nature portfolio

Corresponding author(s): *Angela M. Christiano*

Last updated by author(s): **2021/10/25**

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- [ ] [x] The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] [x] A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [ ] [x] The statistical test(s) used AND whether they are one- or two-sided
  Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- [x] [ ] A description of all covariates tested
- [ ] [x] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [ ] [x] A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- [ ] [x] For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
  Give P values as exact values whenever suitable.
- [x] [ ] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [x] [ ] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [x] [ ] Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software used to collect the data
Data analysis	Methods used for analysis (ATAV and TraP-score) are reported in the methods section and were published by Ren et al. at doi: https://doi.org/10.1101/2020.06.08.136507 and Gelfman et al. https://doi.org/10.1038/nrg281. The code used in this manuscript for gene-level burden analysis of the exome data was previously published in Gelfman et al18. The ATAV software was constructed by the Institute for Genomic Medicine and the code is published online at https://github.com/nickzren/atav

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Exome sequencing data of KRT82 that support the findings in this study have been deposited in dbGaP repository with accession codes phs002632.v1.p1 [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002632.v1.p1]. The raw exome sequencing BAM files submitted in Sequence Read Archive(SRA) under accession number PRJNA768921 (Link: https://www.ncbi.nlm.nih.gov/bioproject/768921). The processed phenotypic files of human exome data are available at dbGaP (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002632.v1.p1). Users can gain access by following the dbGaP data access protocol explained here: https://www.ncbi.nlm.nih.gov/projects/gapprev/gap/cgi-bin/GetPdf.cgi?document_name=GeneralAAInstructions.pdf. Full immunoblots for Fig. 6D are provided in the Source Data File.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The initial sample consisted of 873 AA cases and 18,581 controls. Final sample size for each cohort (cases/controls, Caucasian and multi ancestry) were established by counting reducing from the total number of samples the samples that did not pass QC for a total number of 849 AA cases and 15,640 controls. We used SKAT software package to perform power calculations with 500 AA cases and 500 controls, and estimated more than 80% power to identify significant associations at 2.5E-06 (exome wide significance). Therefore we determined we had robust power to identify associations
Data exclusions	Sample data was excluded based on contamination and population stratification. Samples reporting >8% contamination according to VerifyBamID53 were excluded. KING54 was used to ensure only unrelated (up to third-degree) individuals contributed to the analysis. Furthermore, to be eligible for analysis, we required samples to have >87% of the consensus coding sequence (CCDS) covered at 10X. Once controlled for low quality and related samples, we applied an additional analysis to control for population stratification by using EIGENSTRAT 55 to remove samples that were considered as genetic outliers. This ensured that the main cluster of samples was of similar genetic origins and would not be biased by sub-population that are genotypic outliers.
Replication	For sequencing, we used the full set of case samples available and did not use an additional case cohort to replicate the results. Immunoblot, IHC and IF were repeated with similar results a varying number of times, specifics are included in the figure legends
Randomization	For sequencing, the samples were allocated to experimental groups based on case control status of Alopecia Areata and by PCA analysis of ancestry in case of Caucasian only analysis. For other experiments in this study such as expression analysis and staining, samples were also allocated by their disease status or hair cycle stage (mice)
Blinding	During human sample collection, blinding was not possible due to the disease status of individuals. However, investigators were blinded to group allocations during analysis.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National*

Research sample	*Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☑ Antibodies
☑	☐ Eukaryotic cell lines
☑	☐ Palaeontology and archaeology
☐	☑ Animals and other organisms
☐	☑ Human research participants
☑	☐ Clinical data
☑	☐ Dual use research of concern

Methods

n/a	Involved in the study
☑	☐ ChIP-seq
☑	☐ Flow cytometry
☑	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-KRT82 (Progen cat# GP-HHB2; 1:200 IF, 1:2000 WB); anti-GP (Invitrogen cat# A-11073, 1:1000); HRP-conjugated anti-guniea pig IgG (Jackson ImmunoResearch cat# 106-035-033, 1:1000); HRP-conjugated anti-B-actin (Santa Cruz cat# sc-47778, 1:5000); anti-CD8 (abcam cat#ab101500; 1:100); biotinylated anti-rabbit Ig, H+L (VECTOR cat#BA-1000); biotinylated anti-guinea pig Ig (VECTOR cat#BA-7000)
Validation	anti-KRT82 Progen cat# GP-HHB2: Langbein, L. et al. The catalog of human hair keratins: II. Expression of the six type II members in the hair follicle and the combined catalog of human type I and II keratins. J. Biol. Chem. 276 anti-CD8 abcam Cat#ab101500: Abpromise Guarantee for IHC. Antibody reacts with human. Referenced in 20 applications listed on the antibody website (https://www.abcam.com/cd8-alpha-antibody-sp16-ab101500.html). Abpromise guarantee covers product applications and species that were tested in Abcam labs by suppliers or trusted collaborators

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*

Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	C57BL/6J male mice at various stages of the hair cycle / and age (Postnatal(p)19, p29, p31, p43, p45, p60, p90). Ambient temperature was maintained at 22-26C with a humidity of 40-60%. Mice were housed in individual cages with a maximum density of 5 mice per cage, and were kept on a 12-hour dark/light cycle. Mice between the ages of postnatal day (PND) 14-PND80 were analyzed for the study.
Wild animals	There were no wild animals used in this study
Field-collected samples	No field samples were used in this study
Ethics oversight	Columbia University's Institutional Animal Care And Use Committee (IACUC)

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Samples were prioritized on self-reported Caucasian ethnicity and positive family history of AA disease. A small number of non-Caucasian samples were included in sequencing and removed from analysis using Eigenstrat and principal component analysis. 50 control DNA samples of Caucasian ethnicity and mized genders were ascertained from Northwell Health. IGM controls were selected from >45,000 WES individuals. We excluded all controls with a known diagnosis of infectious disease, primary immune deficiency and endocrine disorders, but they were not explicitly screened for AA. Of the 849 AA patients, 607 were female, 241 male, and 1 unspecified. Average age was 32.5.
Recruitment	Participants presenting to one of five clinical practices associated with the National Alopecia Areata Registry were invited to participate if they received a diagnosis of alopecia areata by a dermatologist and if they completed screening and enrollment by clinical collaborators. Thus, our cohort includes people with alopecia areata who have access to healthcare in five geographic regions in the US and are willing to participate in research. This study was not likely to ascertain participants who chose to self-manage disease symptoms or did not have access to healthcare. We do not expect that these forms of self-selection have introduced bias into our study. Specifically, we expect that the inclusion of such "hidden populations" would not have altered our study results.
Ethics oversight	The study was approved by the local IRB committees from the five institutions that participate in the National Alopecia Areata Registry, including the University of Texas, MD Anderson Cancer Center; Columbia University Irving Medical Center; University of Minnesota, Minneapolis; University of Colorado, Denver; and University of California, San Francisco.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.