nature research

Corresponding author(s): Pinkal Desai and Duane Hassane

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

Describe how sample size was determined.

All available cases with centrally adjudicated AML diagnosis were used from the Women's Health Initiative Cohort. Controls were matched to cases on a one to one basis based on age, follow up time, sample collection, and medical history.

2. Data exclusions

Describe any data exclusions.

Participants with a history of myeloid disorders at baseline were excluded.

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

Longitudinal mutation data were compared to reference baseline and demonstrated 95% mutation rediscovery as shown in figures S10 and S11, indicating reproducibility of findings.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Participants were allocated based on diagnosis of AML as cases who were diagnosed with AML and controls who did not develop AML in the cohort.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators were blinded to case control status and clinical data during the experiments and these data were available for analyses only after genomic sequencing data was released to WHI.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ 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.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	R software v3.4.0, ggplot2 v2.2.1, plyr v1.8.4, dplyr v0.7.4, reshape2 v1.4.3, logistf v1.22, ComplexHeatmap v1.12.0, circlize v0.4.3, maftools v1.4.25, MutationalPatterns v1.0, PyMol 2.0 , Trimmomatic v0.32 , AdapterRemoval v2, BWA MEM v0.7.12, SamBlaster v0.1.21, MarkDupsByStartEnd v0.2.1, VarDictJava v1.4.6, VerifyBamID 1.1.3, vt v0.5, Picard Tools v2.6.0, CNVkit v0.8.6, Variant Effect Predictor v85, SnpEff v4.1g

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.	No restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell lines were used
b. Describe the method of cell line authentication used.	Not applicable, no eukaryotic cell lines were used
c. Report whether the cell lines were tested for mycoplasma contamination.	Not applicable, no eukaryotic cell lines were used
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	Not applicable, no eukaryotic cell lines were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.	No animals were used

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	Women aged 50-79 years at baseline were enrolled in the WHI regardless of ethnicity.