"A Holistic Approach to Scalp and Hair Health: Discovery of Novel Genetic Markers and Lifestyle Factors"

Jihye Kim* [1] , Eunbi Ko[1] , Min Cho[1] , Soojung Ko[1] , Jiye kang[1] , Jiyei Choi[1] , Dajeong Lee[1] , Eunjoo Kim [1]

Affiliation 1; Amorepacific R&I Center, Seoul, Korea, South

1. Introduction

Scalp and hair health are influenced by a complex interplay of genetic, biological, and behavioral factors [1,2]. While lifestyle and environmental exposures can modulate scalp and hair conditions over time, underlying genetic predispositions are the primary determinants of conditions such as hair thinning, hair loss, graying, and scalp sensitivity[3,4].

Traditionally, scalp and haircare strategies have focused on addressing visible symptoms or self-reported concerns after the manifestation of problems [5]. However, such reactive approaches often fail to prevent the early deterioration of scalp and hair health. In contrast, leveraging genetic information offers a unique opportunity to predict vulnerability before clinical symptoms emerge, enabling preventive and personalized interventions tailored to each individual's inherent risk profile [6].

Current beauty industry scalp and haircare solutions are predominantly designed for the general population, with limited consideration of individual genetic variability. Moreover, consumer segmentation is often based on subjective self-assessment (e.g., perceived hair loss severity) or temporary external factors (e.g., seasonal dryness), rather than stable biological indicators [7,8]. There are few cases of utilizing genetic information to provide personalized scalp and hair care solutions [9]. This paper is novel in that it applies a holistic approach, utilizing a wider range of genetic information including markers related to skin, nutrients, and exercise, as well as scalp and hair-specific genes.

In this study, we aimed to bridge this gap by adopting a holistic approach that integrates genetic, biological, and behavioral dimensions. We performed genotype-based clustering of Korean individuals using scalp, hair, skin aging, and metabolism-related genetic markers. Our goal was to identify genetically distinct groups that exhibit differential scalp and hair health profiles — both objectively (e.g., hair density, scalp sensitivity) and subjectively (e.g., primary hair concerns, care behaviors). By characterizing these clusters, we sought to demonstrate that genetic information, rather than transient symptoms, can serve as a stable and predictive tool for segmenting individuals and guiding precision scalp and haircare solutions.

2. Materials and Methods

2.1. Study Population

Data were collected from 1,300 Korean individuals aged 20 to 60 years between July 2023 and October 2024. Among them, 706 participants who completed scalp assessments, lifestyle surveys, and genetic profiling were included in the final analysis. Skin assessment data, although collected, were not utilized in this study.

2.2. Scalp and Hair Assessments

Scalp and hair conditions were assessed using Aram Huvis (hair density and thickness evaluation). Data from six scalp regions were used: frontal hairline(fonthairline), right hairline, left hairline, mid-frontal scalp (frontcenter), vertex (center), and occipital scalp (back).

2.3. Self-Reported Questionnaire

Participants completed a structured survey comprising 16 items on scalp and hair health, covering:

- **Hair primary concerns**: hair loss, itchiness, dandruff, limp hair, dryness, graying.
- **Hair care behaviors**: shampoo timing (morning, before going out, nighttime), frequency, drying behavior.
- **Scalp sensitivity**: product reactions, environmental triggers (e.g., dryness, fine dust).

Most responses were rated on a 4- or 6-point ordinal scale.

2.4. Genetic Profiling

Genetic profiling was conducted using oral epithelial cell samples. DNA was extracted and genotyped by Lab Genomics (Seongnam, Korea) following Korean DTC guidelines. A total of 66 SNP markers related to scalp, hair, skin, metabolism, and lifestyle traits were analyzed. Genetic rank scores (1st = most favorable; 100th = attention-needed) were calculated relative to a Korean reference cohort.
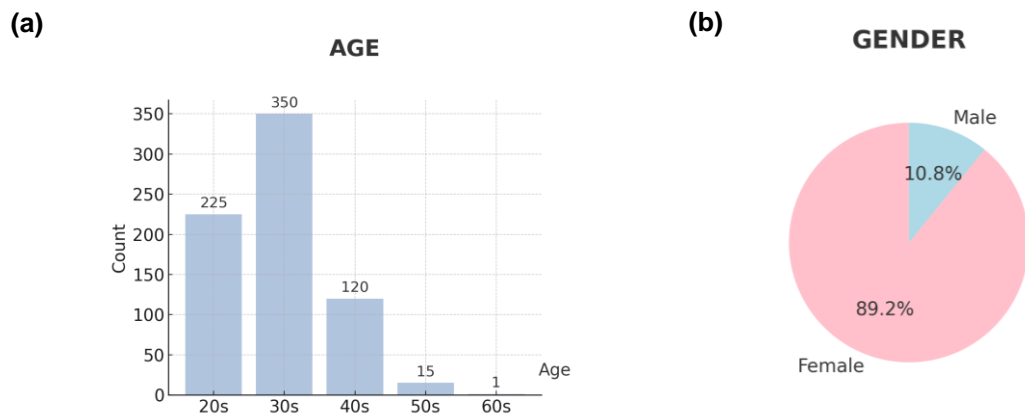
2.5. Genetic Clustering

K-means clustering was applied to standardized genetic rank scores using SPSS 20.0 (IBM Corp., Armonk, NY, USA). The optimal number of clusters was determined via the elbow method. Three distinct genetic clusters were established.

2.6. Statistical Analysis

Associations between clusters and outcomes were analyzed using: A p-value < 0.05 was considered statistically significant.

- One-way ANOVA (hair density, thickness) with Tukey's post-hoc tests
- Chi-square tests (scalp type, hair concern, shampoo behavior)
- Kruskal-Wallis tests (ordinal-scaled sensitivity and lifestyle variables)

3. Results

(a)

(b)

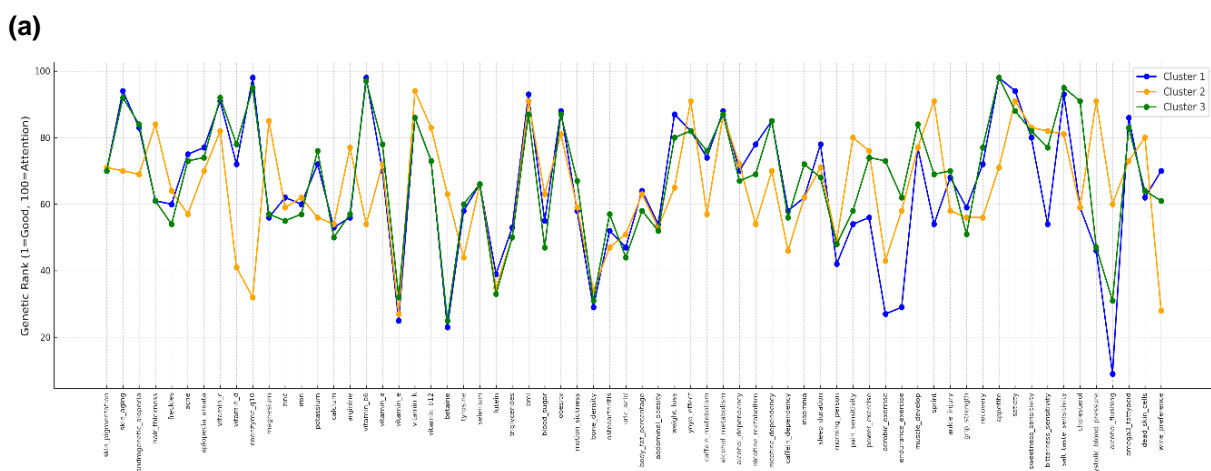**Figure 1. Demographic Characteristics of Participants.**

(a) Bar chart showing the distribution of participants across five age groups (20s–60s), with the majority in their 30s and 20s.
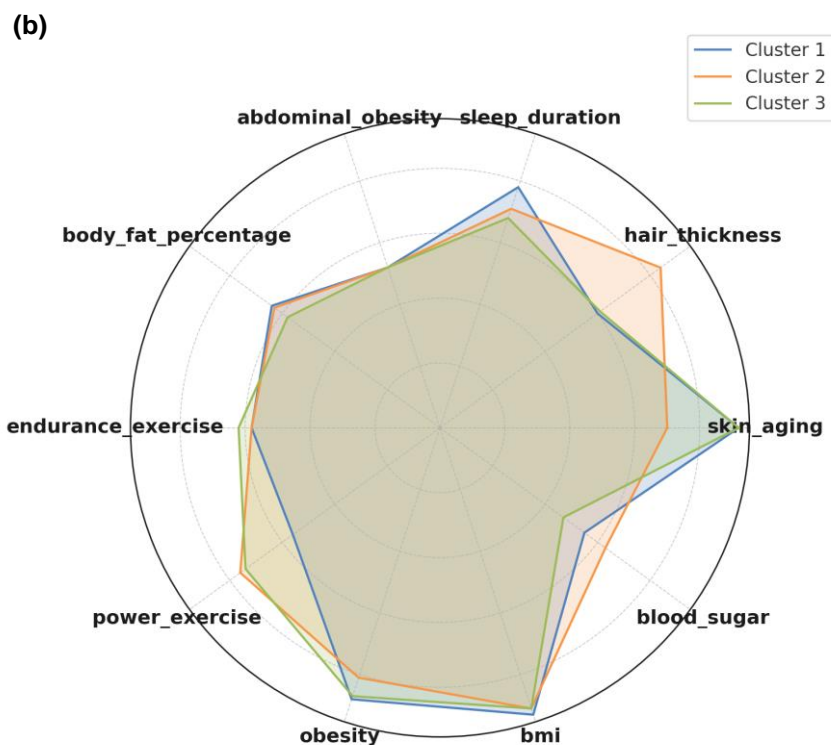**(b) Pie chart depicting gender distribution, with females representing approximately 89% of the sample. These demographic characteristics provide context for the subsequent clustering analysis based on genetic and lifestyle factors.**

A total of 706 Korean individuals aged 20 to 60 years were included in the analysis.
As shown in Figure 1a, the age distribution was skewed toward younger groups, with the majority of participants in their 20s (n ≈ 225) and 30s (n ≈ 350), followed by smaller numbers in their 40s (n ≈ 120), 50s, and 60s. Figure 1b shows that the cohort was predominantly female (89.1%), with only 10.9% male participants. Despite the skewed sex ratio, the large sample size and broad age range provided sufficient demographic diversity to support robust clustering based on genetic and lifestyle characteristics.

Genetic K-means clustering was applied to 66 SNP-derived rank scores representing percentile-based genetic risk or advantage (1 = favorable, 100 = attention-needed) across domains such as scalp/hair traits, skin aging, nutrient metabolism, obesity, and sleep. Although a 2-cluster solution initially offered clear separation, a 3-cluster model was selected to allow more granular profiling suitable for future consumer-targeted solution strategies.
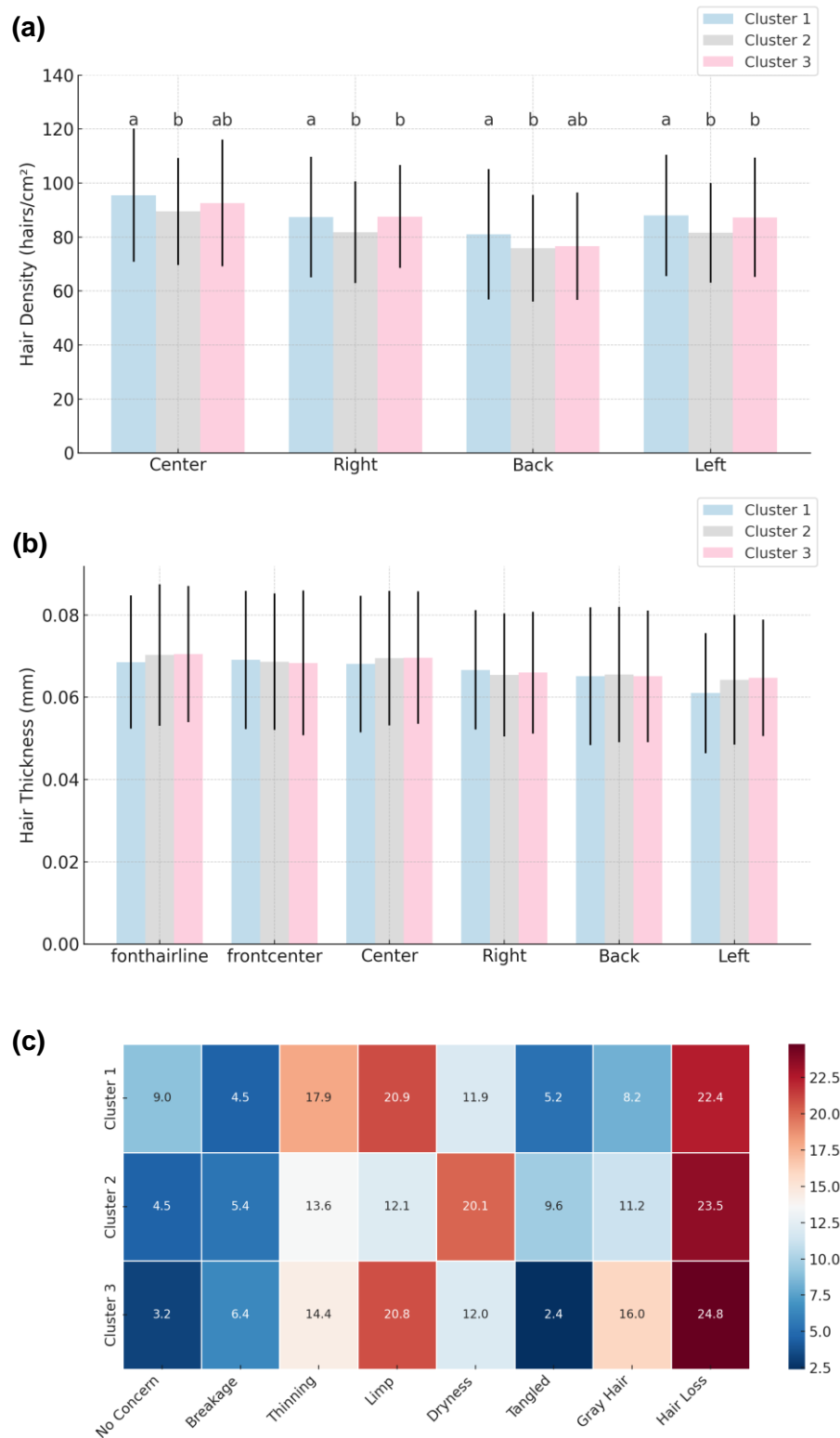
(a)

(b)

Figure 2. Genetic Cluster Profiles Across 66 Trait-Linked Markers. (a) Line plot showing average genetic rank scores (1 = favorable, 100 = attention-needed) across 66 trait-linked markers for each of the three identified clusters. Cluster 1 and Cluster 3 shared generally favorable scores in skin aging, obesity, and metabolic traits, but differed in areas such as hair thickness, vitamin metabolism, and caffeine sensitivity.
(b) Radar chart summarizing representative domains (skin, hair, metabolism, obesity, and exercise capacity) by cluster.   Cluster 2 demonstrated a consistently elevated genetic risk profile relative to Clusters 1 and 3, particularly in domains related to scalp condition, hair integrity, and metabolic health.

As shown in Figure 2a, Cluster 1 and Cluster 3 shared several common traits, including strong rankings in skin aging, obesity-related traits, and metabolic markers. However, Cluster 1 showed relatively stronger rankings in *hair thickness, muscle development, and alcohol metabolism*, suggesting higher resilience in dermatologic and performance-related traits. In contrast, Cluster 3, while similar in skin-related scores, showed relatively lower scores in pigmentation, vitamin C/D and, caffeine-related metabolism, indicating different intervention priorities. Cluster 2 was distinctively disadvantaged across multiple axes: it had the lowest rankings in skin aging, hair thickness, sleep duration, and blood sugar, along with higher scores (worse ranks) in obesity and fatigue-related pathways. This overall trend suggests Cluster 2 represents a higher-risk group requiring more comprehensive support strategies for scalp, hair, and general health optimization. These trends are further visualized in Figure 2b, which summarizes representative traits across major domains. Cluster 2 consistently appears outside the performance envelope of Clusters 1 and 3, reinforcing its characterization as the most vulnerable group across dermatologic and metabolic markers.

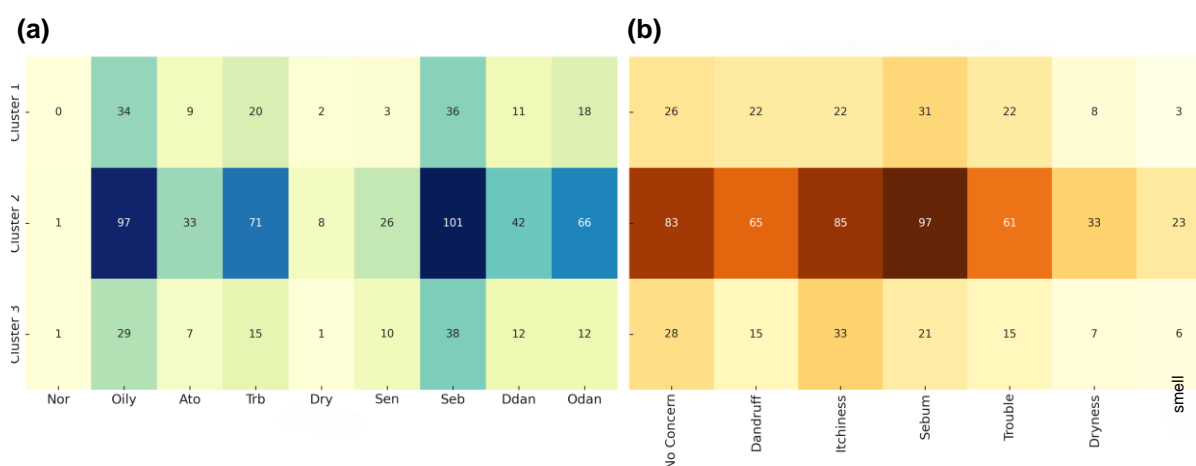**Figure 3. Hair Health and Concern Distribution Across Genetic Clusters.**
(a) Bar plots of hair density (mean ± SD) by cluster are shown for four scalp regions where statistically significant differences were observed, selected from six measured sites. Tukey HSD post hoc group letters (a, b, ab) indicate intergroup significance.
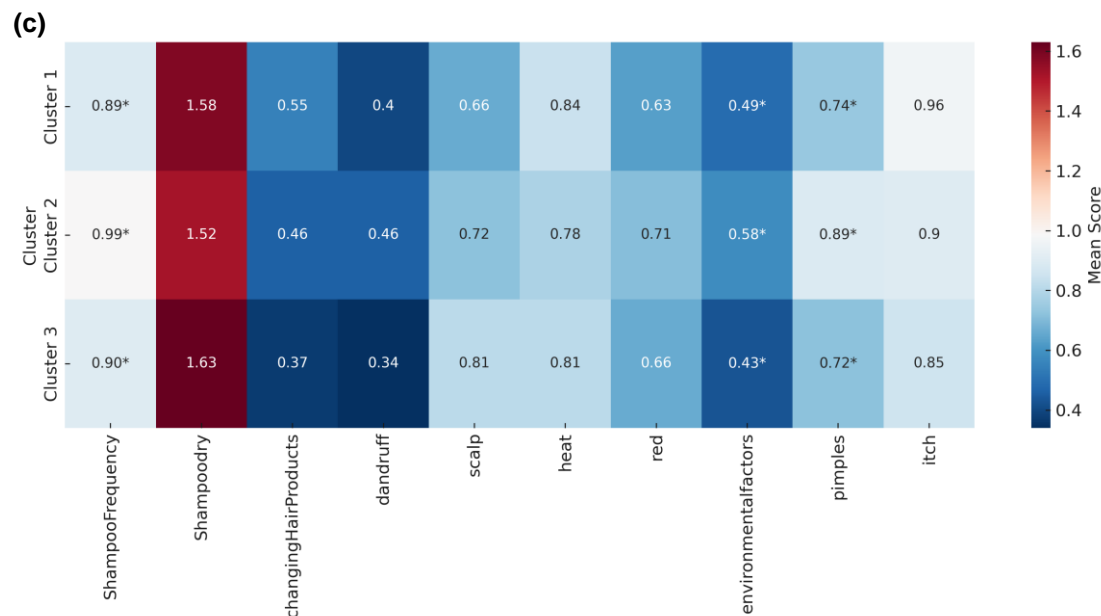**(b) Hair thickness (mean ± SD) by region and cluster; no significant differences were found.**

(c) Heatmap showing cluster-wise proportions of primary hair concern types (e.g., hair loss, thinning, gray hair).

To evaluate differences in objective and subjective hair-related characteristics, we analyzed scalp hair density and thickness across six predefined regions: the frontal hairline (fonthairline), right hairline, left hairline, mid-frontal scalp (frontcenter), vertex (center) and occipital scalp (back). As shown in   Figure 3a, among the six scalp regions assessed, four regions — the center (vertex top), right hairline, left hairline and back (occipital) — demonstrated significant differences in hair density across clusters (ANOVA p-values < 0.05). Cluster 1 consistently exhibited the highest density across these regions, while Cluster 2 showed the lowest density. Tukey HSD post hoc comparisons indicated that Cluster 1 differed significantly from Cluster 2 in all four significant regions, whereas Cluster 3 typically showed intermediate values, overlapping with either Cluster 1 or 2 depending on the region. In contrast, hair density in the frontal hairline and mid-frontal scalp regions did not show statistically significant differences among clusters. Hair thickness did not significantly differ among clusters(Figure 3b).

In addition to objective hair density and thickness measurements, participants' self-reported primary hair concerns were compared across genetic clusters. As shown in Figure 3c, a heatmap visualizing the distribution of primary hair concerns revealed distinct patterns among clusters. A chi-square test confirmed that the distribution of top concerns differed significantly across clusters (χ^2 = 33.090, p = 0.003). Specifically, Cluster 2 showed the highest proportion of individuals citing hair loss as their primary concern. In contrast, Cluster 1 participants more frequently reported concerns about thinning hair and lack of volume. Cluster 3 participants predominantly reported concerns about gray hair and limpness. These results indicate that genetic clustering is not only associated with measurable biological traits such as hair density, but also reflects subjective perceptions of hair-related issues.

(a)

	Nor	Oily	Ato	Trb	Dry	Sen	Seb	Ddan	Odan
Cluster 1	0	34	9	20	2	3	36	11	18
Cluster 2	1	97	33	71	8	26	101	42	66
Cluster 3	1	29	7	15	1	10	38	12	12

(b)

	No Concern	Dandruff	Itchiness	Sebum	Trouble	Dryness	smell
Cluster 1	26	22	22	31	22	8	3
Cluster 2	83	65	85	97	61	33	23
Cluster 3	28	15	33	21	15	7	6

(c)

	ShampooFrequency	Shampoodry	changingHairProducts	dandruff	scalp	heat	red	environmentalfactors	pimples	itch
Cluster 1	0.89*	1.58	0.55	0.4	0.66	0.84	0.63	0.49*	0.74*	0.96
Cluster 2	0.99*	1.52	0.46	0.46	0.72	0.78	0.71	0.58*	0.89*	0.9
Cluster 3	0.90*	1.63	0.37	0.34	0.81	0.81	0.66	0.43*	0.72*	0.85

Figure 4. Scalp Type and Scalp Concern Distributions Across Genetic Clusters.
(a) Heatmap showing the distribution of nine scalp types across clusters (normal, oily, atopic, troubled, dry, sensitive, seborrheic, dry dandruff-prone, oily dandruff-prone). The numbers within each cell represent the frequency of participants with each scalp type in the respective cluster. No significant differences were observed.
(b) Heatmap depicting the distribution of primary scalp concerns (e.g., itchiness, dryness, sebum) by cluster. Cell values indicate the number of participants within each cluster who selected the corresponding concern as their top priority. Again, no significant differences were found, although minor descriptive trends were noted.
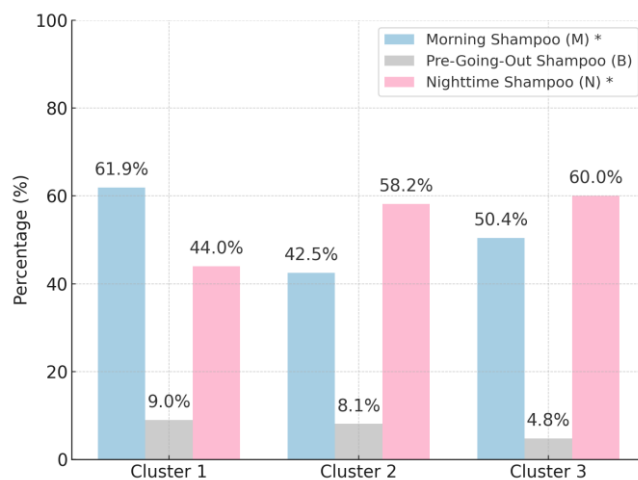**(c) Heatmap visualizing the average scores (0–3 scale) across 10 scalp sensitivity–related survey items for each genetic cluster. Higher scores indicate greater frequency or intensity of symptoms such as dandruff, redness, and itching. Cluster 2 showed consistently elevated values, particularly for pimples, itchiness, and environmental reactivity. Values marked with * indicate variables showing statistically significant differences between clusters (Kruskal–Wallis test, $p < 0.05$).**

To investigate whether genetically defined clusters were associated with specific scalp types or scalp-related concerns, we analyzed both scalp type classification and primary scalp concern selection across clusters. As shown in Figure 4a, a heatmap comparing scalp type distributions (normal, oily, atopic, troubled, dry, sensitive, seborrheic, dry-dandruff, oily-dandruff types) across clusters did not reveal statistically significant differences (Pearson χ^2 test, p = 0.727). Although formal statistical significance was not achieved, descriptive trends suggested that Cluster 2 had relatively higher frequencies of oily and seborrheic scalp types compared to Clusters 1 and 3. Similarly, Figure 4b illustrates the distribution of primary scalp concerns such as itchiness, sebum, dryness, and inflammation. Again, no statistically significant differences were observed across clusters (χ^2 p = 0.595); however, visual inspection suggested that Cluster 2 participants tended to report scalp inflammation and sebum-related

concerns more often, while Cluster 1 participants showed a slightly higher proportion reporting dryness as a top concern. Overall, while scalp type and scalp concern profiles showed minor descriptive variations among clusters, these differences were not statistically robust, in contrast to the clearer patterns observed for hair density and hair concerns. We further compared participants' self-reported scalp sensitivity and scalp-related lifestyle factors across genetic clusters using the Kruskal-Wallis nonparametric test, as the data were ordinal-scaled. To explore cluster-level differences in scalp sensitivity traits, both descriptive and nonparametric analyses were conducted.

As shown in Figure 4c, Cluster 2 exhibited the highest average scores across most survey items, especially in scalp pimples (0.89), itchiness (0.90), and environmental reactivity (0.58). These trends were reinforced by the Kruskal–Wallis test, where Cluster 2 showed the highest mean ranks in pimples (371.15) and environmental sensitivity (366.83), indicating statistically significant differences ($p < 0.05$). Participants in Cluster 2 tended to shampoo more frequently ($p = 0.027$) and reported greater scalp reactivity to environmental factors such as dryness, fine dust, and seasonal change ($p = 0.032$), along with higher frequencies of scalp pimples compared to Clusters 1 and 3 ($p = 0.002$). These findings suggest that although general scalp sensitivity measures were largely comparable among clusters, Cluster 2 individuals demonstrated a more sensitive and environmentally reactive scalp profile, consistent with their overall disadvantageous genetic and phenotypic clustering.

Figure 5. Preferred Shampooing Times Across Genetic Clusters.
Bar plots showing the proportion of participants in each cluster who reported shampooing in the morning, before going out, or at nighttime. Significant differences were observed for morning ($p < 0.001$) and nighttime ($p = 0.009$) shampooing, but not for pre-going-out shampooing ($p = 0.394$).

Beyond physiological traits, we further examined behavioral patterns by analyzing participants' preferred shampooing times categorized as morning (shampoo_M), before going out (shampoo_B), or nighttime (shampoo_N). As shown in Figure 5, significant differences were found in shampoo timing preferences among clusters. Participants in Cluster 1 were more likely to shampoo in the morning compared to Clusters 2 and 3 ($p < 0.001$), whereas nighttime shampooing was more frequently reported by Clusters 2 and 3 ($p = 0.009$). No significant differences were observed for shampooing before going out ($p = 0.394$), suggesting that the major behavioral distinctions lie between morning and nighttime routines. These results imply that genetic clustering may influence not only physiological traits but also daily hair care behaviors, particularly regarding preferred shampooing times.

4. Discussion

This study, through a holistic approach integrating genetic, biological, and lifestyle data, demonstrated that genotype-based clustering can meaningfully stratify individuals with respect to both physiological traits and behavioral patterns. Unlike conventional symptom-based segmentation, which relies on transient or subjective indicators, our approach utilized stable genetic markers to identify underlying predispositions—enabling the design of long-term, preventive, and personalized scalp and hair care strategies.

4.1. Genetic Constitution as a Stable Stratification Framework

Through unsupervised clustering of genetic ranking scores across 66 SNPs, individuals were grouped according to their intrinsic susceptibilities related to scalp sebum production, inflammatory responses, dandruff sensitivity, and metabolic traits. These genetic profiles remain constant over time, unlike environmental exposure or self-perceived symptoms, allowing for a more reliable and proactive framework for consumer segmentation and care intervention.

4.2. Cluster-Specific Scalp Profiles and Concerns

The clustering revealed distinct scalp condition patterns. Cluster 2 emerged as a high-risk group characterized by oily scalp, high sebum production, and frequent complaints related to pimples and irritation. In contrast, Clusters 1 and 3 exhibited tendencies toward dryness, barrier dysfunction, or aging-related structural decline. These findings highlight that individuals with similar symptoms (e.g., itchiness or inflammation) may have fundamentally different genetic underpinnings, calling for customized solutions even within symptomatically similar groups.

4.3. Association with Objective Hair Metrics

Hair density and thickness metrics across six scalp regions further validated the cluster differences. Cluster 2 showed lower values in both metrics, particularly in the vertex and left parietal areas—indicating a greater genetic risk for hair loss. Meanwhile, Clusters 1 and 3 showed moderate-to-high hair density but diverged in patterns: Cluster 1 exhibited early-stage thinning, whereas Cluster 3 showed signs of aging-related graying and structural weakening. These anatomical and structural distinctions suggest potential differences in responsiveness to hair growth or recovery interventions.

4.4. Lifestyle and Behavioral Patterns

Beyond biology, behavioral tendencies also varied meaningfully across clusters. Cluster 2 participants shampooed more frequently and reported higher reactivity to environmental triggers such as fine dust and seasonal changes. This may reflect compensatory behaviors in response to heightened sebaceous activity or inflammation, but also raises concern about aggravation due to over-cleansing. Such patterns underscore the need for genetic × lifestyle matched strategies—e.g., mild cleansers and barrier-supporting routines for Cluster 2, versus volume-enhancing care for Cluster 1.

4.5. Implications for Personalized Care Solutions

The stratified risk and phenotype profiles uncovered in this study offer actionable insights for targeted product development.

- **Cluster 1**: Early-stage thinning and cosmetic sensitivity → *volume-boosting, density-enhancing products*.

- **Cluster 2**: Inflammatory-prone, oily scalp → *anti-inflammatory, mild-cleansing, barrier-repair routines*.

- **Cluster 3**: Aging-associated concerns → *pigmentation maintenance, nutrition- and density-supporting solutions*.

The alignment between genetic vulnerability and subjective concern further supports the utility of genotype-driven segmentation in guiding personalized interventions. Not only can such insights inform product formulation, but they also enable more effective consumer communication, expectation management, and long-term adherence.

Limitations of this study include the female-skewed sample and reliance on self-reported lifestyle data, which may introduce bias. Future research should validate the current findings in more gender-balanced and diverse populations, and ideally incorporate longitudinal designs to evaluate predictive power over time.

5. Conclusion

This study applied genotype-based clustering to a Korean cohort of 706 individuals using 66 SNP-derived markers associated with scalp physiology, hair characteristics, skin aging, and metabolic health. By integrating these genetic profiles with clinical measurements, self-reported concerns, and behavioral patterns, we identified biologically and behaviorally meaningful subgroups that extend beyond conventional symptom-based classification.

Among the three clusters, Cluster 2 emerged as the most genetically vulnerable group, exhibiting poor genetic rankings across multiple domains including skin aging, sebum metabolism, and inflammation. This genetic disadvantage was reflected in phenotypic traits such as lower hair density, increased scalp sensitivity, and frequent compensatory behaviors including excessive shampooing.

Clusters 1 and 3 shared several favorable genetic traits and displayed relatively stable scalp and hair conditions compared to Cluster 2, suggesting a more resilient genetic background. However, further analysis revealed important distinctions in specific genetic domains:

Cluster 3 showed elevated risk scores related to hair graying and pigmentation as well as structural hair integrity, while Cluster 1 showed mild elevations in androgenetic alopecia susceptibility and sebum regulation genes. When considered together with biological and behavioral phenotypes—including differences in scalp concern profiles and shampooing patterns—these genetic distinctions enabled more granular differentiation between the two clusters. This underscores how genotype-driven segmentation can be further refined through integrated analysis of clinical and behavioral data, supporting a more precise stratification framework.

These findings highlight the potential of genotype-informed segmentation as a foundation for precision scalp and haircare strategies. Moving beyond reactive approaches, this framework enables early identification of risk-prone individuals and the development of long-term, preventive, and personalized care aligned with each individual's intrinsic biological predispositions.

6. References

[1] Gokce N, Basgoz N, Kenanoglu S, Akalin H, Ozkul Y, Ergoren MC, Beccari T, Bertelli M, Dundar M. An overview of the genetic aspects of hair loss and its connection with nutrition. *J Prev Med Hyg.* 2022 Oct 17;63(2 Suppl 3).

[2] Cuevas-Diaz Duran R, Martinez-Ledesma E, Garcia-Garcia M, Bajo Gauzin D, Sarro-Ramírez A, Gonzalez-Carrillo C, et al. The biology and genomics of human hair follicles: A focus on androgenetic alopecia. *Int J Mol Sci.* 2024 Feb 22;25(5):2542.

[3] Wang S, Kang Y, Qi F, Jin H. Genetics of hair graying with age. *Ageing Res Rev.* 2023;89:101977.

[4] Sinclair R. Genetics of hair loss: polygenic and beyond. *Br J Dermatol.* 2007;157(3):483–7.

[5] Trüeb RM, Henry JP, Davis MG, Schwartz JR. Scalp condition impacts hair growth and retention via oxidative stress. *Int J Trichology.* 2018 Nov-Dec;10(6):262–70.

[6] Vila-Vecilla L, Russo V, Torres de Souza G. Genomic markers and personalized medicine in androgenetic alopecia: A comprehensive review. *Cosmetics.* 2024;11(5):148.

[7] Draelos ZD. Cosmetic dermatology: Products and procedures for skin and hair. *J Am Acad Dermatol.* 2010;62(6):1057–70.

[8] Draelos ZD. The biology of cosmetic procedures for scalp and hair: objective vs. subjective outcomes. *J Cosmet Dermatol.* 2012;11(3):230–5.

[9] Fujimoto A, O J, et al. A replication study confirmed the EDAR gene to be a major contributor to population differentiation regarding head hair thickness in Asia. *Hum Genet.* 2008 Sep;124(2):179–85.