Supplemental Online Content

Gudobba C, Mane T, Bayramova A, et al. Automating hair loss labels for universally scoring alopecia from images: rethinking alopecia scores. *JAMA Dermatol.* Published online December 14, 2022. doi:10.1001/jamadermatol.2022.5415

This supplemental material has been provided by the authors to give readers additional information about their work.

eMethods 1. Image datasets and 'ground-truth' (GT) labeling

Images from the 1484 image dataset were taken using mobile phones and ranged from 750x750 to 1600x1600 pixels in size. All images were resized to 384x384 for the network input, guaranteeing a minimum 250-pixel head width for all images, which is the minimum resolution needed for clinicians to identify regions of alopecia correctly.[28]

For the coarse-level GT labels, we first had three experts (2 board-certified dermatologists, 1 dermatology research assistant) independently label a smaller set of 250 images and computed their pairwise inter-rater variability (via the Dice score). We then chose the person with the highest mean pairwise Dice score, *i.e.*, the one with the smallest variability with respect to the other two, as the annotator for the full dataset.

All image labels (C, F) were double-checked by the 3 researchers (2 medical students, 1 dermatology faculty) who provided our clinical baseline scores.

For the images used in the top correlation studies, 162 were taken from the 1484 set and 121 from Trichy, representing 139 and 111 unique subjects respectively. The exact demographic breakdown is provided in eTable 1.

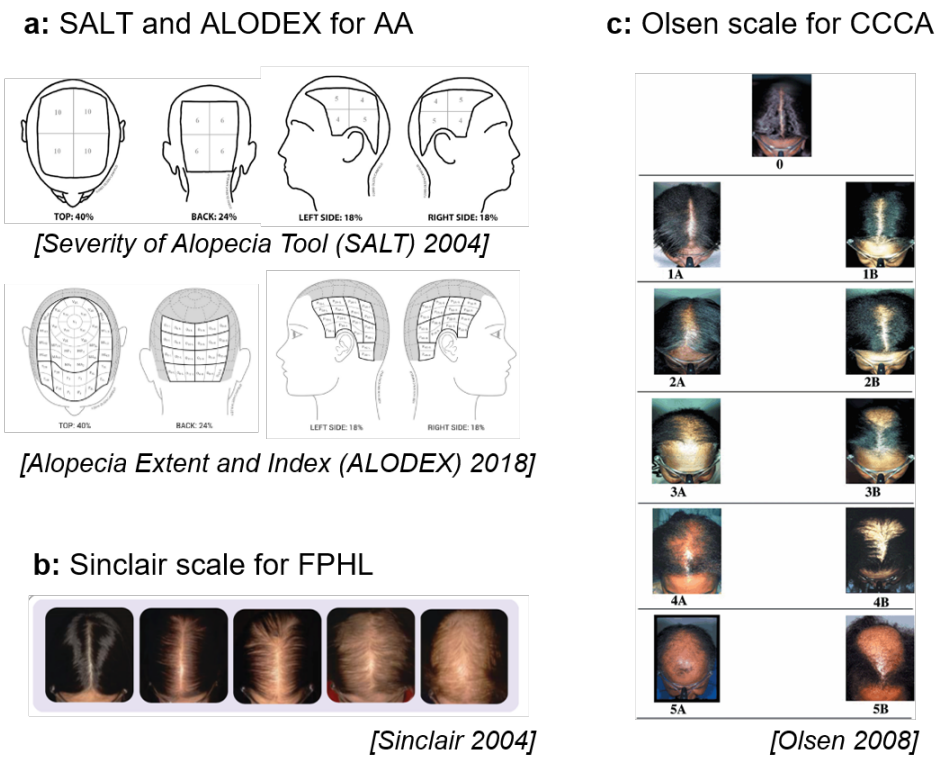
eTable 1. Demographics of Study Sample for Scoring Systems Correlations. Participant demographics are outlined both in aggregate and for the Clinic and Trichy image sets individually. Note, age information was unavailable for 28 participants (27 in the clinical set, 1 in the Trichy set). Min age was 3, max age was 77.

	Total Sample (n=250)	**Clinic Image Set (n=139)**	**Trichy Image Set (n=111)**
Sex			
Male	74 (29.6%)	21 (15.1%)	53 (47.7%)
Female	176 (70.4%)	118 (84.9%)	58 (52.3%)
Race			
Asian	20 (8.0%)	11 (7.9%)	9 (8.1%)
Black	70 (28.0%)	58 (41.7%)	12 (10.8%)
White	160 (64.0%)	70 (50.4%)	90 (81.1%)
Diagnosis			
AGA	75 (30.0%)	16 (11.5%)	59 (53.2%)
AA	66 (26.4%)	59 (42.4%)	7 (6.3%)
CCCA	50 (20.0%)	39 (28.1%)	11 (9.9%)
Other	59 (23.6%)	25 (18.0%)	34 (30.6%)
Age			
Avg (± SD)	35.3 ± 18.7	38.2 ± 22.4	32.4 ± 13.4

eMethods 2. AA, FPHL, CCCA scoring systems

The Severity of Alopecia Tool (SALT) [16], which is commonly used to measure AA, defines specific scalp areas on 4 head views to compute an overall percent hair loss. This percent is

then weighted to obtain an adjusted percent loss over the entire scalp (eFigure 1a). The Sinclair scale [23] is designed to assess FPHL via a model in which the top of the patient's head is compared to five images representing five stages of progressive hair loss (eFigure 1b). It specifically measures loss of hair density at the crown region using scores from 1 (normal) to 5 (advanced hair loss), with the development of a bald spot occurring at stage 4. In the Olsen scale for CCCA [19] (eFigure 1c), the severity of CCCA is quantified based on variations in hair density coupled with the total loss extent at the middle part line. Much like the Sinclair scale, a photographic scale for central hair loss is used to grade the pattern and severity of hair loss from 0 to 5, where 0 is no hair loss, 3-5 are probable cases of CCCA, and each increment corresponds roughly to a widening of the part. In addition to severity, the scale accounts for two variations in overall hair loss pattern: A, for anterior accentuation, and B, for vertex.



eFigure 1. Established photographic scoring systems for AA, FPHL, and CCCA. a: Severity of Alopecia Tool (SALT) [16] and its refined version the Alopecia Index and Extent (ALODEX) for AA [20]; **b:** Sinclair visual analogue scale for female pattern hairloss (FPHL) [23]; and **c:** Olsen's scale for CCCA [19]. Scales are intended to serve as a reference to which patient images can be compared and assessed in severity.

eMethods 3. Scoring systems analysis

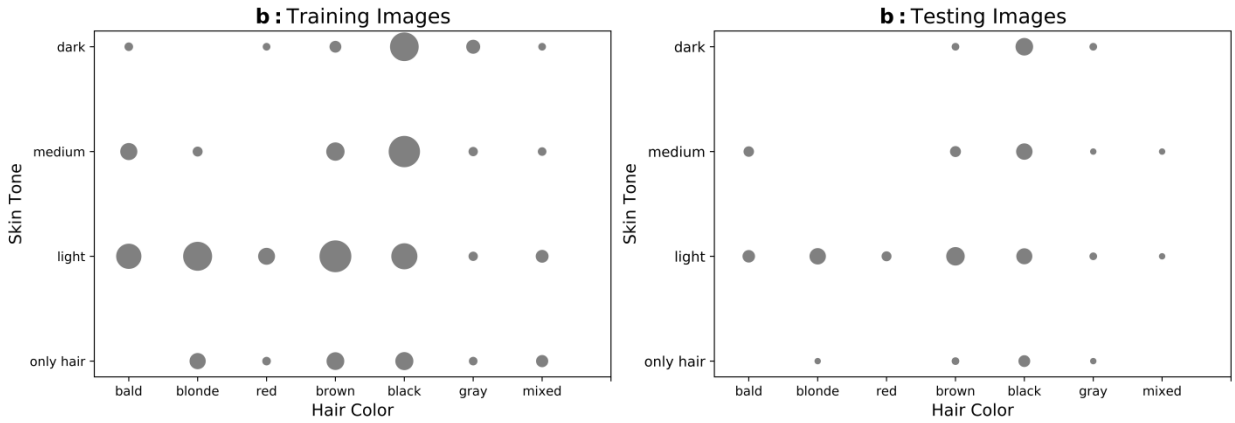
Our ICCs were calculated using the statistical program R in accordance with the guidelines set by Koo and Li.[33] Because we randomly selected our raters from a larger population of raters with similar characteristics, we used a *two-way random effects* model in our ICC calculation. We chose a *mean of k raters* type because we used the mean value of our three raters as our assessment basis. Finally, we used an *agreement* definition because systematic differences between raters could be relevant. A full report of ICC values, both by score and within score by diagnosis, is included in eTable 2.

eTable 2. ICC Values Reported Overall by Score and by Diagnosis. For each scoring system (Olsen, Sinclair), we include intraclass correlations (ICC) with associated 95% confidence intervals by diagnoses. Correlation between the three scorers was high for both for all images regardless of underlying alopecia type, with all ICCs above 0.9.

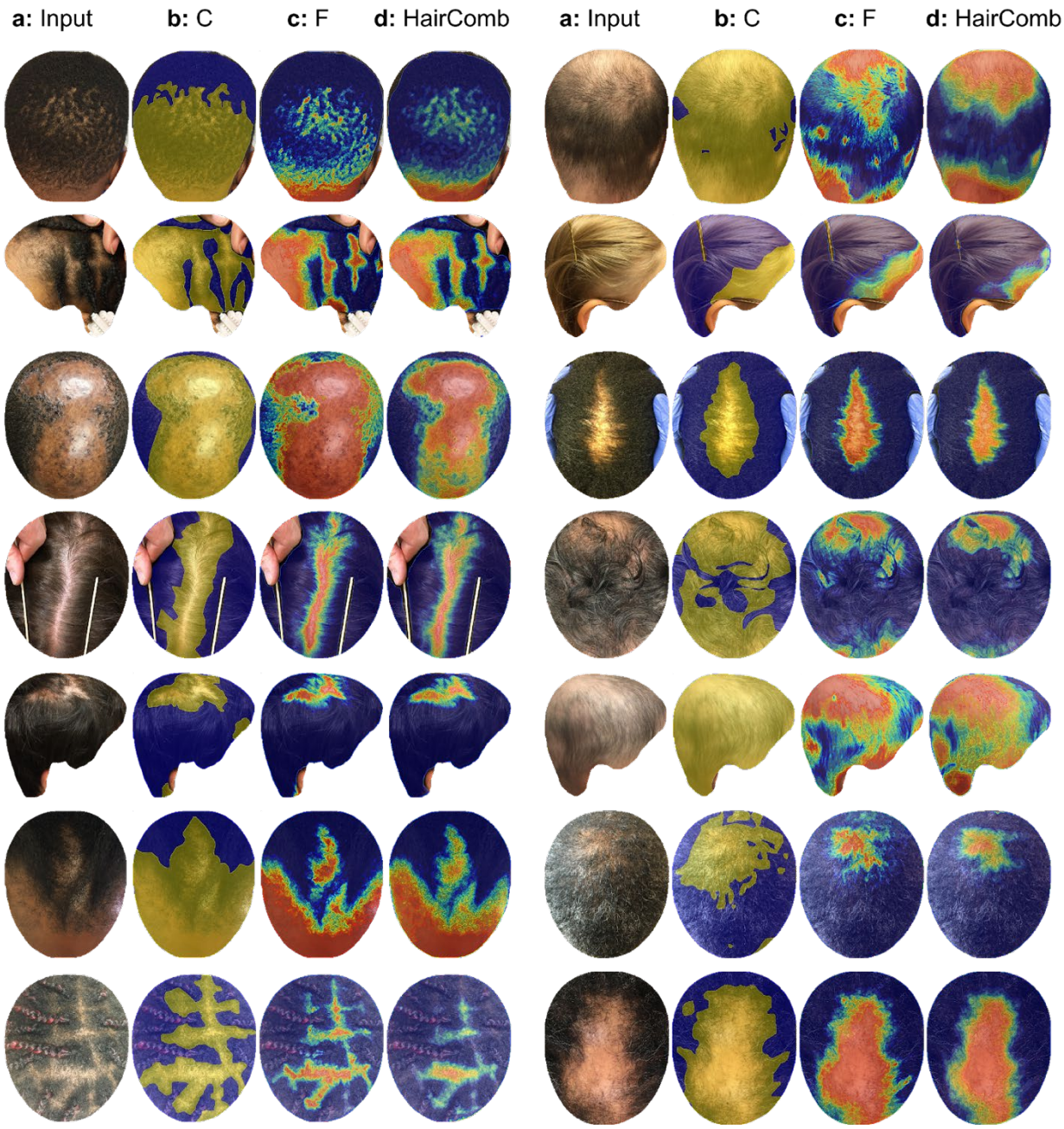
Olsen		
	ICC	**95% Confidence Interval**
Total	0.965	[0.919, 0.981]
AA	0.961	[0.883, 0.982]
AGA/FPHL	0.958	[0.897, 0.979]
CCCA	0.963	[0.768, 0.988]
Other	0.933	[0.760, 0.974]
Sinclair		
	ICC	**95% Confidence Interval**
Total	0.980	[0.975, 0.984]
AA	0.980	[0.971, 0.986]
AGA/FPHL	0.981	[0.972, 0.987]
CCCA	0.959	[0.929, 0.976]
Other	0.968	[0.942, 0.983]

eMethods 4. HairComb algorithm design

The HairComb network consists of two parallel encoder-decoder branches based on the UNet [34] and ResNet50 [35] architectures: 1) a segmentation branch (ResNet) to encode the overall hair loss information learning from coarse-level labels; 2) a regression branch (UNet) to estimate local hair probability information by extracting a hair *vs.* skin probability pixel mask, learning from the fine-level labels.



eFigure 2. Skin tone vs. hair color combinations for HairComb. For training (a) and testing (b), we show skin tone vs. hair color pairs for all 1484 images used. Marker size denotes number of images.



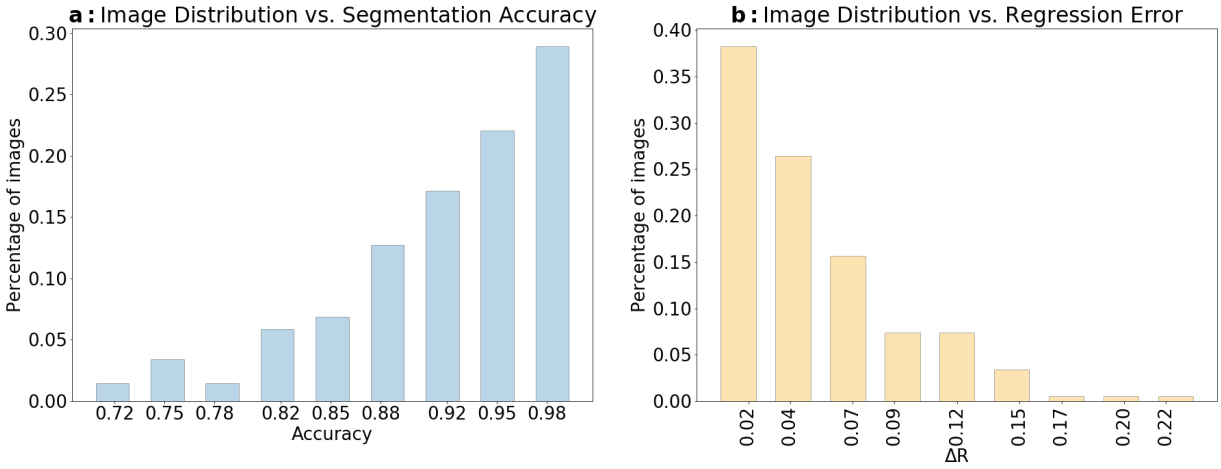
eFigure 3. Sample Qualitative Results on Multi-View Images. For each input image, we show: 1) input image; 2) coarse-level labels C^{GT}; 3) fine-level labels F^{GT}; 4) final automated output HairComb labels H. As can be noted in the examples, the algorithm performs well also in cases of poor/saturated illumination or presence of external objects (*e.g.,* fingers, swab handles, hair appliances).

Let C^{GT}, F^{GT} and C^{P}, F^{P} denote the GT and predicted fine/coarse labels respectively. The segmentation branch, trained on C^{GT}, was designed to segment the RGB images into 3 regions:

1) *normal* hair density (*i.e.*, no alopecia); 2) *abnormal* hair density (*i.e.*, alopecia, including bald scalp); and 3) *background* areas containing everything else (*e.g.* hair pins, ears, fingers/gloves). The regression branch, trained on F^{GT}, is similar to the segmentation branch, but the output layer uses a linear activation function. The final output H is computed by updating F^P such that the normal and background regions, detected in C^P, are set to 1 and 0 in F^P respectively, and then mean filtering the updated F^P with image patch size of 30 pixels.

The network was trained to update the weights using the Adam Optimizer [38] with a loss function equal to the losses for the two individual F and C branches, Λ^F and Λ^C. Λ^C is the binary cross entropy and $\Lambda^F = \sqrt{(F^P - F^{GT})\text{^}2 + \epsilon\text{^}2}$, where ϵ is a small value to guarantee differentiability during learning. [39] The value of ϵ in the loss function was chosen as 10^{-6}. We stopped training if the learning rate dropped below 10^{-5} or if the network's weights did not change over 3 epochs. We used a batch size of 5 images for training the network with a maximum of 100 epochs.

We used 1280 images for training and 204 for testing from the clinical 1484 image set (Table and eFigure 2). To train the model, we augmented the data by adding 3 rotations for each image (90, 180 and 270 degrees) with random gamma correction values ranging from 0.5 to 1.5. The final model takes less than 1 second to produce the final hair loss maps with for each image on an Intel Xeon processor with Nvidia Quadro GPU. Qualitative and quantitative results are included in eFigures 3 and 4 respectively.



eFigure 4. Image distributions *vs.* HairComb output metrics. For all output HairComb labels H, we include the percent of images *vs.* segmentation accuracy (**a**) and regression error (**b**) breakdowns.

eMethods 5. An application: building Olsen and Sinclair prediction models

In this section, we show how to use the HairComb outputs to create simple prediction models for both Sinclair and Olsen. For both scores, we started by identifying a set of possible features from the ones used in the clinical scoring: percent bald scalp (B) and percent low (but non-zero) hair density area (L). More precisely, analogously to h_{low}, we let h_{bald} be the threshold between areas of partial loss and complete baldness. We then defined the low density and bald areas by letting $L = \{h_{scalp} \leq H^{GT} \leq h_{low}\}$ and $B = \{0 \leq H^{GT} \leq h_{bald}\}$, where H^{GT} is the baseline hair loss probability at each pixel created by combining F^{GT} and C^{GT} in the same way the outputs C^P and F^P were combined to obtain H. Noting that Sinclair scores are a direct measure of hair density change at the part line, we determined h_{low} and h_{bald} via a univariate grid search optimization on the Lasso regression [37] model accuracy (with just $\%L$ and $\%B$ as input features) w.r.t. the average manual Sinclair scores. Thus, analyzing automated *vs.* manual scores revealed that raters perceived the alopecia areas when hair densities fell below 80% and baldness areas when below 10%.

We then added the average hair loss percent (δ_{avg}), computed from averaging hair loss percent values of all pixels in $L + B$, as well as more localized information such as maximum and average widths (w_{max} and w_{avg}) of the main affected alopecia area (*i.e.,* the largest alopecia area determined by pixel count). Finally, given the different focus of the two scales, we made different adjustments for each. For the Sinclair score, we added $\delta_{avg} w_{max}$ to capture the density and geometry of the hair part. For Olsen score, we empirically estimated the imaginary horizontal split line to determine whether the alopecia pattern was frontal (Olsen's 'A'), vertex (Olsen's 'B'), or spread over the top (AB), and extracted the restricted region percent affected area ($\Pi_{H(A,B,AB)}$) corresponding to the percent affected area of the relevant region.

For each model, we performed feature analysis using a Lasso regression model to determine the optimal set of features and built two support vector machine models to predict the scores. All input features were normalized (0-1) and the best features were determined using a Lasso regression model (L1 linear regression), by selecting the ones with highest recursive feature elimination ranking. From the feature analysis, L and B, together with δ_{avg}, were the best indicators for predicting both scores. As expected, information regarding the alopecia region of higher prevalence, $\Pi_{H(A,B,AB)}$, was a strong indicator for Olsen's score, and $\delta_{avg} w_{max}$ was a

eTable 3. Olsen and Sinclair prediction model quantitative evaluation. Table includes average score error using different models to predict Olsen and Sinclair scores using the automated HairComb labels H on the top-view images testing set, *i.e.*, 57 images of the tops set that were not included in the training models. As a baseline, we include errors obtained by training/testing the models on H^{GT}. Support vector machine gave the best results in both scenarios (in bold, prediction errors of the results presented in the main manuscript).

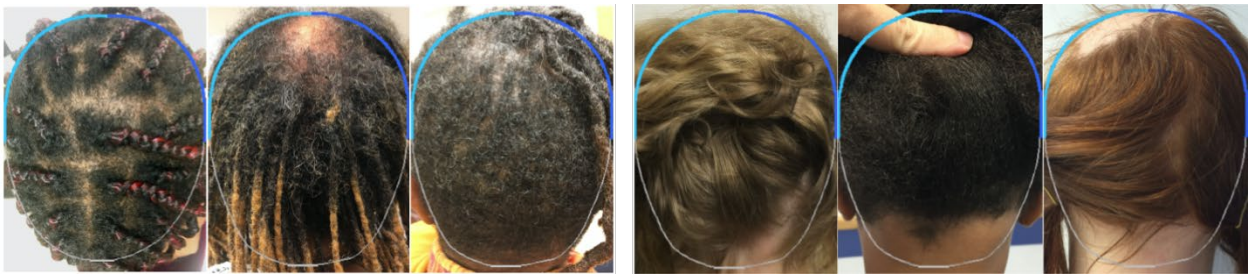
	GT labels H^{GT}		**Output HairComb labels H**	
	Sinclair Error	**Olsen Error**	**Sinclair Error**	**Olsen Error**
Prediction Model				
Lasso Regression	0.45	0.55	0.53	0.66
Ordinary Least Squares	0.44	0.53	0.49	0.68
Support Vector Machine	0.44	0.52	**0.48**	**0.60**

high indicator for Sinclair's score. Finally, we found the optimal horizontal split value for Olsen's score to be 45% of the height of the head.

To conclude, we experimented with various regression models, including Ordinary Least Squares, Lasso regression, and Support Vector Machine (SVM) for both the Olsen and Sinclair score models. In eTable 3, we show errors between predicted and clinical scores as well as provide a baseline of these errors by comparing them with the ones obtained by using H^{GT} instead to train/test the models. For all models, we used a 75-25 split for creating training and testing sets. Support vector machine gave the best results in both scenarios (eTable 3, in bold).

eDiscussion. Image taking guidelines

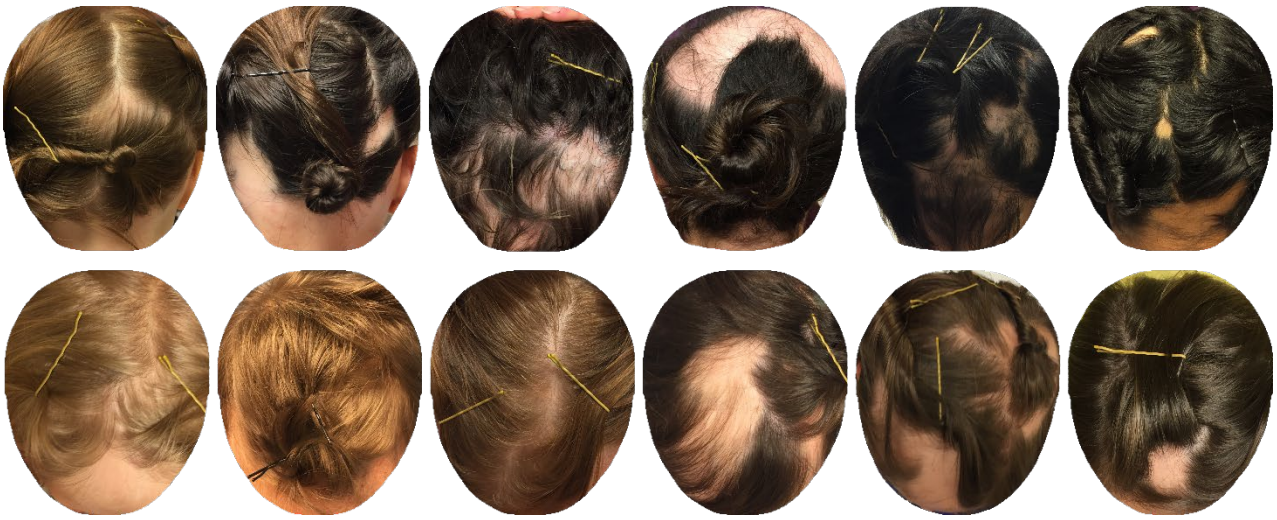
This final section includes styling instructions (eTable 4, eFigures 5-6) and scalp capture guidelines recommended for different alopecia types (eTables 4-5).



eFigure 5. Hair styling examples included in Trichy for AA and CCCA. Left: Sample acceptable hair preparations for CCCA; box braids, twists, locs, weaves, etc. can remain untouched for image capture. Right: Sample hair styling solutions for AA to expose areas of hair loss.

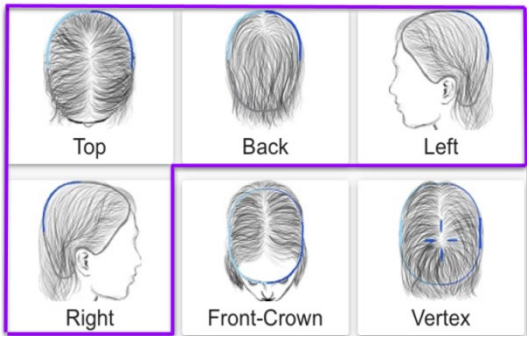
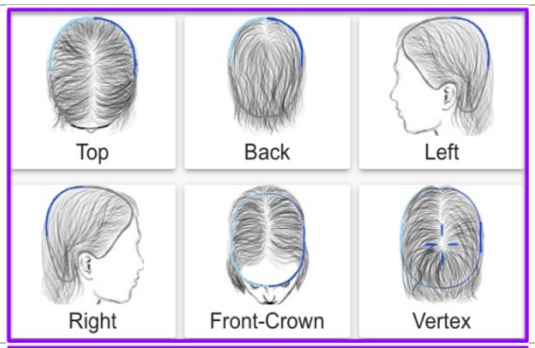
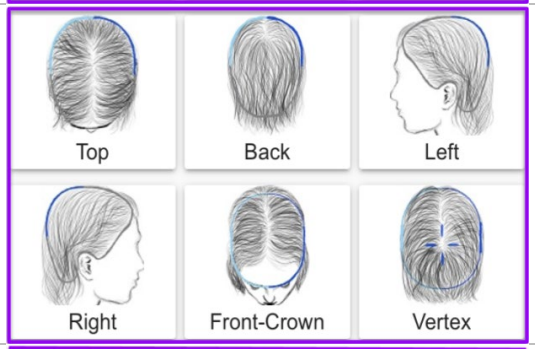
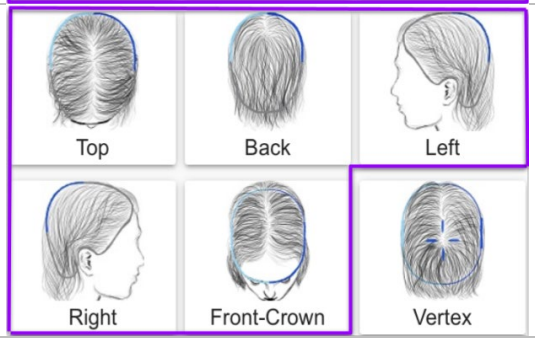
eTable 4. Hair styling instructions for different alopecia types. For each alopecia type, we include preparations tips to help style the hair to capture the areas of hair loss.

General instructions for all hair types	● If your hair is tied back to cover any hair loss, let it down. ● Remove wigs/detachable extensions. ● If you have box braids, twists, locs, weaves, leave them as they are.
Additional instructions for Alopecia Areata (AA)	● Expose areas of involvement/patches with the help of pins/ties by twisting and securing strands of hair. ● You may have to change styling based on the view to achieve maximum exposure of the patches.
Additional instructions for Male/Female Pattern Hair Loss (AGA/FPHL)	● Part hair in the middle. This is necessary to reveal the true density of hair, especially in individuals with long hair. ● Coarser hair types may require consistent application of force to keep the hair parted, please use hair pins/ties for this purpose. ● Unless otherwise instructed by your doctor, keep the middle part the same for all views.
Additional instructions for Central Centrifugal Cicatricial Alopecia (CCCA)	● Part hair in the middle. This is necessary to reveal the true density of hair, especially in individuals with long hair. ● Coarser hair types may require consistent application of force to keep the hair parted, please use hair pins/ties for this purpose. ● Unless otherwise instructed by your doctor, keep the middle part the same for all views.
Additional instructions for Frontal Fibrosing Alopecia (FFA)	● Expose areas of involvement by pulling the hair back and securing it with hair pins/ties. ● You may have to change styling based on the view to achieve maximum exposure.



eFigure 6. Sample hair preparation for photographing alopecia under longer hair. To uncover hidden areas of alopecia, pins and picks should be used to twist the hair and pull it up. This is particularly useful in cases of AA as shown in these images.

eTable 5. Guidelines for image taking using standardized views. Similarly to the 4 scalp views required for computing the SALT score for AA, the illustrations below provide guidelines on viewpoints needed for common types of alopecia.

	Standardized image views recommended for each alopecia type outlined in purple	**Special View-dependent Instructions**
Alopecia Areata (AA)		**Include:** Top, Back, Left, Right. **Left/Right/Back views:** If you experience hair loss on the sides and lower scalp in band-like pattern (ophiasis), pull your hair up to expose the loss.
Male/Female Pattern Hair Loss (AGA/FPHL)		**Include:** Top, Back, Left, Right, Front-Crown, and Vertex. **Top/Front-Crown:** For longer hair, part hair in the middle. If needed, press down on each side of the part with hair pins/ties to keep the hair flat.
Central Centrifugal Cicatricial Alopecia (CCCA)		**Include:** Top, Back, Left, Right, Front-Crown, Vertex. **Top/Back/Front-Crown/Vertex:** Part hair in the middle. If needed, press down on each side of the part with hair pins/ties to keep the hair flat. **Left/Right:** Part hair to expose areas of involvement using similar tools if needed.
Frontal Fibrosing Alopecia (FFA)		**Include:** Top, Back, Left, Right, Front-Crown. **Top/Front-Crown:** pull hair straight back to expose the front hairline. **Left/Right/Back:** If you experience hair loss also on the sides and lower scalp in a band-like pattern, pull your hair up to expose the loss.