EMBO reports

Manuscript EMBO-2016-42395

Identification of novel genes and networks governing hematopoietic stem cell development

Tianxu Han, Chao-Shun Yang, Kung-Yen Chang, Danhua Zhang, Farhad B Imam, Tariz M Rana

Corresponding author: Tariq Rana, UCSD School of Medicine

Review timeline:

Submission date:	17 March 2016
Editorial Decision:	14 April 2016
Revision received:	04 August 2016
Editorial Decision:	30 August 2016
Revision received:	13 September 2016
Accepted:	30 September 2016

Editor: Achim Breiling

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

No Peer Review Process File is available with this article, as the authors have chosen not to make the review process public in this case.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

Corresponding Author Name: Tariq M. Rana
Journal Submitted to: EMBO reports
Manuscript Number: EMBOR-2016-42395V1

Reporting Checklist For Life Sciences Articles (Rev. July 2015)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.
- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

Please ensure that the answers to the following questions are reported in the manuscript itself. We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

In the pink boxes below, provide the page number(s) of the manuscript draft or figure legend(s) where the information can be located. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable).

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	For the fish experiment, at least 13 fish were counted for analysis as shown in Fig 5D and Appendix Figure S3F.
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	The numbers of fish in each group were shown in Figure 5D and Appendix Figure S3F.
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	Page 18, 19 For DNA-seq data analysis, a dataset of shRNA library sequences was generated from 57,600 22-bp hairpin stem sequences. Next, we removed the low-quality reads and converted the FASTQ file format to FASTA format. Blast analysis of the sequenced reads (100 bp) against the shRNA library was carried out, and perfectly matched shRNAs were used for normalization. Page 19, An shRNA read in ESCs of <10 reads per million total reads was set as the cut-off value. For GO and enrichment analysis, the ANOVA file was uploaded to the Metacore server. The cut-off values for the fold-change and p-value were set as 2 and 0.05, respectively.
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	No. The cells or fish embryos are indistinguishable. Therefore, they were considered as identical.
For animal studies, include a statement about randomization even if no randomization was used.	NA. The cells or fish embryos are indistinguishable. Therefore, they were considered as identical.
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	No. For assessing results from cell culture based experiments, the cells were indistinguishable and were considered as identical. For fish in situ hybridization, the fish were grouped at one cell stage and hence no more steps were taken when assessing results.
4.b. For animal studies, include a statement about blinding even if no blinding was done	NA. For assessing results from cell culture based experiments, the cells were indistinguishable and were considered as identical. For fish in situ hybridization, the fish were grouped at one cell stage and hence no more steps were taken when assessing results.
5. For every figure, are statistical tests justified as appropriate?	Yes
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	NA
Is there an estimate of variation within each group of data?	NA
Is the variance similar between the groups that are being statistically compared?	NA

C- Reagents

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	Page 18 Detailed information on the antibodies used in this study is provided in Table EV4 Expanded View Table 4. Antibodies for flow cytometry and FACS. Antibodies Vender Cat. No Clone Anti-Mouse CD309 (FLK1) APC ebioscience 17-5821 Avas12a1 Anti-Mouse CD184 (CXCR4) PerCP-eFluor® 710 ebioscience 46-9991 2B11 Anti-Human/Mouse SSEA-1 PE ebioscience 12-8813 eBioMC-480 (MC-480) BD Lineage Antibody Cocktail BD Biosciences 558074 Anti-Mouse CD117 (c-Kit) PE ebioscience 12-1171 2B8 Anti-mouse Ly-6A/E PE-Cy7 BD Biosciences 558162 D7 Anti-Mouse CD41 APC ebioscience 17-0411 eBioMWReg30 (MWReg30) Anti-Mouse CD45 PE ebioscience 12-0451 30-F11 Anti-human CD34 APC ebioscience 17-0349 4H11 anti-human CD90 PE-Cy7 ebioscience 25-0909 eBio5E10 (5E10)
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	Page 16 Materials and Methods Mouse embryonic stem cell culture The mouse ESC line iHoxB4 (kindly provided by Dr. Michael Kyba) and E14tg2a (a gift from Dr. Chuan He) were used in this study.

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	Page 20. Zebrafish studies General maintenance, collection, and staging were performed as previously described [65]. All animal work was approved by the Institutional Review Board at the University of California, San Diego and was performed in accordance with Institutional Animal Care and Use Committee guidelines.
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	NA
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	NA

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	NA
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	NA
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	NA
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	NA
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	NA
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	NA
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	NA

F- Data Accessibility

18. Provide accession codes for deposited data. See author guidelines, under 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	The GEO accession numbers for the DNA-seq and RNA microarray data are GSE86898 and GSE86853, respectively.
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	NA
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	NA
21. As far as possible, primary and referenced data should be formally cited in a Data Availability section. Please state whether you have included this section. Examples: **Primary Data** Wetmore KM, Deutschbauer AM, Price MN, Arkin AP (2012). Comparison of gene expression and mutant fitness in Shewanella oneidensis MR-1. Gene Expression Omnibus GSE39462 **Referenced Data** Huang J, Brown AF, Lei M (2012). Crystal structure of the TRBD domain of TERT and the CR4/5 of TR. Protein Data Bank 4O26 AP-MS analysis of human histone deacetylase interactions in CEM-T cells (2013). PRIDE PXD000208	Page 13. For example, comparison of our Group X target genes with those previously reported to play roles in HSPC differentiation [27] revealed that only 67 of the 351 genes in Group X were previously identified. Referenced Data 27. McKinney-Freeman S, Cahan P, Li H, Lacadie SA, Huang HT, Curran M, Loewer S, Naveiras O, Kathrein KL, Konantz M, et al. (2012) The transcriptional landscape of hematopoietic stem cell ontogeny. Cell Stem Cell 11: 701-14
22. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	NA

G- Dual use research of concern

23. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	NA