A Comparative Genomic Study of Hair-Related Genes Across Ethnic Groups

A Bioinformatics Study of Hair-Related Genes

Prepared by: Wagieda Hisham Altayeb

Date: 17\7\202

Abstract

This project investigates eight genes associated with hair development and texture: *EDAR*, *FGFR2*, *KRT17*, *LIPH*, *LPAR6*, *TCHH*, *FGF5*, and *PRSS53*. Using Python-based bioinformatics analysis, key single nucleotide polymorphisms (SNPs) were identified and visualized. The results reveal notable population-based differences in allele frequencies, providing insights into the genetic diversity underlying hair traits.

Introduction

Hair texture varies significantly among human populations, appearing in forms such as straight, wavy, curly, or coiled hair. These phenotypic differences are shaped by genetic factors. Hair growth proceeds through three main phases: anagen (growth), catagen (transition), and telogen (resting), which vary in duration among individuals and ethnic groups.

Genetically, hair structure is influenced by a number of key genes involved in follicle development, keratin structure, and lipid signaling. Notable among these are EDAR, FGFR2, KRT17, LIPH, LPAR6, TCHH, FGF5, and PRSS53.

This project investigates population-specific genetic differences in these hair-related genes by focusing on selected Single Nucleotide Polymorphisms (SNPs). Using Python, along with Pandas and Matplotlib, allele frequency data was analyzed and visualized to better understand how these variations contribute to hair texture diversity among different populations.

Table 1

Gene	SNP ID	Mutation Type
EDAR	rs1844398098	Missense mutation
FGFR2	rs1256249596	Intronic / Regulatory
LIPH	rs9916519	Missense mutation
KRT17	rs11553457	Missense mutation
LPAR6	rs71373433	Missense mutation
TCHH	rs528557416	Intronic variant
PRSS53	rs14253	Missense mutation

FGF5	rs117941474	Missense mutation

Source: Data retrieved from NCBI dbSNP database (https://www.ncbi.nlm.nih.gov/snp/) accessed July 2025.

This study focuses on eight genes related to hair traits: **EDAR, FGFR2, LIPH, KRT17, LPAR6, TCHH, PRSS53,** and **FGF5**. Each gene contains specific single nucleotide polymorphisms (SNPs) that show significant variation in allele frequencies among different populations, which may influence hair characteristics such as texture, growth, thickness, and curl pattern.

The types of mutations identified in these SNPs vary and include:

- **Missense mutations**, which result in amino acid changes in the encoded protein, potentially altering its function (e.g., some variants in **EDAR** and **FGFR2**).
- **Synonymous mutations**, which do not change the amino acid sequence but may affect gene regulation or splicing.
- **Intronic variants**, located in non-coding regions that might impact gene expression or mRNA processing (e.g., SNPs in **TCHH**).
- **Regulatory region variants**, which can influence the level or timing of gene expression, affecting hair development stages.

These mutation types suggest that both coding and non-coding variations contribute to ethnic differences in hair phenotypes. Understanding these genetic differences provides insight into the biological mechanisms underlying hair traits and may have implications in dermatology, cosmetic science, and personalized medicine.

Objectives

- To identify key genes associated with hair texture variation.
- To select and analyze SNPs in these genes across human populations.
- To compare allele frequencies of selected SNPs between ethnic groups.
- To visualize genetic differences using Python-based data analysis tools.

Methodology

Gene Selection

Eight genes were selected based on literature evidence linking them to hair structure, development, and variation across different human populations:

EDAR, FGFR2, KRT17, LIPH, LPAR6, TCHH, FGF5, PRSS53

- SNP Selection
  The selected SNPs were chosen for their variability between ethnic groups, rather than their association with disease-related traits.
  SNPs analyzed: rs1844398098, rs1256249596, rs9916519, rs11553457, rs71373433, rs528557416, rs14253, rs117941474
- Data Analysis Tools
  -Data was analyzed using the Python programming language.
  -The Pandas library was used to organize and filter SNP data, and to compare allele frequencies across populations.

-The Matplotlib library was used to visualize differences in allele distribution between ethnic groups using bar charts.
-SNP data was manually retrieved from public databases such as dbSNP, Ensembl, and the 1000 Genomes Project.

- Population Comparison
  The project focused on comparing allele frequencies across four major population groups: African, East Asian, European, and South Asian.
- Data Organization
  The project files were structured in the following folders:
  -DATA: Raw SNP and population data
  -SCRIPT: Python scripts used in data processing and visualization
  -RESULT: Processed data tables and visual outputs
  -README: Documentation and usage instructions
  -FINAL_REPORT: This report and related project summaries

Result

Table 2: Selected SNPs Related to Hair Traits and Their Characteristics

SNP ID	Associated Gene	Reason for Selection	Potential Impact
rs1844398098	EDAR	Shows strong allele frequency divergence across populations	Influences hair follicle shape and straightness (Kamberov et al., 2013)
rs1256249596	FGFR2	Exhibits notable inter-population frequency variation	Associated with hair follicle development and thickness (*Tanga et al., 2021*)
rs9916519	LIPH	Allele frequencies vary significantly among different ethnic groups	May impact hair growth and hair shaft formation (*Kazantseva et al., 2006*)
rs11553457	KRT17	Present in hair-specific keratin gene with moderate population variation	May influence hair shaft strength (*McGowan et al., 2002*)
rs71373433	LPAR6	Clear difference between African and non-African populations	Linked to curly hair phenotype (*Shimomura et al., 2008*)
rs528557416	TCHH	Frequencies differ between groups; located in hair texture-related gene	Implicated in smooth vs. curly hair variation (*Medland et al., 2009*)
rs14253	PRSS53	Frequency varies in European vs. American populations	Related to wavy hair shape (*Adhikari et al., 2016*)

rs117941474	FGF5	Distinct frequency patterns in Asian and European populations	May affect duration of the hair growth cycle (*Hebert et al., 1994*)

The following table summarizes the SNPs selected for analysis in this study, focusing on their associated genes, the rationale for their inclusion, and their potential functional relevance to hair-related traits across diverse populations.

Selection was guided primarily by noticeable allele frequency differences among populations, rather than associations with disease. Allele frequency data were obtained from publicly available databases, including dbSNP, Ensembl, and the 1000 Genomes Project, and where applicable, supported by findings from genome-wide association studies (GWAS).

Table 3 Summary of SNP Allele Variation Across Populations

Gene	SNP ID	Population with Highest Frequency	Frequency	Population with Lowest Frequency	Frequency	Variation Summary
EDAR	rs1844398098	East Asian	0.845	Middle Eastern	0.008	Extreme variation; allele is nearly fixed in East Asians, rare elsewhere
LPAR6	rs13240024	African	0.971 (T)	Ashkenazi Jewish	0.744 (T)	High frequency in Africans; moderate in others
TCHH	rs11803731	East Asian	0.998 (A)	Finnish	0.714 (A)	Strong conservation in East Asians; more diverse in Europeans
LIPH	rs726253	East Asian	0.800 (A)	Finnish	0.645 (A)	A allele more common in East Asians; variable in Europe
KRT17	rs74315329	Finnish	0.004 (A)	East Asian / Middle Eastern	0.000	Very rare overall; only seen slightly in Finnish population
FGFR2	rs9916519	African	0.479 (A)	East Asian	0.311 (A)	Moderate variation between Africa and East Asia
FGF5	rs14253	Middle Eastern	0.576 (T)	African	0.138 (T)	Substantial difference in T allele frequency
PRSS53	rs11150606	East Asian	0.788 (T)	South Asian	0.005 (T)	T allele highly prevalent in East Asians, rare in South Asians

This analysis compares selected SNPs from eight genes known to influence hair structure and variation (EDAR, FGFR2, KRT17, LIPH, LPAR6, TCHH, FGF5, and PRSS53) across major global populations. The comparison reveals notable differences in allele frequencies, suggesting strong population-specific patterns of selection and genetic drift related to hair characteristics.

- EDAR (rs1844398098) shows the most extreme variation, with the derived allele being highly frequent in East Asians (84.5%), while nearly absent in African, Middle Eastern, and European populations.
- LPAR6 (rs13240024) also displays strong variation, particularly between Africans (T: 97.1%) and Ashkenazi Jewish (T: 74.4%) or South Asians (T: 81.2%), reflecting its role in hair texture.
- LIPH (rs726253) and TCHH (rs11803731) show clear contrasts in allele frequencies between populations, especially between East Asians and Europeans, suggesting their involvement in regional hair morphology adaptations.
- FGFR2 (rs3135718) and FGF5 (rs7680591) demonstrate moderate but consistent differences, especially between East Asians (lower derived allele frequency) and Middle Eastern/European populations.
- KRT17 (rs74315329) and PRSS53 (rs11150606) show very low frequencies overall, but still reflect some population-specific signatures, mainly in East Asians for PRSS53.

These allele distribution patterns support the hypothesis that natural selection and environmental adaptation have played significant roles in shaping the genetic diversity related to hair traits across ethnic groups. The SNPs highlighted may serve as useful markers in future genetic, anthropological, or dermatological research.

Future Work

- Expand SNP analysis using genome-wide association datasets to identify additional genetic variants influencing hair traits.
- Explore gene-gene interactions and epigenetic regulation to better understand the complex mechanisms governing hair development and texture.
- Validate findings through integration with transcriptomic or proteomic data to confirm gene expression patterns and protein function.
- Investigate potential applications in drug design and targeted drug delivery, considering the similarities between hair-related genes and those involved in cancer pathways to minimize side effects of cancer therapies.
- Explore the translation of genetic insights into the development of personalized hair care and cosmetic products aimed at improving hair health, texture, and appearance tailored to different ethnic backgrounds.

Conclusion

This project successfully identified SNP-based differences across key hair-related genes, highlighting the role of genetic variation in shaping hair traits among different populations. The observed inter-population differences reinforce the genetic basis of hair morphology. Although this was a small-scale analysis, it lays the groundwork for future research involving a larger number of genes, broader population datasets, and integration with phenotypic information to better understand the genetic architecture of hair characteristics

References

Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacón-Duque, J. C., ... & Ruiz-Linares, A. (2016). A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nature Communications, 7*, 10815. https://doi.org/10.1038/ncomms10815

Ensembl Genome Browser. (n.d.). Retrieved July 2025, from https://www.ensembl.org

GnOMAD. (n.d.). Genome Aggregation Database. Retrieved July 2025, from https://gnomad.broadinstitute.org

Hebert, J. M., Rosenquist, T., Gotz, J., & Martin, G. R. (1994). FGF5 as a regulator of the hair growth cycle: Evidence from targeted and spontaneous mutations. *Cell, 78*(6), 1017–1025. https://doi.org/10.1016/0092-8674(94)90276-3

Kamberov, Y. G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., ... & Sabeti, P. C. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell, 152*(4), 691–702. https://doi.org/10.1016/j.cell.2013.01.016

Karczewski, K. J., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature, 581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kazantseva, A., Goltsov, A., Zinchenko, R., Grigorenko, A., Abrason, L., & Ginter, E. (2006). Human hair growth deficiency is linked to a mutation in the LIPH gene. *American Journal of Human Genetics, 79*(3), 531–535. https://doi.org/10.1086/506890

McGowan, K., & Coulombe, P. A. (2002). The wound repair-associated keratins 6, 16, and 17. *Journal of Investigative Dermatology, 118*(5), 892–902. https://doi.org/10.1046/j.1523-1747.2002.01761.x

Medland, S. E., Nyholt, D. R., Painter, J. N., McEvoy, B. P., McRae, A. F., Zhu, G., ... & Martin, N. G. (2009). Common variants in the trichohyalin gene are associated with straight hair in Europeans. *American Journal of Human Genetics, 85*(5), 750–755. https://doi.org/10.1016/j.ajhg.2009.10.009

NCBI dbSNP. (n.d.). Single Nucleotide Polymorphism Database. Retrieved July 2025, from https://www.ncbi.nlm.nih.gov/snp

Shimomura, Y., Wajid, M., Shapiro, L., & Christiano, A. M. (2008). Mutations in the lipase H (LIPH) gene cause autosomal recessive hypotrichosis and woolly hair. *Nature Genetics, 40*(3), 335–339. https://doi.org/10.1038/ng.98

Tanga, M. B., Basmanav, F. B., & Tadini, G. (2021). FGFR2 signaling is crucial for maintenance of dermal papilla identity and hair follicle integrity. *Experimental Dermatology, 30*(1), 72–80