nature research

Corresponding author(s): David J Adams, Benedict Paten, Thomas M Keane

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.

   Fig 1e: Sample size for the genomic enrichment is the number of repeat elements (Repeatmasker) in the mouse genome.

   Fig. 3d: For the Efcab3-like knockout mouse, sample size is 6 (3 wild types, and 3 controls)

2. Data exclusions

   Describe any data exclusions.

   No data was excluded

3. Replication

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

   This was a genomics study primarily. All raw sequencing data and assembled genomes have been deposited in relevant public databases (accessions provided in Supplementary Table 8, 17, 18).
   The knock out mouse model used was replicated across multiple biological replicates, and is available via the KOMP/IMPC projects.

4. Randomization

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

   No randomisation of samples/organisms/participants was applied.

5. Blinding

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

   No blinding was carried out in this study.

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.

Software name Version URL
SGA v0.9.43 https://github.com/jts/sga
bwa v0.7.5 & 0.7.12-r1039 https://github.com/lh3/bwa
bedtools v2.25.0 https://github.com/arq5x/bedtools2
GATK MarkDuplicates v3.4 https://software.broadinstitute.org/gatk/
SOAP2 r240 https://github.com/aquaskyline/SOAPdenovo2
Dovetail HiRise v0.75 https://dovetailgenomics.com/webstore/plant-animal/hi-rise-software/
SNAP mapper v0.15.4 http://snap.cs.berkeley.edu/
progressiveCactus v0.0 http://blaxter-lab-documentation.readthedocs.io/en/latest/progressivecactus.html
Ragout v2.0 https://github.com/fenderglass/Ragout
Repeatmasker open-4.0.5 http://www.repeatmasker.org/
PseudoPipe n/a
RCPedia n/a
samtools v1.2 http://www.htslib.org
bcftools v1.2 http://www.htslib.org
Sanger-pathogens/assembly-stats v1.0.0 https://github.com/sanger-pathogens/assembly-stats
CEGMA v2.5 http://korflab.ucdavis.edu/datasets/cegma/
PantherDB 12 http://www.pantherdb.org/
BLASTall v.2.2.25 https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download
Knickers v1.5.5 http://www.bnxinstall.com/knickers/Knickers.htm
Geneious R8 https://www.geneious.com/

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).

In terms of validation, all these panels were validated and optimised from mouse (via pilots and titration experiments) by internally at Sanger. These antibodies and panels are based on the commonly used panels developed for the large scale mouse phenotyping corsortia like IMPC and EUMODIC. References for previous use of these antibodies is provided on the supplier websites.

Antibody and fluorochrome Channel Dilution Supplier Catalogue # Clone
CD44 FITC FITC 2000 BD 561859 IM7
CD25 PE PE 500 Biolegend 102008 PC61
CD62L PE-CF594 PE-Texas Red 2000 BD 562404 MEL-14
TCRαβ PerCP-Cy5.5 PerCP-Cy5.5 600 Biolegend 109228 H57-597
KLRG1 PE-Cy7 PE-Cy7 500 Biolegend 138416 2F1
CD161/NK1.1 BV421 Pacific Blue 600 Biolegend 108732 PK136
CD4 BV510 AmCyan 3000 Biolegend 100553 RM4-5
TCRγδ APC APC 600 Biolegend 118116 GL3
CD45 Alexa 700 Alexa 700 600 Biolegend 103128 30-F11
CD8a APC-H7 APC-Cy7 200 BD 560182 53-6.7

Ly6B FITC FITC 1000 Serotec MCA771FB 7/4
I-A/I-E PE PE 4000 Biolegend 107608 M5/114.15.2
CD19 PE-CF594 PE-Texas Red 2000 BD 562291 ID3
Ly6C PerCP-Cy5.5 PerCP-Cy5.5 5000 Biolegend 128012 HK1.4
CD11b PE-Cy7 PE-Cy7 2000 Biolegend 101216 M1/70
Ly6G V450 Pacific Blue 600 BD 560603 1A8
IgD BV510 AmCyan 2000 BD 563110 11-26c.2a
CD115 APC APC 500 Biolegend 135510 AF598

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used in sections a-d.

b. Describe the method of cell line authentication used.

Describe the authentication procedures for each cell line used OR declare that none of the cell lines used have been authenticated OR state that no eukaryotic cell lines were used.

c. Report whether the cell lines were tested for 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 OR state that 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.

Provide a rationale for the use of commonly misidentified cell lines OR state that no commonly misidentified 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.

The DNA for the 16 mouse strains was obtained from the Jackson Laboratory, strain identifiers are:

Strain Jax stock no. Generation of sequenced animal
C57BL/6NJ 005304 ?+F8
FVB/NJ 001800 F95pF98
A/J 000646 F280
AKR/J 000648 F256
BALB/cJ 000651 F226
C3H/HeJ 000659 F258pF262
CBA/J 000656 F275
CAST/EiJ 000928 F90pF93
DBA/2J 000671 F219pF224
LP/J 000676 F195
NOD/ShiLtJ 001976 F117pF121
NZO/HlLtJ 002105 ?+F41
PWK/PhJ 004660 F69+3+17
SPRET/EiJ 001146 F78
WSB/EiJ 001145 ?+F4
129S1/SvImJ 002448 F63pF65

Details of the knockout mouse model are given in Supplementary methods:
Strain: CBAxC57BL/6J, bred with C57BL/6NTac.
Phenotyping: 16 weeks
Sex: 7 male 8 female

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.	This study did not involve human participants.