase
ATP citrate lyase	ACLY	-0.531	0.585	Cytoplasm	enzyme
Myeloperoxidase	MPO	-0.493	0.581	Cytoplasm	enzyme
Involucrin	IVL	-0.476	0.804	Cytoplasm	other
Eukaryotic translation initiation factor 3 subunit A	EIF3A	-0.295	0.226	Cytoplasm	other
Alpha-2-macroglobulin like 1	A2ML1	-0.254	0.923	Cytoplasm	other
Scinderin	SCIN	-0.187	0.375	Cytoplasm	other
Heat shock protein family A (Hsp70) member 1A	HSPA1A/HSPA1B	-0.122	0.573	Cytoplasm	enzyme
POF1B actin binding protein	POF1B	0.094	0.875	Plasma Membrane	other
Histone H3.4	H3-4	0.139	0.938	Nucleus	other
Histidine ammonia-lyase	HAL	0.175	0.522	Cytoplasm	enzyme
COPI coat complex subunit zeta 1	COPZ1	0.225	0.536	Cytoplasm	transporter
Tubulin alpha 1c	TUBA1C	0.245	0.314	Cytoplasm	other
H3 clustered histone 1	H3C1	0.249	0.562	Nucleus	other
Adenosylhomocysteinase	AHCY	0.333	0.202	Cytoplasm	enzyme
Plectin	PLEC	0.441	0.256	Cytoplasm	other
Ubiquinol-cytochrome c reductase complex III subunit VII	UQCRQ	1.415	0.0976	Cytoplasm	enzyme
H2B clustered histone 9	H2BC9	1.423	0.221	Nucleus	other
Alström syndrome protein 1, centrosome and basal body associated protein	ALMS1	1.754	0.0214*	Cytoplasm	other

Note: Girls showed higher spectral counts than boys for several proteins (shaded orange), whereas boys had higher spectral counts for other proteins (shaded blue). CSN2 was significantly higher in boys, whereas ALMS1 was significantly higher in girls. Significance was based on Kruskal-Wallis ANOVA with *post hoc* Benjamini Hochberg corrections for multiple comparisons (*p-value ≤ 0.05).
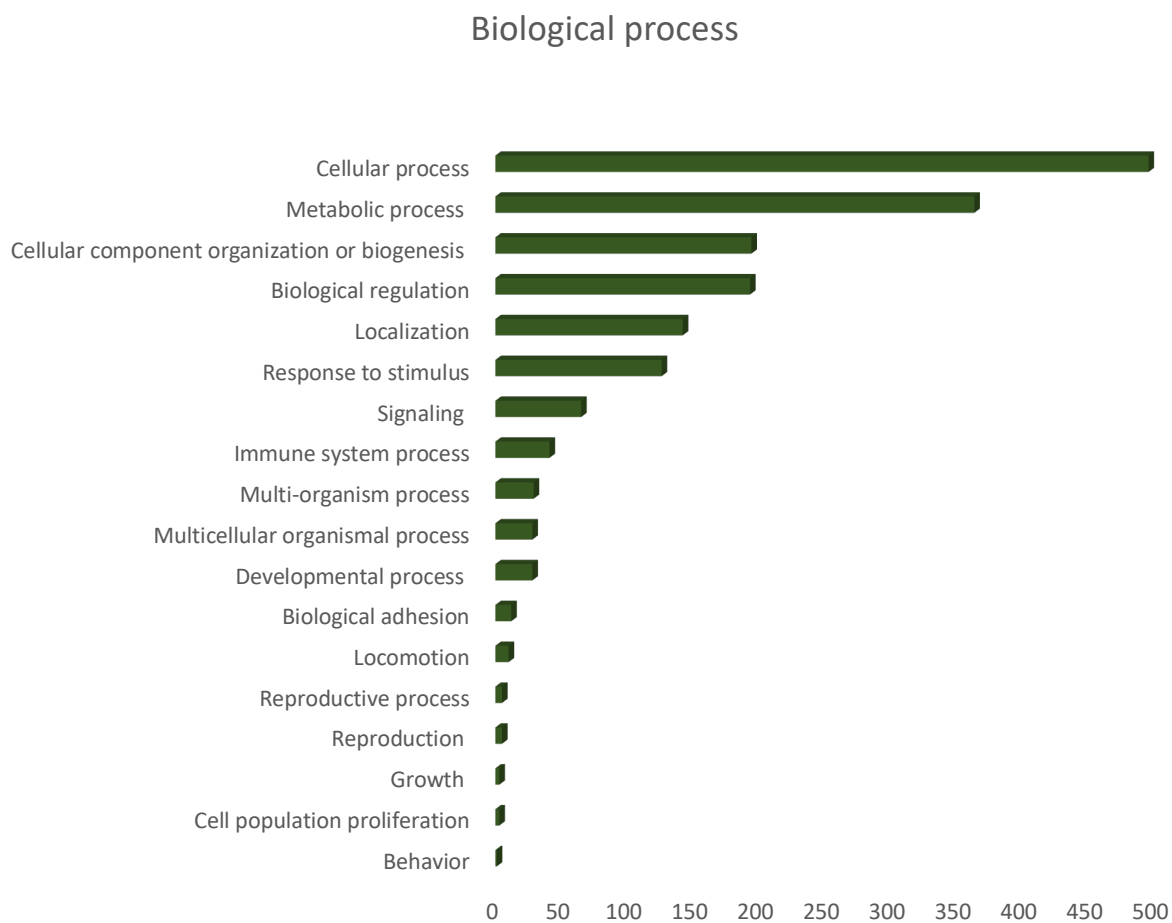
Biological process

Figure S1: Number of proteins associated with specific biological processes identified by PANTHER classification system. Most of the proteins are associated with cellular functions and metabolic processes.
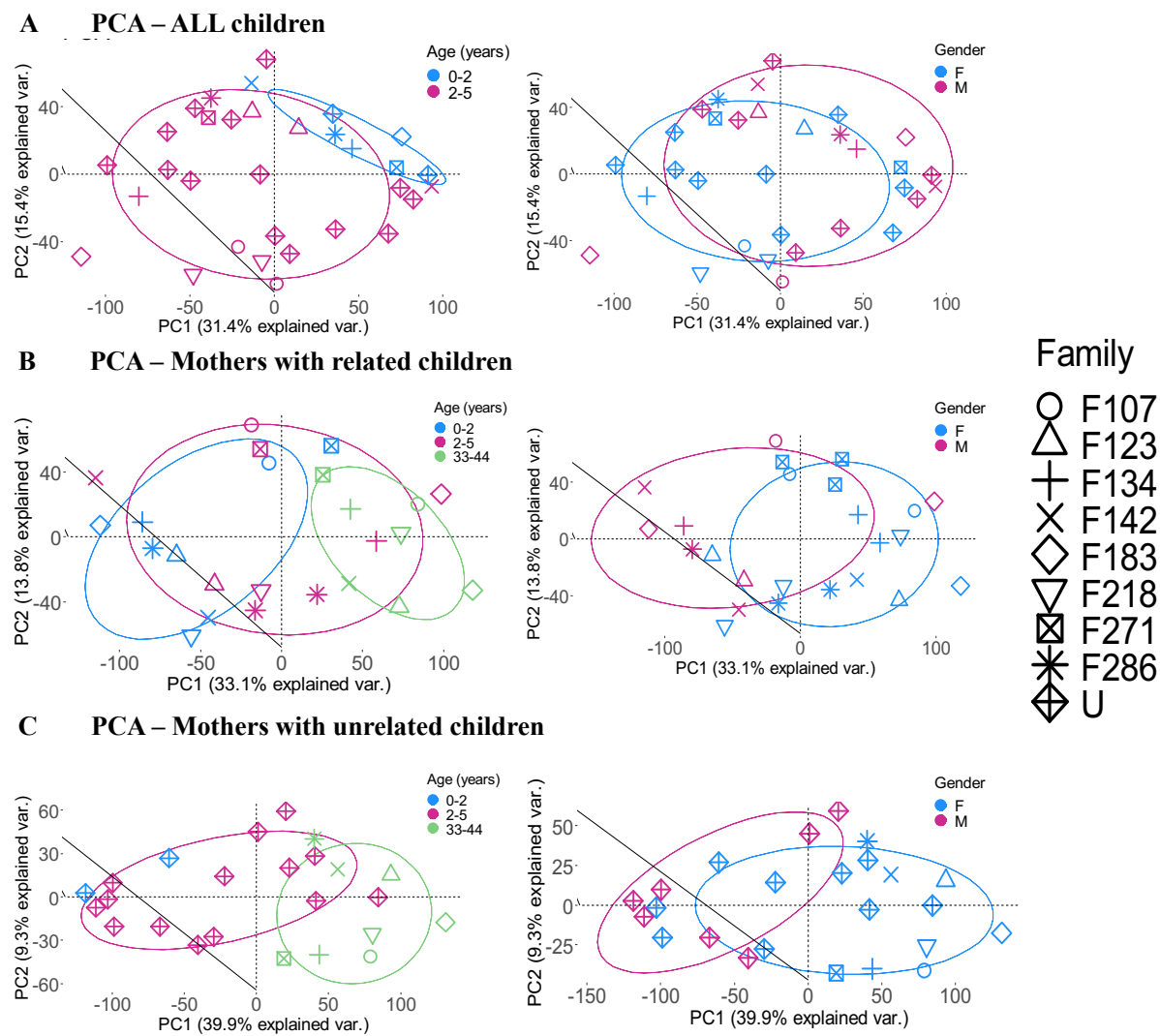
Figure S2: Spatial plots of individuals based on PCA analyses for different population combinations by Age (left side) and Sex (right side). The younger children (aged 0-2) could be separated out from older populations in age-based analyses for all 3 populations.