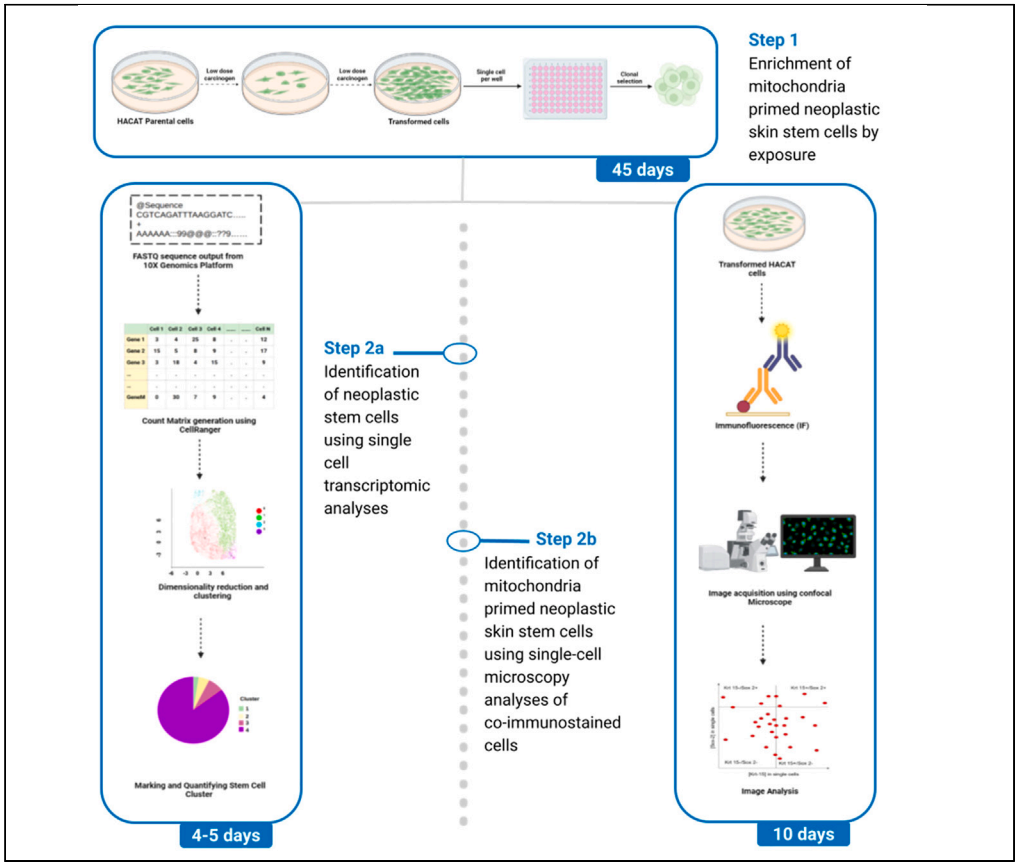
STAR Protocols

CellPress
OPEN ACCESS

Protocol

Enrichment of carcinogen-driven "mitochondria-primed" human skin stem cells and their identification using single-cell analyses



Previous work has shown that mitochondria play a critical role in priming stem cells to self-renew and proliferate. Here, we describe a protocol for enriching and identifying the mitochondria-primed stem cells (mpSCs) for their characterization and applications. We describe steps for enriching mpSCs with the environmental carcinogen 2,3,7,8-tetrachlorodibenzo-p-dioxin in a skin keratinocyte lineage and for identifying mpSCs using single-cell transcriptomics and single-cell microscopy analyses of expression of relevant stem cell markers.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Swati Agarwala, Mayank Saini, Danitra Parker, Kasturi Mitra

kasturi.mitra@ashoka.edu.in

Highlights

Enriching "mitochondria-primed" neoplastic skin stem cells with TCDD exposure

Identifying neoplastic skin stem cells by single-cell transcriptomics

Identifying "mitochondria-primed" neoplastic skin stem cells by single-cell microscopy

Agarwala et al., STAR Protocols 4, 102545 September 15, 2023
https://doi.org/10.1016/j.xpro.2023.102545

STAR Protocols

CellPress
OPEN ACCESS

Protocol

Enrichment of carcinogen-driven ''mitochondria-primed'' human skin stem cells and their identification using single-cell analyses

Swati Agarwala,[1,3,4] Mayank Saini,[1,3,4] Danitra Parker,[2] and Kasturi Mitra[1,2,5,*]

[1]Department of Biology, Trivedi School of BioSciences, Ashoka University, Sonipat, Haryana 131029, India

[2]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA

[3]These authors contributed equally

[4]Technical contact: swati.agarwala@ashoka.edu.in; mayank.saini_phd21@ashoka.edu.in

[5]Lead contact

*Correspondence: kasturi.mitra@ashoka.edu.in
https://doi.org/10.1016/j.xpro.2023.102545

SUMMARY

Previous work has shown that mitochondria play a critical role in priming stem cells to self-renew and proliferate. Here, we describe a protocol for enriching and identifying the mitochondria-primed stem cells (mpSCs) for their characterization and applications. We describe steps for enriching mpSCs with the environmental carcinogen 2,3,7,8-tetrachlorodibenzo-p-dioxin in a skin keratinocyte lineage and for identifying mpSCs using single-cell transcriptomics and single-cell microscopy analyses of expression of relevant stem cell markers.

For complete details on the use and execution of this protocol, please refer to Spurlock et al.[1]

BEFORE YOU BEGIN

We have previously described flow cytometry dependent isolation of neoplastic mitochondria primed stem cells (mpSCs) from ovarian cancer cells.[2] Here, we describe the protocol for a series of experiments demonstrating enrichment and identification of neoplastic mpSCs from skin keratinocytes. In the main publication, we showed that low dose of the environmental carcinogen 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) enriches the mpSCs with greater self-renewal ability that are sustained by fine-tuned repression of mitochondrial fission.[1] In that model, the fine-tuned repression of mitochondrial fission happens by either repressing Drp1 phosphorylation (in neoplastic situation) or by repressing Drp1 protein (in non-neoplastic situation). This protocol is a combination of three parts: Part 1 - Enrichment of mitochondria primed neoplastic skin stem cells, Part 2a - Identification of neoplastic stem cells using single cell transcriptomic analyses and Part 2b - Identification of mitochondria primed neoplastic skin stem cells using single-cell microscopy. Part 1 describes enrichment of mpSCs by neoplastic transformation of skin keratinocyte cell line-HACAT by low dose of TCDD. Part 2a describes their identification using single cell RNA sequencing (scRNASeq). Part 2b describes single cell immunofluorescence analyses through microscopy. The same protocol can be appropriately modified to be applied to the use of any other agents causing neoplastic transformation on any other cell type (including primary cells). Similarly, any confocal microscope with appropriate hardware and software setup can be used for data acquisition for Part 2b. For Part 1 and 2b, basic knowledge and expertise of cell culture and confocal microscopy are prerequisites. For Part 2a, scRNASeq analyses will be executed using the terminal and R on an Ubuntu system, and thus a basic understanding of both the system and R as an environment is a prerequisite. Part 2a is compatible with any relevant scRNASeq datasets with all relevant files ready.

Enrichment of mitochondria primed neoplastic skin stem cells by exposure to 1 nM TCDD

Reagent preparation for cell culture and transformation

Timing: 3–4 h

1. Preparation of reagents for HACAT cell culture.
   a. Complete Dulbecco's Modified Eagle's Medium (DMEM-10% FBS) (500 mL).
      i. In a disposable vacuum filter flask unit (0.22 μm), add all the reagents as per relevant table in materials and equipment section.
      ii. Filter the complete media (DMEM-10% FBS).

   Note: One can also use powdered DMEM or without supplements DMEM and supplement it with glucose (4.5 g/L), sodium pyruvate (1 mM), L-Glutamine (4 mM), Penicillin (100 μg/mL), Streptomycin (100 μg/mL), and 10% FBS using standard techniques. The appropriate medium should be made for any other cell type used.

      iii. Aliquot the filtered medium in multiple 50 mL sterile tubes and store them at 4°C, to be used when required.

   ⚠ CRITICAL: The complete medium can be stored at 4°C for a maximum of 1 month. Therefore, the volume of the medium can be scaled up or down as per the usage.

   b. 10× Phosphate Buffer Saline (PBS) (1 L).
      i. Take 500 mL of dd-H_2O in a 1 L screw cap bottle.
      ii. Follow the recipe mentioned in relevant table in materials and equipment section.
      iii. Mix them well using a magnetic stirrer.
      iv. Make the volume up to 1 L using dd-H2O.
   c. 1×-PBS working solution.
      i. Take 10 mL of 10× PBS and add 90 mL of dd-H2O.
      ii. Filter sterilizes using 0.22 μm syringe filters.
      iii. Autoclave the solution.
      iv. Open the autoclaved solution inside the hood and use it as and when required.
   d. 0.25% Trypsin.
      i. Aliquot directly from the main bottle to multiple 15 mL sterile tubes.
      ii. Store them at −20°C and use single tubes as and when required.
2. Preparation of medium with TCDD for cell transformation.
   a. Make 1 nM, 10 nM TCDD solution and Toluene solution (as vehicle control) in the complete pre-warmed DMEM, to be used fresh.
   b. Add the reagents as mentioned in relevant table in materials and equipment section from the stock solution and make the volume up to 5 mL using DMEM.

   ⚠ CRITICAL: TCDD is a known carcinogen. Please wear appropriate PPE when handling TCDD or TCDD containing media.

Identification of neoplastic stem cells using single cell transcriptomic analyses

Installation of Ubuntu, R, CellRanger and Seurat on your system

Timing: 1.5 h

Note: Recommended Ubuntu version 20+ LTS. All instructions provided here are executeable on Ubuntu system. However, it can run on Windows and macOS provided the system requirements of the software are met and the user has the knowledge of corresponding basic functions.

3. Updation/Installation of Ubuntu.
   a. Install/update the Ubuntu packages to resynchronize the package index files from their sources, run the code in the box:

```
#Open the Terminal by long pressing Ctrl+Alt+T

Black screen opens with a $

#Update the Ubuntu packages

$ sudo apt update
```

Note: This is valid for Terminal environment for Linux and macOS. However, for Windows, the terminal can be run using virtual machine, installable using standard steps (Not recommended).

4. Installation of R.
   a. Install the latest version of R by running the following code on the Terminal.

```
> $ sudo apt install --no-install-recommends r-base
```

   b. Check if R is installed properly using the below mentioned code:

```
> $ sudo -i R
```

Note: scRNASeq packages are R version sensitive and works best at version >4.2.0.

Once R is installed, the following message will appear on the terminal screen:

```
R version 4.2.2 Patched (2022-11-10 r83330) -- "Innocent and Trusting"

Copyright (C) 2022 The R Foundation for Statistical Computing

Platform: x86_64-pc-linux-gnu (64-bit)

Type 'demo()' for some demos, 'help()' for online help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

>
```

Note: '>' indicates that the user is inside R environment

5. Installation of Cell Ranger.
   a. Check for minimum systems requirements for running Cell Ranger pipeline as mentioned below:
      i. 8-core Intel or AMD processor (16 cores recommended).
      ii. 64GB RAM (128GB recommended).
      iii. 1TB free disk space.
      iv. 64-bit CentOS/RedHat 7.0 or Ubuntu 14.04 (deprecated November 2021).
   b. Setup Folder for Cell Ranger Installation.

i. Create a folder named cell ranger in the working directory using:

```
> $ mkdir cellranger
```

ii. Execute the following code to enter the directory:

```
> $ cd cellranger
```

c. Installation file for Cell Ranger.

i. Fill out and submit the "10× Genomics End User Software License Agreement" information form by clicking on the following link:
10× Genomics End User Software License Agreement: https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest

Note: Upon submission, you will be redirected to a page with Cell Ranger Installer link.

ii. On the redirected page, select and copy the link for 'wget' enclosed in a box under the heading 'Cell Ranger (tar.gz compression)' using the code in the box.

```
wget -O cellranger-7.1.0.tar.gz "https://cf.10xgenomics.com/releases/cell-exp/cellranger-
7.1.0.tar.gz?Expires=1688047529&Policy=eyJTdGF0ZW1lbnQiOlt7IlJlc291cmNlIjoiaHR0cHM6Ly9jZi
4xMHhnZW5vbWljcy5jb20vcmVsZWFzZXMvY2VsbC1leHAvY2VsbHJhbmdlci03LjEuMC50YXIuZ3oiLCJDb25kaX
Rpb24iOnsiRGF0ZUxlc3NUaGFuIjp7IkFXUzpFcG9jaFRpbWUiOjE2ODgwNDc1Mjl9fX1dfQ__&Signature=XW-
YuUYczAbgdNOJZ5HZTm70BeHYzY7~aLJSMqa0BOXabAFRR3kupJnbHX9mb9rfWucHM1WP5lyre7UprxbIMMIFWKZ
1a0juemowVU98tZB3z7Iq5YwhaAfS3gACzekgQnS07-Tx-XdVsV3ucQbU2gd9w8hUMjrclkUFTpdK3-W7w0tVZ9D
wDoK8gvlVw-bfPuffMPhC8Ehe4oKxWNxyUqso1YfJroNBrHTx30POAvykd4uiexTVHxEbI6Iiy2WXE1W44oAcMb
~fejCRX2q13-IhMyaINXZynH1xRLRJB26MfaiykPhmOm~AN0O7jcME31CoVWdTkMkzEK-Yk9fUmQ__&Key-Pair-
Id=APKAI7S6A5RYOXBWRPDA"
```

iii. Paste the copied link in the previous step to download the Cell Ranger installer link by running the code inside the previously created "cellranger" directory.

iv. Once the installation is complete, output on the terminal screen will display the message as mentioned in the box below:

```
Connecting to cf.10xgenomics.com (cf.10xgenomics.com)|104.18.1.173|:443... connected.

HTTP request sent, awaiting response... 200 OK

Length: 669216050 (638M) [application/x-tar]

Saving to: 'cellranger-7.1.0.tar.gz'

cellranger-7.1.0.ta 100%[==================>] 638.21M 2.80MB/s in 6m 3s

2023-04-02 19:21:16 (1.76 MB/s) - 'cellranger-7.1.0.tar.gz' saved [669216050/669216050]
```

v. Unpack the installer file using the following code:

```
> tar -xzvf cellranger-7.1.0.tar.gz
```

d. Download and unpack reference genome , in this case human reference genome named "refdata-gex-GRCh38-2020-A" using the code in the box below:

```
wget https://cf.10xgenomics.com/supp/cell-exp/refdata-gex-GRCh38-2020-A.tar.gz

tar -xzvf chttps://cf.10xgenomics.com/supp/cell-exp/refdata-gex-GRCh38-2020-A.tar.gz.
```

Note: If needed, find the other required reference genomes (containing their indices and other essential data) and updated Cell Ranger installer link from 10× Genomics downloads page: https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest

e. Add to the Path.
   i. To be able to use the "cellranger" command from anywhere on your system, add the Cell Ranger directory to your PATH by running the following code.

```
>$exportPATH=/home/username/cellranger/cellranger-7.1.0:$PATH
```

Note: Replace the 'username' with the respective username of the system in use.

f. Verify the installation by running the following code:

```
> $ cellranger testrun --id=tiny
```

Note: Output will look like this:

```
Running Cell Ranger in test mode...

Martian Runtime - v4.0.3

...

Pipestance completed successfully!

yyyy-mm-dd hh:mm:ss Shutting down.

Saving pipestance info to "tiny/tiny.mri.tgz"
```

6. Install Seurat and other package dependencies in R environment.
   a. Install Seurat in the R environment using:

```
> install.packages("Seurat") |troubleshooting problem 2|
```

b. Load the library to check the installation using:

```
> library(Seurat)
```

c. Once installed, the following message appears on the Terminal screen:

```
> Attaching SeuratObject
```

Note: For further details on Seurat, refer to Stuart et al.[3]

d. Install all the necessary packages listed in the "key resources table" to ensure smooth running of Seurat and other pipelines for scRNASeq analysis. For this, execute the following code.

```
> install.packages(c("Package1","Package2","Package3"))
```

Note: Replace the terms package1, package2 and so on with package names from the key resources table. For example, to install 'dplyr' and 'cowplot', the code will be: >install.packages(c("dplyr","cowplot")

e. Check if the packages are installed properly by using:

```
> sessionInfo()
```

Note: Check for the presence of the installed packages in the output list.

7. scRNA Sequencing Data.
   a. Raw scRNA sequence output data obtained from the sequencer are in the FASTQ format, which is the point of beginning for scRNASeq analysis in this protocol.

Example of a single entry in a FASTQ file:

```
@TF1

TTTGGCATATTCCAAACGAACGAACGATACATGGTTTCCCACACGACGAAAAATTACGAGCAAAGCAG

+

AAAAAAAAAAAABBBAAAAAAAAAAAEEEEEEEEEEEE@::::::::'''BBAABAB@''':::::BB
```

Important points and R setup for analyses tools

Timing: 30 min

8. Points to be noted for the step-by-step method details section.
   a. Each step describes the Objective, the INPUT and OUTPUT files with their respective paths followed by the executable code contained within a numbered box.
   b. Create the input directory 'in' and output directory 'out' inside the 'scrna_seq' folder as per the input and output paths mentioned in each step.
   c. Copy and paste codes from the numbered boxes to the Terminal/R environment as directed in each step.
   d. In the boxes with the codes:
      i. The text following '#' indicates comments describing the action of the executable code.
      ii. Replace italicized part of the code with input file locations as chosen.
      iii. Each code in a box can be copied and pasted on the Terminal/R environment as directed to achieve the output at each step.
9. Setup R to run Seurat pipeline.

Specify the working directory and output filename for each generated output using the code in the box.

```
Open The Terminal

# Create a folder named "scrna_seq" in your home directory

mkdir scrna_seq/

#Activate R on the Terminal

sudo R

# Print current working directory

getwd()

# Set working directory to a desired location

# Replace 'username' with the respective username of the system in use

setwd("/home/username/")

# Set up code chunk options for knitr

knitr::opts_chunk$set(echo = TRUE, cache = TRUE)

# Set analysis ID as per your choise, like date etc.

VERSION <- 20210113

# Define the function to create an output name using the analysis ID and Cell Ranger sample filename

outfile_name <- function (filename) {

return(paste(VERSION, filename, sep = "-"))}

# Define the function to save an RDS file compressed with xz

saveRDS.xz <- function(object, file) {

  con <- pipe(paste0("xz -T0>", file), "wb")

  saveRDS(object, file = con)

  close(con)}
```

Identification of mitochondria primed neoplastic skin stem cells using single-cell microscopy analyses of co-immunostained cells

Reagent preparation for immunocytochemistry

Timing: 1.5–2 h

10. Fixing solution.
    a. From the 16% paraformaldehyde (PFA) vial, take 2.5 mL of paraformaldehyde and mix in 5 mL of 1× PBS.
    b. Add 0.4 g of sucrose to the solution and dissolve it.
    c. Make the final volume up to 10 mL using 1× PBS.

⚠ CRITICAL: Formaldehyde is hazardous. Use personal protective equipment while preparing. Chemical fume hoods should be used for formaldehyde processing.

Note: After opening the PFA ampule, it should be stored at 4°C and to be used within 15–20 days for the best results.

11. 10% Bovine serum albumin solution.
    a. Weigh 2 g of BSA and dissolve in 10 mL of 1× PBS.

b. Make the volume up to 20 mL. This will yield a 10% BSA solution.
c. Store it at 4°C.

12. 1% BSA (Blocking solution).
    a. Add 1 mL of 10% BSA solution to 9 mL of 1× PBS.
    b. Store it at 4°C.
13. Permeabilizing solution (0.1% Triton-X).
    a. In a 15 mL tube add 5 mL of 1×-PBS and add 5 μL of Triton-X 100 solution to it.
    b. Dissolve it properly by gentle vortexing.

⚠ CRITICAL: Pipette and dispense Triton-X 100 solution very carefully and slowly as it is highly viscous. Cutting the tip of the micropipette tips diagonally may facilitate this process.

14. Antibody cocktail solution (Primary-1:100 dilution).
    a. The working antibody solution should be made in 1% BSA as per the required dilution.
    b. In 800 μL of 1% BSA add 8 μL each of antibody 1 and antibody 2 respectively. (See table).
    c. Use the cocktail solution for primary antibody staining.

Antibody combinations

	Antibody-1(Localization)	Antibody-2 (localization)
Combination-1	Sox-2(nucleus)	Krt-15 (cytosol)
Combination-2	Sox-2(nucleus)	pDrp-1(S616) or total Drp1 (cytosol)
Combination-3	Sox-2(nucleus)	pCyc-E(T62) (nucleus)

15. Antibody cocktail solution (Secondary-1:1000 dilution).
    a. The working antibody solution should be made in 1% BSA as per the required dilution.
    b. In 800 μL of 1% BSA, add 0.8 μL each of antibody 1 and antibody 2 respectively.
    c. Use the cocktail solution for secondary antibody staining.

Note: Secondary antibodies should be chosen based on the kind of primary antibodies we are using.

16. Mounting medium.
    a. Take 1 mL of Fluoromount G in a microcentrifuge tube.
    b. Add 1 μL of Hoechst-33342 (10 μg/mL) into it.
    c. Gently mix the solution by inverting the tube manually.

⚠ CRITICAL: Pipette and dispense Fluoromount-G solution very carefully and slowly as it is highly viscous. Cutting the tip of the micropipette tips diagonally may facilitate this process.

Culturing HACAT skin keratinocyte cell line (volume applies to T25 flask or equivalent)

Timing: 4–5 days

17. Aspirate all media from the flask of a growing culture of cells.
18. Wash the cells using 5 mL of 1× PBS and aspirate completely.
19. Add 2 mL of 0.05% 1× Trypsin and keep it at 37°C for 5–10 min (Mix 400 μL of 0.25% trypsin in 2 mL of 1×-PBS to get 0.05% 1× trypsin).
20. Flush the plate with trypsin to detach all cells.
21. Transfer cells to a 15 mL conical tube with 5 mL of growth media.
22. Centrifuge at 150–200 × *g* for 3 min at 22°C.

23. Remove the supernatant and resuspend in 1 mL of growth media to count the cells via the preferred counting method and plate cells at the desired cell density.
24. Spread the cells evenly on the dish surface.
25. Incubate the cell plate at 37°C with 5% CO2 and monitor daily for confluence, stress, or contamination.

Note: Cells take 30–45 min to passage. Cells should be grown to 80% confluency. HACAT cells can be safely passaged twice a week when seeded at 100,000 cells/mL in a 100 mm dish. Cells should be passaged at least two times after thawing to allow full revival before doing any experiments.

⚠ CRITICAL: HACAT cells are often difficult to detach from tissue culture plate surfaces. The incubation time after the addition of trypsin may be increased to 15 min, but cells must be carefully monitored to prevent death. A higher concentration of trypsin can be used for faster detachment while taking utmost care to prevent over-trypsinization that will cause cell death.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Sox2 (mouse, clone 030–678) (1:100)	BD Pharmingen	561,469 RRID: AB_10694256
Cytokeratin15 (rabbit, clone EPR1614Y) (1:200)	Abcam	ab52816 RRID: AB_869863
Cy5 conjugated goat anti-mouse	Jackson ImmunoResearch Laboratories	115-175-146
Alexa Fluor 488 conjugated goat anti-rabbit	Jackson ImmunoResearch Laboratories	111-545-144
Drp1 (mouse, clone 8/DLP1) (1:100)	BD Transduction	611,112 RRID: AB_398423
pCyclin E- T62 (rabbit polyclonal) (1:50)	Cell Signaling Technology	4,136 RRID: AB_2071080
pDrp1 S616 (rabbit, clone D9A1)	Cell Signaling Technology	4,494 RRID: AB_11178659
Chemicals, peptides, and recombinant proteins		
2,3,7,8-tetrachlorodibenzo-p-dioxin solution	Merck-Supelco	48599
DMEM-high glucose	Hi-Media	AL219A
Fetal bovine serum	Hi-Media	RM10681-500mL
Pen/Strep	Hi-Media	A001A
Trypsin-EDTA (0.25%)	Hi-Media	TCL049-100mL
Ethanol	AR China	Amazon
Toluene	Sigma-Aldrich	179418
Sodium chloride	Sigma-Aldrich	S9888
Potassium chloride	Sigma-Aldrich	P3911
Sodium phosphate dibasic	Sigma-Aldrich	S9763
Potassium phosphate monobasic	Sigma-Aldrich	P0662
Triton X-100	SRL	64518 (2024271)
Fluoromount G	Southern Biotech	0100–01
Hoechst-33342	Invitrogen	H3570
BSA	Sigma-Aldrich	A9647
Sucrose	Sigma-Aldrich	S8501
Paraformaldehyde (EM grade)	Electron Microscopy Solutions	15711
Nail polish	MAC	Amazon
Deposited data		
Raw, analyzed, and meta data deposited in Gene Expression Omnibus	GSE171772	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE171772

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Cell lines		
Cell line (human)	HACAT	M.Athar PMID: 27725709 (gift), originally obtained from AddexBio Technologies
Software and algorithms		
Cell Ranger 10× Genomics	RRID: SCR_017344	https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest
R (v4.2.2)	-	https://www.r-project.org/
Seurat (v4.2.0)	RRID: SCR_016341	https://satijalab.org/seurat/
dplyr (v1.0.8)		https://dplyr.tidyverse.org/
tidyverse (v1.3.1)	-	https://www.tidyverse.org/
patchwork (v1.1.1)	-	https://patchwork.data-imaginist.com/
ggplot2 (v3.3.6)	-	https://ggplot2.tidyverse.org/
RColorBrewer (v1.1-3)	-	https://r-graph-gallery.com/38-rcolorbrewers-palettes.html
remotes (v2.4.2)	-	https://cran.r-project.org/web/packages/remotes/index.html
devtools (2.4.5)	-	https://cran.r-project.org/web/packages/devtools/index.html
future	-	https://cran.r-project.org/web/packages/future/index.html
metap	-	https://cran.r-project.org/web/packages/metap/index.html
BiocManager	-	https://cran.r-project.org/web/packages/BiocManager/index.html
Others		
60 mm Cell Culture Dishes laboratory	NEST	705001
TC-treated 6-well cell culture plates	NEST	703001
T-25 flask	NEST	707001
Serological pipettes	Corning	4487
Tissue roll	-	Amazon
Auto-pipette tips	Fisher Scientific	-
Disposable filter units	Thermo Scientific	564–0020
Falcons, 15 mL and 50 mL	Corning	430055, 430829
Coverslip	Fisher Scientific	22X60-1; #062915-9
Syringe	DispoVan sterilized 50 mL	Amazon
MF-Millipore membrane filter, 0.22 µm pore size	Merck	GSWP04700
Cell counting slides for TC20 cell counter, dual chamber	Bio-Rad	1450015
Nunc Lab-Tek chamber slides (convenient but expensive, simple glass slides with coverslips can also be used after optimization)	LabTek-II	154534
LSM-700 axio-observer confocal Microscope	Carl Zeiss	-
Autoclave	Tomy-SX-700	-
Shaker rocker	Rocking shaker II, BOEKEL	-
Class-II bio-safety hood with laminar flow	Microfilt	-
Auto-pipettes	Eppendorf	-
Incubator	Eppendorf	-
Cell counter TC20	Bio-Rad	-
Dry bath	Lab Armor	-
Pipette gun	Drummond Scientific	-
Bright field microscope	Olympus	-
Fluid aspiration system	Vacuubrand-BVC professional	-
Centrifuge 5425/5425 R - microcentrifuge	Eppendorf	-

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
System for scRNASeq analysis	A machine with at least 64 GB of memory is recommended. However, using up to approximately 100–128 GB can still be beneficial. This protocol should be run with at least 8 CPU cores. We used 32 cores in ours.	-

MATERIALS AND EQUIPMENT

Complete media DMEM (500 mL)		
Reagents	Final concentration	Amount
Dulbecco's Modified Eagle Medium (DMEM), High glucose with 4.5 g Glucose per liter, L-Glutamine, Sodium pyruvate and Sodium bicarbonate without Calcium chloride	N/A	445 mL
Fetal Bovine Serum (FBS)	10% v/v	50 mL
Penicillin and streptomycin cocktail mix	10 UI/mL and 100 µg/mL respectively	5 mL
Total	N/A	500 mL
Storage conditions: Temperature-4°C, Maximum storage time-1 month		

10×-Phosphate Buffer Saline (PBS) (1 L)		
Reagents	Final concentration	Amount
NaCl	1.37 M	80 g
KCl	27 mM	2 g
Na_2HPO_4	10 mM	14.4 g
KH_2PO_4	18 mM	2.4 g
Storage conditions: Temperature-4°C, Maximum storage time-1-2 month		

Medium with TCDD		
Reagents	Final concentration	Amount
DMEM Complete Media	N/A	5 mL
TCDD	1 nM	0.12 µL (From 30 µM stock vial)
	10 nM	1.2 µL (From 30 µM stock vial)
Toluene	10 nM	50 µL (From 100 mM stock solution)
Total	N/A	5 mL
Storage conditions: To be used fresh.		

STEP-BY-STEP METHOD DETAILS

Enrichment of mitochondria primed neoplastic skin stem cells

Timing: ~ 45 days

The aim of this part is to accomplish enrichment of mitochondria primed stem cells in HACAT cell population by transforming them with low dose of the carcinogen TCDD.

1. Grow HACAT skin keratinocytes to 80% confluency. Passage as indicated in the 'before you begin' section.

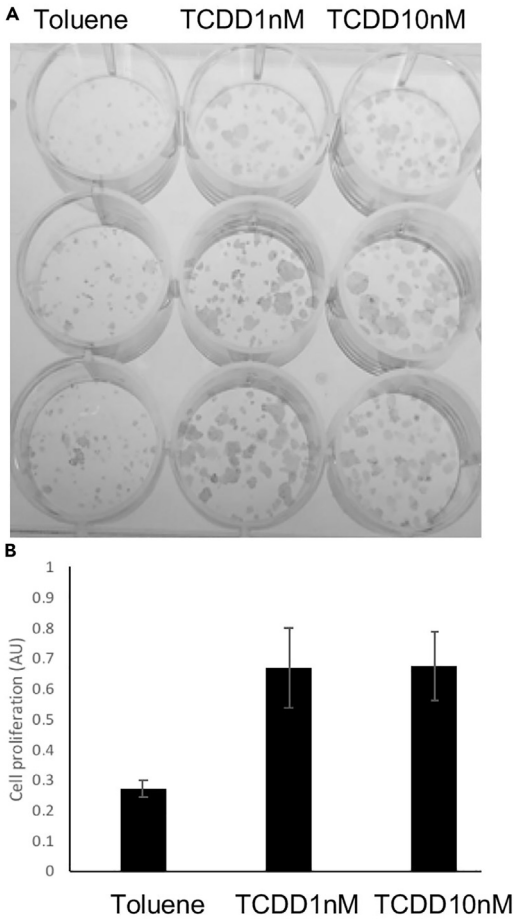


Figure 1. Neoplastic transformation with TCDD

(A) Crystal violet of cells showing more area of the culture plate covered by cells after exposure to TCDD 1 nM and 10 nM in comparison to Toluene vehicle control.

(B) Quantification of A, the error bar in the plot represents the standard deviation between multiple experimental replicates. An appropriate statistical test has to be followed in order to arrive at meaningful conclusions.

2. Resuspend cells in 1 mL of DMEM complete media and calculate the cell density of live cells using trypan blue vital dye and a cell counter.
3. Plate 1,000 live HACAT cells in each well of a 6-well plate containing 1.5 mL of DMEM complete media.
4. Incubate the cell plate in an incubator at 37°C with 5% CO2 for 2 days.
5. Aspirate media from each well of the 6-well plate.
6. Add 2 mL of 1 nM TCDD media into each 6 well plate, keeping 2-3 replicates.

   Note: For controls, use media with TCDD 10 nM and Toluene 10 nM (vehicle for TCDD) for reduced and no enrichment of mitochondria primed neoplastic stem cells, respectively.

7. Incubate the cell plate at 37°C with 5% CO2 for 16–20 days.
8. Replenish with the respective fresh media every 3–4 days to remove the dead cells.

   Note: There will be substantial cell death during the process of neoplastic transformation with TCDD. However, the total number of cells detected is expected to be substantially higher after exposure to both 1 and 10 nM TCDD as a result of increased cell proliferation due to neoplastic transformation (Figure 1).

9. Trypsinize the population of the emergent transformed cells.
10. Count the number of trypsinized cells and make a suspension of 1 cell per 200 μL. (troubleshooting problem 3).
11. Seed 200 μL cell suspension in each 96 well plate.
12. Incubate the cell plate at 37°C with 5% CO2 for 16–20 days or until the single transformed colonies emerge in each of the 96 well.
13. Select as many colonies as required and grow them individually in a flask/plate, after naming them appropriately.

Note: Not all the wells of 96 well plate will form colonies.

14. Cryopreserve the cultures obtained from individual clones separately.

Note: Successful transformation can be checked by the standard techniques of *in vitro* colony formation assay and *in vivo* xenograft assay.[1]

15. Quantify the abundance of neoplastic stem cells using a tumorsphere assay with Extreme Limiting Dilution Analysis statistics, as described in our previously published step by step protocol.[2]
16. Enrichment of neoplastic stem cells in the TF1 population can be confirmed by comparing the quantified abundance of stem cells (frequency) obtained in step 15 between clones transformed with 1 nM TCDD (TF1) or 10 nM TCDD (TF10) and the Parental population.
17. To confirm the unique mitochondrial structure related to mitochondrial priming in the TF1 population, perform the single cell mitochondrial structural analyses (described in our previously published step by step mito-SinCe[2] protocol[4]) on Mitotracker green stained cells (described in our previously published step by step protocol[5]).

Note: From this step onwards TF1 population will be considered as the mpSC enriched population, while the TF10 and P populations will be considered as neoplastic and parental non-neoplastic controls, respectively.

⚠ CRITICAL: For any experimental manipulation with Parental or TF1 and TF10 HACAT cells, allow 2 days after seeding on the appropriate culture surface for cells to achieve their healthy cell morphology. Failing to do this may introduce cell stress induced irreproducibility in the results.

Identification of neoplastic stem cells using single cell transcriptomic analyses

Timing: ~ 8 days or more

The aim of this part is to accomplish identification of distinct stem cell clusters (and others) within the population of TF1 (mpSC enriched), P (parental non-neoplastic control) and TF10 (neoplastic control) HACAT cells. This step is achieved by single cell RNA-seq approach.

18. Grow and process the TF1 (mpSC enriched), P (parental non-neoplastic control) and TF10 (neoplastic control) HACAT cells for performing scRNASeq on a 10× genomics platform (in house or core/outsourced facilities) as required.
    Refer to original paper[1] for basic details of generation and use of the scRNASeq data.
19. Obtain the FASTQ files to process through the following analytical pipeline.
20. Generate count matrix objects using Cell Ranger.

Note: This step performs alignment, filtering, barcode counting, and Unique Molecular Identified (UMI) counting. Quantitative measurements of the gene expression levels for individual

cells are obtained in the form of count matrix for each sample.

Input Files: FASTQ files for TF1, P, TF10 samples.

Note: Make sure that the FASTQ files are saved in the respective input path.

```
PATH:

/home/scrna_seq/cellranger/create_count_files/in/fastq_files

Output Files: Three files- barcodes.tsv, features.tsv and matrix.mtx for each sample in

directories named TF1, P and TF10 (zipped format)

PATH:

/home/scrna_seq/cellranger/cell_ranger_count/out/TF1/Filtered_features_by_bc_matrix/

/home/scrna_seq/cellranger/cell_ranger_count/out/P/Filtered_features_by_bc_matrix/

/home/scrna_seq/cellranger/cell_ranger_count/out/TF10/Filtered_features_by_bc_matrix/
```

a. To obtain the count matrix for each sample, run 'cellranger_count' function using the code in the box on the input files.

```
for i in 'ls /home/username/scrna_seq/cell_ranger_count/in/fastq_files/*_L001_I1_001.
fastq.gz'

do

filename='basename $i'

#echo $filename

samplename='echo $filename | cut -d_ -f 1'

#echo $samplename

qsub -V -cwd -o $samplename.log -b y -j y -pe smp 8 "cellranger count --id=$samplename --/home/user
name/scrna_seq/cellranger/refdata-gex-GRCh38-2020-A --fastqs=/home/scrna_seq/cell_ranger_
count/in/H2G7JBGXH --sample=$samplename --localcores=8 --localmem=32"

done
```

Note: It takes 1–2 h of wall time as a function of memory for a 20k cells high-throughput dataset run by allocating more than the minimum of 64GB memory to the pipeline.

b. Pull up 'filtered_feature_by_bc_matrix' directory from the 'out' folder and compress the 'outs' folder of each sample using the code in the box on the files generated in the previous step.

```
for i in 'find . -maxdepth 1 -mindepth 1 -type d'

do

  bn='basename $i'

  rm -rf $i/_*

  rm -rf $i/$i.mri.tgz

  rm -rf $i/SC_RNA_COUNTER_CS

  #if [ ! -e /home/username/scrna_seq/cellranger/cell_ranger_count/$bn ]; then
```

```
  #  ln -sr $i/outs/filtered_feature_bc_matrix /home/username/scrna_seq/cellranger/cell_-
ranger_count//$bn

  #fi

  cp -vr $i/outs/filtered_feature_bc_matrix $i/

  tar -cv $i/outs/* | xz -T0 >$i/$i.outs.xz

done
```

21. Set up unprocessed Seurat objects.

Note: This step creates a representation of single-cell expression data for R

Input Files: Compressed count matrix objects for each sample- TF1, P and TF10 as generated from previous step.

```
PATH:

/home/scrna_seq/cellranger/cell_ranger_count/out/TF1/outs.gz

/home/scrna_seq/cellranger/cell_ranger_count/out/P/outs.gz

/home/scrna_seq/cellranger/cell_ranger_count/out/TF10/outs.gz

Output Files: Individual Seurat objects- TF1, P and TF10 (rds format)

PATH:

/home/scrna_seq/Seurat/Create_individual_SeuratObject/out/TF1.rds

/home/scrna_seq/Seurat/Create_individual_SeuratObject/out/P.rds

/home/scrna_seq/Seurat/Create_individual_SeuratObject/out/TF10.rds
```

To create the individual Seurat objects run the following set of codes from the box on the input files.

```
library(Seurat)

files<-list.dirs("/home/username/scrna_seq/cellranger/cell_ranger_count/out/",recursive=
FALSE)

data <- list()

for ( i in 1:length(files)){d <- Read10X(files[[i]])

data[[i]] <- CreateSeuratObject(counts = d, project = "scrna_seq", min.cells = 3, min.
features = 200)}

names(data) <- basename(files)
```

22. Pre-Process the objects.

Note: This step transforms raw data into a format that is suitable for downstream analysis, such as clustering, dimensionality reduction, and differential gene expression analysis.

Input Files: Individual Seurat objects- TF1, P and TF10 generated from the previous step (rds format)

```
PATH:

/home/scrna_seq/Seurat/Create_individual_SeuratObject/out/

Output Files: Processed individual Seurat objects- TF1, P and TF10 (rds format)

PATH:

/home/scrna_seq/Seurat/Create_individual_SeuratObject/out/TF1.rds

/home/scrna_seq/Seurat/Create_individual_SeuratObject/out/P.rds

/home/scrna_seq/Seurat/Create_individual_SeuratObject/out/TF10.rds
```

To perform standard preprocessing, run the following set of codes from the box on the input files.

```
# Filter datasets based on nFeature_RNA

clean_data <- list()

for (i in 1:length(data)) {clean_data[[i]] <- subset(data[[i]], subset= nFeature_RNA > 2000 )

#clean_data[[i]] <- data[[i]] }

# Normalize data

for (i in 1:length(clean_data)) {clean_data[[i]] <- NormalizeData(clean_data[[i]],verbose = F)

# Detection of Highly Variable Features

clean_data[[i]] <- FindVariableFeatures(clean_data[[i]], selection.method = "vst", nfeatures = 2000,verbose = F)}

names(clean_data) <- basename(files)

#Save individual Seurat Object as 'rds'

for (i in 1:length(clean_data)) {

  name <- names(clean_data)[i]

    outfile <- outfile_name( paste0(name,".rds"))

      saveRDS(clean_data[[i]], file =

paste0("/home/username/scrna_seq/Seurat/Create_individual_Seurat/out/",

outfile), compress = "xz")}
```

⚠ CRITICAL: QC metrics can be customized based on user defined criteria, one of which is based on higher percentage coverage of mt-DNA reads in dying cells. Here, refrain from using this filter to avoid exclusion of detecting mt-DNA gene expression. Instead, use 'nFeature_RNA' function to filter the cells with unique feature counts over 2000. Importantly, confirm that the chosen filtering method excludes cells with >10% of mt-DNA gene expression (indicating dead cells) by using the following code in the box.

```
for (i in 1:length(data)) {data[[i]][["percent.mt"]] <- PercentageFeatureSet(data[[i]],
pattern="^MT-")}

VlnPlot(data[[1]], features = c("nFeature_RNA","nCount_RNA", "percent.mt"))

#Repeat the step highlighted in red for each dataset (saved as compressed count file)
```

23. Dimensionality reduction and Clustering using integrated Seurat Object.

Note: This step compresses the data to low-dimensional space to form biologically meaningful cell clusters based on the similarity of their gene expression.

Input Files: Processed Seurat objects- TF1, P and TF10 from previous output (rds format).

```
PATH:

/home/scrna_seq/Seurat/Create_individual_SeuratObject/out/

Output Files: Integrated Seurat object for TF1, P and TF10 (rds format)

PATH:

/home/scrna_seq/Seurat/integrate_Seurat/out/integrate_seurat_TF1_P_TF10.rds
```

To perform dimensionality reduction, clustering upon integrated Seurat object, run the following set of codes from the box using the input file:

```
# Set options for output file and caching

knitr::opts_chunk$set(echo = TRUE)

knitr::opts_chunk$set(cache = TRUE)

VERSION <- 20210204 (can be changed as per user's choice)

# Function to create output file names with version number

outfile_name <- function (filename) {

  return (paste(VERSION,filename, sep = "-"))}

# Function to save objects in xz format

saveRDS.xz <- function(object,file) {con <- pipe(paste0("xz -T0>",file),"wb")

  saveRDS(object, file = con)

  close(con)}

#Setup for output file

knitr::opts_chunk$set(echo = TRUE)

knitr::opts_chunk$set(cache = TRUE)

VERSION <- 20210204

outfile_name <- function (filename) {

return (paste(VERSION,filename, sep = "-"))}

saveRDS.xz <- function(object,file) {

con <- pipe(paste0("xz -T0>",file),"wb")

saveRDS(object, file = con)

close(con)}

EVAL <- 1

library(Seurat)

files <- list.files("./in", full.names = T)
```

```
names <- basename(files)
names <- gsub("^\\d+-", "", x = names)
names <- gsub(".rds", "", x = names)
data <- list()
#Loop function for reading the input files
for (i in 1:length(files)) {data[[i]] <- readRDS(files[[i]])}
names(data) <- names
for (i in 1:length(files)) {data[[i]][["sample"]] <- names[[i]]}
names(data) <- names
library(future)
plan("multiprocess", workers= 16)
library(ggplot2)
#Integration of processed individual Seurat Objects
anchors <- FindIntegrationAnchors(object.list = data, dims = 1:20)
integrated <- IntegrateData(anchorset = anchors, dims=1:20)
integrated <- ScaleData(integrated, verbose = F)
#Dimensionality Reduction
integrated <- RunPCA(integrated, npcs=20, verbose = F)
integrated <- RunUMAP(integrated, reduction = "pca", dims=1:20)
outfile_name(paste0("integrated_seurat_TF1_P_TF10_", i, ".Rds")))
p <- DimPlot(integrated, reduction = "umap", split.by = "sample", label = TRUE)
integrated <- FindNeighbors(integrated, reduction = "pca", dims = 1:20)
#Clustering
integrated <- FindClusters(integrated, resolution = 0.2)
res <- c(0.1, 0.2, 0.3)
for (i in res) { integrated_tmp <- FindClusters(integrated, resolution = i, verbose = F)
p <- DimPlot(integrated_tmp, reduction = "umap", split.by = "sample", label = T)
ggsave(outfile_name(paste0("combined_plot_relevel_", i, ".pdf")), plot = p, width = 11,
height = 8.5)
#Save the integrated Seurat Objects
saveRDS.xz(integrated_tmp, file =
```

Note: In our analysis, resolution of 0.2 separated out stem cell markers in a well-defined separate cluster. If needed, resolution can be increased to identify the stem cell cluster, maximally to 0.5.

24. Marker Identification for differentially expressed genes (troubleshooting problem 5).

Note: This step identifies positive and negative markers for each cluster defined by differential gene expression, compared to all other cells.

Input Files: Integrated objects from the previous output (rds format)

```
PATH:

/home/scrna_seq/Seurat/integrate_Seurat/out/integrate_seurat_TF1_P_TF10.rds

Output Files: Single table of differentially expressed genes (tsv format)

PATH:

/home/scrna_seq/Seurat/FindAllMarkers/out/all_markers_0.2.tsv
```

To perform FindAllMarkers analysis, run the following set of code from the box on the input files.

```
#FindAllMarkers

output_file_markers <- outfile_name(paste0("integrated_seurat_TF1_P_TF10_", i, ".Rds"))

all_markers <- FindAllMarkers(integrated_data, return.thresh = 0.02, verbose = FALSE)

write.table(all_markers, file = outfile_name(paste0("all_markers_TF1_P_TF10", i, ".tsv")),
quote = FALSE, sep = "\t", row.names = FALSE, col.names = TRUE)
```

25. Identifying the stem cell cluster identity type.

Note: A tailored manual in-house method for cell type identity for clusters was developed as detailed below. This step identifies specific cell type identities for cell clusters.

Input Files: all_markers (tsv format) from FindAllMarker output.
Output Files: Cell identity assigned to each cluster.
Follow these steps to identify the cell type identity for each cluster.

a. Obtain the output of FindAllMarkers in a .tsv format.
b. Import the file in excel.
c. Filter out data for each cluster.
d. Sort the data in descending order of avg_logFC.
e. Check the functionality for topmost candidate genes. This gene set marks the cluster as 'cell type high status.'
f. Likewise, check the functionality for bottommost candidate genes. This gene set marks the cluster as 'cell type low status.'
g. The screenshots show filtered data for cluster 1 arranged in descending and ascending order with their cluster identity as stem-cell^{high} status and cell-cycle^{low} status based on functionality of topmost and bottommost genes respectively (Figure 2).

26. Quantifying stem cell clusters.

Note: Quantification of stem cell clusters can be done by visualization of cell clusters, pie charts, violin, and feature plots and heatmap as detailed in this section (Figure 3).

a. Cell Clusters and pie charts.

Note: This step identifies any differences in the distribution of cell clusters across samples and provides insights into the heterogeneity and composition of cell populations within each sample.

A

p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster	gene
6.53E-36	2.26633151262247	0.531	0.341	1.31E-32	1	SCGB1A1
8.50E-263	2.02062189138781	0.963	0.732	1.70E-259	1	KRT13
0	1.80669759073893	0.994	0.9	0	1	KRT15
8.81E-128	1.72549400653098	0.811	0.621	1.76E-124	1	S100A7
6.04E-223	1.67907357592272	0.878	0.574	1.21E-219	1	S100A8
1.61E-173	1.66410722407121	0.974	0.869	3.21E-170	1	S100A9
1.14E-128	1.54750794690206	0.854	0.64	2.28E-125	1	PI3
8.91E-220	1.42078097067692	0.915	0.632	1.78E-216	1	SERPINB4
2.57E-178	1.31800824568715	0.884	0.587	5.14E-175	1	SERPINB3
1.45E-282	1.26371472511244	0.952	0.808	2.91E-279	1	TNFSF10

Lineage specific functional stem cell Marker

B

6.18E-152	-1.04440422349209	0.767	0.864	1.24E-148	1	TOP2A
1.91E-145	-1.05625952528413	0.781	0.843	3.82E-142	1	FST
4.94E-142	-1.08481151201204	0.751	0.836	9.88E-139	1	CCNB1
6.77E-191	-1.11008005695886	0.809	0.901	1.35E-187	1	TPX2
1.48E-234	-1.11275375038483	0.801	0.922	2.96E-231	1	CDKN3
3.79E-191	-1.12771938012876	0.73	0.847	7.59E-188	1	UBE2S
2.63E-306	-1.15148925648266	0.895	0.978	5.26E-303	1	TUBB4B
2.97E-118	-1.20998230629299	0.778	0.824	5.94E-115	1	UBE2C
0	-1.4167013359753	0.949	0.993	0	1	TUBA1B
3.12E-200	-1.6538093583748	0.961	0.981	6.25E-197	1	HIST1H4C

General cell cycle Marker

Figure 2. Screenshot for identifying the stem cell cluster identity type

(A) Output of FindAllMarker function from Seurat Package in descending order of average log fold change (avg_logFC) to identify lineage specific functional stem cell markers in specific cluster (boxed in red).

(B) Output of FindAllMarker function from Seurat Package in descending order of average log fold change (avg_logFC) to identify general cell cycle markers specific cluster (boxed in red).

```
Input Files: Integrated Seurat object file (rds format)

PATH:

/home/scrna_seq/Seurat/integrate_Seurat/out/integrate_seurat_TF1_P_TF10.rds

Output Files: Pie chart (pdf format)

PATH:

/home/scrna_seq/Seurat/out/piechart/
```

To generate cell clusters and pie charts, run the following set of code from the box using the input files:

```
#Cell Clusters

  seurat <-readRDS("/home/username/scrna_seq/Seurat/integrate_Seurat/out/integrate_seurat
_TF1_P_TF10.Rds")

p <- DimPlot(seurat, reduction = "umap", split.by = "sample", label = TRUE)

ggsave(outfile_name("combined_plot.pdf"), plot = p, width = 11, height = 8.5)

# PieCharts

# Create count table (value for 'integrated_snn_res' will be same as the set resolution)

library(dplyr)

data <- seurat

metadata <- data@meta.data
```

```
counts <- metadata %>%
  group_by(sample, integrated_snn_res.0.2) %>%
  summarise(count = n())
counts
#Generate Pie Chart
library(cowplot)library(scales) library(ggplot2)
#Replace the sample names TF1, P and TF1 with the desired sample names as per your choice
samples <- c ("TF1", "P", "TF10")
my_color <- hue_pal()(length(levels(
     as.factor(counts$integrated_snn_res.0.2)
   )) + 1)
plots <- list()
for (i in samples) {
  d <- counts %>% filter (sample == i)
  plots[[i]]<-ggplot(d, aes(x="", y=count, fill=integrated_snn_res.0.2)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) + theme_void() + ggtitle(i) + labs(fill=
"Cluster") + scale_fill_manual(values = my_color)
}
combined_plot <- plot_grid(plotlist = plots,nrow = 1)
combined_plot
save_plot(paste0("out/piechart/", outfile_name("pie_chart.pdf")), plot =
combined_plot, base_width = 6.6)
```

b. Violin and Feature Plots.

Note: This step compacts visual summary of the distribution of gene expression, including the shape of the distribution and any outliers through violin plots. Also, visualization of the differentially expressed genes between different conditions or treatments through feature plots.

Input Files: Integrated Seurat object file (rds format).
PATH: /home/scrna_seq/Seurat/integrate_Seurat/out/integrate_seurat_TF1_P_TF10.rds.
Output Files: Feature plots and violin plots for a geneset (pdf format).

```
PATH:
/home/scrna_seq/out/vlnplot/
/home/scrna_seq/out/featureplot/
```

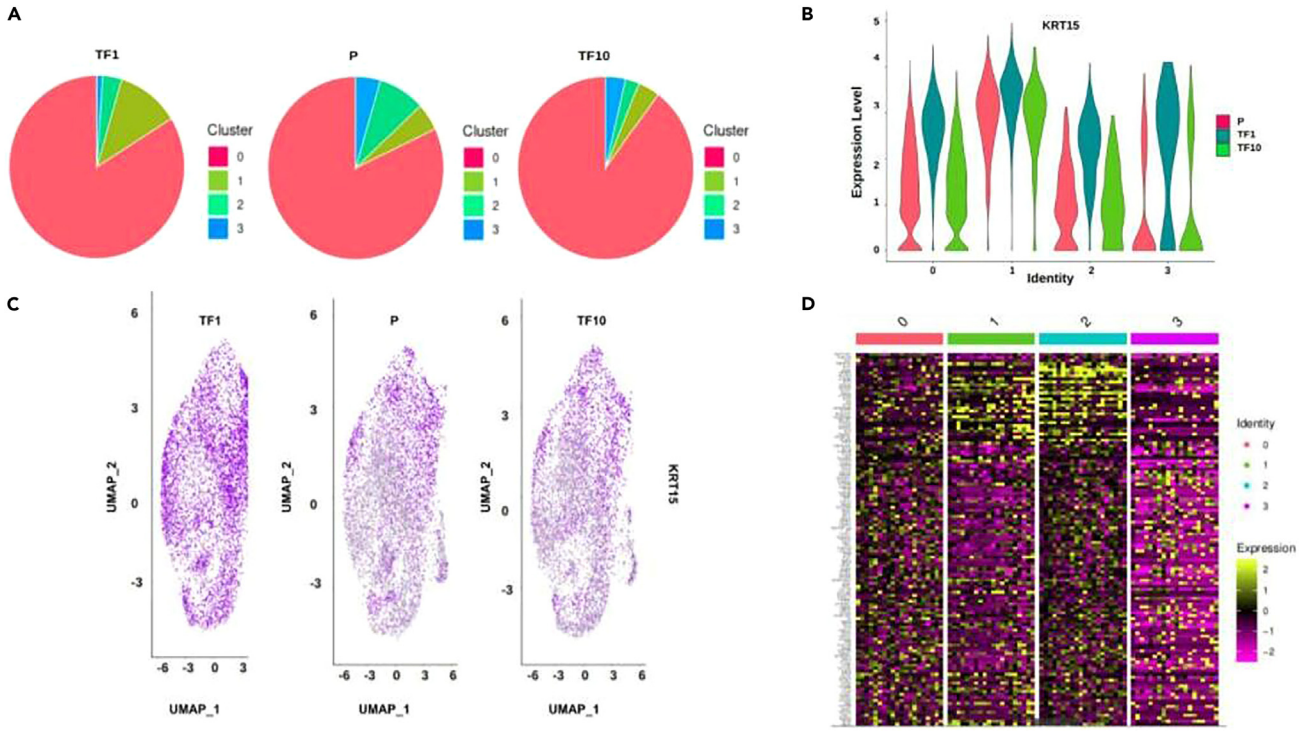


Figure 3. Data visualizations for quantifying stem cell clusters

(A) Pie chart of the percentage distribution of the scRNASeq derived clusters for TF1, P and TF10 populations with color codes.

(B) Violin plot depicting expression of stem cell marker KRT15 for the scRNASeq derived clusters identified in (A).

(C) Feature plot of KRT15 based on UMAP projection on the populations- TF1, P and TF10.

(D) Heat map of significant marker genes for the color-coded scRNASeq derived clusters as obtained in (A).

To generate violin and feature plots, run the following set of code from the box using the input files.

```
#Violin Plot and Feature Plot Generation

mkdir /home/username/scrna_seq/Seurat/out/vlnplot/

/home/username/scrna_seq//Seurat/out/featureplot/

seurat <-

readRDS("/home/username/scrna_seq/Seurat/integrate_Seurat/out/
integrate_seurat_TF1_P_TF10.Rds")

library(patchwork)

library(cowplot)

library(ggplot2)

library(RColorBrewer)

library(Seurat)

# Replace the 'gene1', 'gene2' with genenames of interest

gene_list <- c("gene1","gene2")

data <- seurat

DefaultAssay(data) <- "RNA"
```

```
plot_list <- list()

for (i in gene_list) {plot_list[[i]] <- FeaturePlot(data, features = i, split.by = "sample",
cols = c("grey","blue"))}

vplot_list <- list()

data$clusternum <- Idents(data)

for (i in gene_list) {vplot_list[[i]] <- VlnPlot(data, features = i, split.by = "sample",
group.by = "clusternum", pt.size = 0, combine = FALSE)}

for (i in gene_list) {

ggsave(paste0("/home/username/scrna_seq/Seurat/out/featureplot/","map_", i, "0.2_raw.
pdf"), plot = plot_list[[i]], width = 11, height = 8.5)

pdf(paste0("/home/username/scrna_seq/Seurat/out/vlnplot/","vin", i,

"_0.2_raw.pdf"), width = 11, height = 8.5)

print(vplot_list[[i]])

dev.off()}
```

c. HeatMap.

Note: This step identifies patterns of gene expression that are associated with different cell states or functions.

Input Files: Integrated Seurat object file (rds format).

```
PATH:

/home/scrna_seq/Seurat/integrate_Seurat/out/integrate_seurat_TF1_P_TF10.rds
```

Output Files: Heatmap with a dynamically generated file name (pdf format).

```
PATH:

/home/scrna_seq/Seurat/heatmap/out/
```

To create a heatmap, run the following set of code from the box using the input files:

```
# Heatmap Generation

library(future)

plan("multicore", workers= 16)

options(future.globals.maxSize = 2000 * 1024^2)

markers <- FindAllMarkers(seurat, logfc.threshold = 0.1, return.thresh = 0.05)

markers <-

read.table("/home/username/scrna_seq/Seurat/FindAllMarkers/out/all_markers_0.2.tsv",
header = T)

selected_markers <- markers[ abs(markers$avg_log2FC) > 0.5, ]

library(ggplot2)
```

```
set.seed(1235)

seurat_small <- subset(seurat, downsample = 500)

hm <- DoHeatmap(seurat_small, features = unique(as.character(selected_markers$gene)))

hm <- hm + theme(text = element_text(size = 10))

hm <- hm + theme(axis.text.y = element_text(size = 4))

ggsave(file.path("out", outfile_name("heatmap_all_markers.pdf")), hm, height = 10,
width = 7.5)
```

Note: Using altered codes, heatmaps can be generated for specific cluster sets with select p value. The Seurat object and selected markers should be defined before running this procedure.

Identification of mitochondria primed neoplastic skin stem cells using single-cell microscopy

Timing: 6–7 h

The aim of this part is to accomplish identification of Krt15^{hi}Sox2^{hi} cells within the population of TF1 (mpSC enriched), P (parental non-neoplastic control) and TF10 (neoplastic control) HACAT cells. This step is achieved using single cell microscopy analyses of co-immunostained cells.

27. Immunofluorescence/Immunocytochemistry

Note: All the experimental procedures are done on the sub confluent culture of cells to ensure cytosolic areas do not overlap (but may touch each other). The steps described below should be followed to obtain the optimal dilution of individual antibodies that should be subsequently used for co-immunostaining. An 8-well Nunc Labtek chamber (or equivalent, see KRT) slide requires a minimum of 100 μL of any solution to cover the entire surface, while more than 300 μL is not necessary. All the incubations and washing are done by gently rocking the chamber slide on a rocker.

a. Prepare the culture flasks for TF-1 (mpSC enriched), P (parental non-neoplastic control) and TF-10 (neoplastic control) HACAT cells.
b. Prepare to perform 3 co-immunostaining combinations for each group of cells (TF-1/P/TF-10) as per Table for Antibody Combination mentioned in Step 14 of the before you begin section.
c. Label one 8 well Nunc Labtek chamber slide for each co-immunostaining combination (see Table for Antibody Combination in Step 14 of the before you begin section). For each chamber, label 3–4 wells for co-immunostaining with one combination of antibodies, one well for individual staining with each antibody, one well for secondary antibody control (without primary antibody) and one well for autofluorescence control.

Note: Keeping appropriate controls may help in effective troubleshooting along the way. The individually stained samples are for optimization and thus does not need to be repeated with every experiment that uses the same samples, antibodies, and image settings.

d. Seed around 8k to 10k cells in each well of 8 well Nunc Labtek chamber slide.
e. Allow cells to grow and attain their healthy cell morphology for at least 2 days.

⚠ CRITICAL: It is essential to ensure that the number of cells maintained in each well is comparable (same ballpark) to avoid any inconsistencies of staining intensity that may result from stoichiometric differences in antigen-antibody interactions. This can be objectively estimated after 2 days of seeding by counting the number of cells in 5–6 representative

field of views from each well using standard bright-field or transmitted light tissue culture microscope before processing further. In situations where cells of different proliferation rate are being considered, a smaller number of cells should be seeded for the cell population with higher proliferation rate to achieve the similar cell number after 2 days of seeding between all samples of interest.

f. Fix cells in freshly prepared pre-heated fixing solution supplemented with (4%, w/v) sucrose for 15 min at 37°C.

Note: Preheating should be done at 37°C for no more than 5–10 min.

⚠ CRITICAL: Adding sucrose to the fixing solution and pre-heating it to 37°C helps in maintaining mitochondrial morphology.

g. Wash *cells* with 1×-PBS thrice for 5 min each. (troubleshooting problem 6).
h. Permeabilize cells in freshly prepared Triton X-100 (0.1%) for 15 min at 37°C.
i. Wash *cells* with 1×-PBS thrice for 5 min each.
j. Block cells by incubating them in freshly prepared blocking solution for at least 1 h at 37°C.

Pause point: At this point protocol can be paused and blocking can be done for 12–14 h at 4°C

k. Incubate cells in appropriate primary antibody solution for 1–2 h (or as optimized) at 37°C.

Pause point: At this point, the protocol can be paused, and primary antibody incubation can be done for 16–24 h at 4°C.

l. Wash cells with 1×-PBS thrice for 5 min each.
m. Incubate in appropriate secondary antibody solution for 1 h at 37°C.

⚠ CRITICAL: The choice of primary and secondary antibodies should be made carefully to avoid cross-reactivity between the antibodies and spectra overlapping of the fluorophores conjugated to the secondary antibody. All dilutions, incubation times, and blocking solution can vary from antibody to antibody and thus should be optimized for any other primary and secondary antibody employing necessary controls.

n. Wash cells with 1×-PBS thrice for 5 min each.
o. Remove the chamber wells using the tool provided in the box of the chamber by following the manufacturer's instructions.
p. Put 2 to 3 drops of mounting medium on the slide and cover the slide using a coverslip.

⚠ CRITICAL: Ensure all the wells are properly covered by a coverslip, if not the cells on the slide might dry.

q. Allow it to dry and seal the edges of the coverslip using transparent nail polish.
r. Allow the nail polish to dry.

Pause point: At this point protocol can be paused and the final slide can be kept at 4°C for some days to a couple of months.

28. Confocal Microscopy.

Note: Confocal microscopy was performed on a laser scanning confocal microscope, Zeiss LSM700 microscope equipped with 40× Plan apochromat 1.4NA/Oil objective and

405 nm, 488 nm, 555 nm, and 639 nm lasers, PMT detectors with appropriate emission filters. Any other confocal microscope with the corresponding software for image acquisition can be used while maintaining the parameters described below.

a. Switch on the microscope at least 15 min before image acquisition to allow the system to stabilize.
b. Prepare to perform confocal microscopy on 3 slides with 3 co-immunostaining combinations in each group of cells. (see Table for Antibody Combination in Step 14 of the before you begin section).

Pause point: Pauses can be allowed after scanning of each slide

c. Design image acquisition parameters for 8-bit images of 512 × 512 pixel density, an optical zoom of 0.5 with 1 airy unit pinhole and about 1 μm Z interval with the total number of slices covering the cells from top to the bottom (~10–15 μm).
d. Design 3-channel image acquisition with appropriate lasers and filters to excite the DNA dye and the two secondary antibody fluorophore conjugates for detecting the molecule of interest(s).
e. The laser intensity and detectors should be adjusted such that minimal or no signal is obtained from the autofluorescence and the secondary only control samples, and detectable signal from all the cells that is expected to vary between cells.
f. Adjust offset parameters for each channel to set the baseline for signal intensity.
g. Ensure minimal cross-talk and cross-excitations between channels.
h. Use the optimized parameters to acquire 3D images (Z-stacks) from 5 to 10 different fields of view of the immunostained cells to include at least 600–700 cells for analysis from each experimental and control group.
i. Save *these* acquired Z-stacked images to be used for image analysis. (troubleshooting problem 7 & 8).

⚠ CRITICAL: Ensure to label each image appropriately to be able to track the source of the signal, given the same fluorophore combinations are being used for all the three co-immunostaining combinations.

29. Image processing and statistical analyses.

Note: Image processing and analysis of relative protein abundance and localization were performed using proprietary Zen Black and Zen Blue software and Microsoft Excel. It can also be performed using available open-source image analyses or statistical software and packages like IMAGE-J, MATLAB, R etc. and can also be automated, while maintaining the basic steps described below. The individually stained samples are for optimization and thus do not need to be repeated with every experiment that uses the same samples, antibodies, and image settings.

a. Prepare to analyze 3 co-immunostaining combinations for each TF-1, P and TF-10 group, where either the cytosolic or nuclear signal must be quantified for each molecule of interest (see Table for Antibody Combination in Step 14 in before you begin section).
b. Make an image subset from each of the acquired Z stack images to include 3 consecutive optical slices providing maximum coverage of the cells.
c. Obtain a maximum intensity projection (MIP) image of the Z subset.
d. In each MIP image, draw ROIs around single cells using a free-hand tool.
e. For each cell, draw one ROI to demarcate the nucleus based on DNA stain and one ROI to demarcate the cytosolic region (without nucleus) based on the immunostain of the cytosolic molecule of interest.

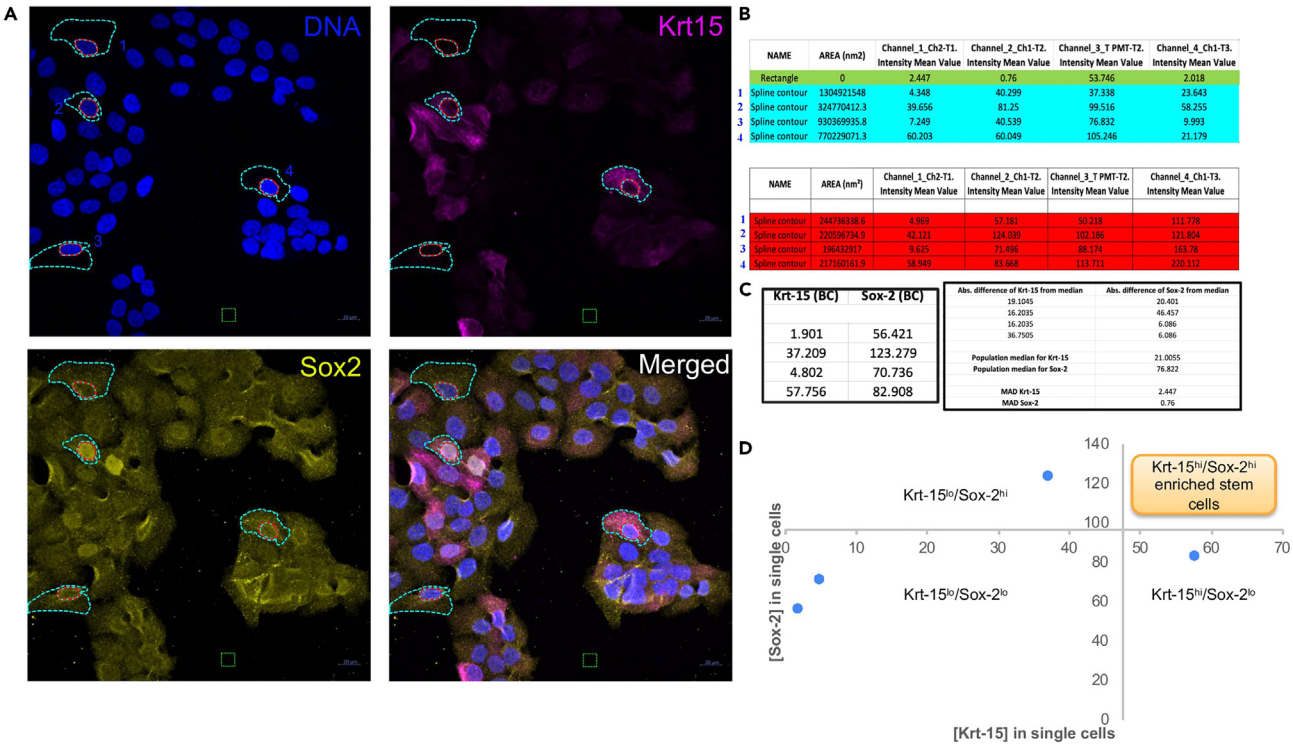


Figure 4. Single cell immunofluorescence analyses of stem cell markers

(A) Representative maximum intensity projections of a field of view HACAT cells co-immunostained for Krt15 and Sox2; analyses are exemplified by 4 randomly chosen cells are marked with color coded ROIs (turquoise: cytosolic without nucleus and nucleus: pink); scale bar is 20 μm.

(B) MFI of various channels from color coded ROIs calculated from Zen-blue software. (Channel-1 Krt-15; Channel-2 Sox-2, Channel-3 Bright Field Channel-4 Hoechst/Nucleus).

(C) Example table showing background corrected MFI along with tables used to calculate MAD values as described in the protocol.

(D) Graph for identification of Krt-15^{hi}Sox-2^{hi} cells generated based on calculations of only 4 randomly chosen cells along with thresholding values plotted for quadrant division (actual statistics will be obtained with analyses of all the cells from all the micrographs acquired for any co-immunostaining experiment).

Note: If necessary, the image intensity can be adjusted appropriately to reveal the very weak signal in cells.

f. Draw a rectangular ROI anywhere between cells where there is a dark background to obtain background signals from different channels.
g. Obtain the mean fluorescent intensity (MFI) from each channel of each ROI, including the background ROI.
h. See example, cited (Figure 4) for Krt-15 and Sox-2.
i. To obtain background-corrected mean fluorescent intensity for a single channel, subtract the mean intensity of the background for that particular channel from the mean intensity from the region of interest (ROI).
j. Confirm that the mean and SD of the signal intensity of individually stained samples are comparable to that in the co-stained sample.
k. Generate a scatter plot with the background corrected MFIs of marker signal in channel(s) 1 and 2.
l. Decide the groups for comparative analyses for performing statistical calculations described below. The comparative groups must be considered as a single population.
m. To obtain the signal thresholding value for identifying the marker defined cell group of interest (ex: Krt15^{hi}/Sox2^{hi}), use the Median Absolute Deviation (MAD) value of each relevant marker channel.

Note: MAD is the median of the absolute difference of individual value from the population median

n. MAD can be calculated using the following steps.
   i. Calculate the median of the population for both markers of interest from background corrected mean intensity.
   ii. Get the absolute difference from the median of each ROI of both the marker signals by subtracting each value from the respective population median.
   iii. Take the median of this absolute difference for the entire population of each signal separately to get MAD values for the respective marker.

⚠ CRITICAL: Groups to be compared must be considered as a single population to be able to obtain and apply the same thresholding values to the scatter plot for each analytical group in comparison.

o. Threshold the signal from both markers using the formulae = [Median] + 1.5× [MAD]. Based on the values, marking both axes to get the 4 quadrants in the graph.
p. Perform the Chi-square analysis to obtain the p values for the distribution of the cells in the 4 quadrants using the following steps:
   i. Organize data in a contingency table for the variables of interest by listing all the levels of one variable as rows and correspondingly all the levels of other variables as columns.
   ii. Create a table with the observed frequencies for each combination of variables.
   iii. Calculate the expected frequencies for each cell in the contingency table using the formula: expected frequency = (row total × column total) / grand total.
   iv. Calculate the chi-square statistic using the formula: chi-square = SUM (observed frequency - expected frequency)^2/ expected frequency) to get the p-value for the chi-square test.

EXPECTED OUTCOMES

For Part 1, the primary outcome is generation of a transformed cell population with enrichment of stem cells by exposing the parental cells with low dose of carcinogen. Two weeks exposure of HACAT cells to TCDD 1 or 10 nM colonies increases cell proliferation in comparison to similar exposure to Toluene vehicle control (Figure 1). Such an outcome is indicative of the first round of neoplastic transformation *in-vitro*. In the next step, the freshly transformed clones selected from the 96 well plate (in presence of 1 or 10 nM TCDD) and expanded in larger culture dishes should be used for characterization at various levels before cryofreezing for storage. Increase in abundance of self-renewing cells from that of the parental cell population will confirm stem cell enrichment. The transformed cells also should maintain enhanced cell proliferation from that of the parental cell population.

For Part 2a, the primary outcome is identification of distinct cell clusters using scRNASeq analyses, one of which is marked by enhanced levels of Krt15 (and other lineage specific stem cell markers) within the population of Parental and TCDD transformed cells as quantitatively monitored in data exported to a spreadsheet (Figure 2A) or visually depicted in a UMAP plot.[1] The reduced levels of cell cycle markers in the same cluster signifies cell cycle quiescence of the stem cell cluster (Figure 2B). Importantly, the clones transformed by 1 nM TCDD are expected to have significantly higher abundance of the Krt15 marked stem cell cluster, as depicted by the pie chart (Figure 3A). Moreover, violin plots and feature plots would demonstrate that 1 nM TCDD transformed samples maintains elevated Krt15 levels (and other relevant stem cell markers) across all the clusters, while the Krt15 levels are prominently high in the designated stem cell cluster even in the control transformed and non-transformed populations (Figures 3B and 3C). The heat map shows the overall transcriptomics profile of individual cell cluster (Figure 3D).

For Part 2b, the primary outcome is identification of Krt15^{hi}Sox2^{hi} cells using single cell microscopy analyses of immunofluorescence, where the Sox2^{hi} cells have pDrp1(S616)^{lo} and pCycE(T62)^{lo} status (see Table for Antibody Combination in Step 14 of the before you begin section) . The confocal micrographs from the samples successfully processed for immunofluorescence would yield images with two antibody fluorophore channels depicting the relevant antibody combinations, and one DNA stain (Figure 4A). Appropriate ROIs drawn on the images would yield signal intensity values of the cells and the background (Figure 4B). The background corrected intensity values from the nuclear or cytosolic ROIs (as appropriate for each protein marker) would yield the threshold values (Figure 4C) that would determine the 4 quadrants towards identifying the Krt15^{hi}Sox2^{hi} cells (Figure 4D), Sox2^{hi}pDrp1(S616)^{lo} or Sox2^{hi}pCycE(T62)^{lo} cells (not shown).

LIMITATIONS

The protocol describes the use of one carcinogen on one cell line in enriching the mpSCs. There may be other methods of achieving the same goal in other cell types that would enrich the yield of mpSCs with higher self-renewal ability and fine-tuned repression of Drp1 driven mitochondrial fission. Successful execution of the whole protocol requires other protocols that have been published elsewhere by our group. scRNASeq makes it more difficult to detect low abundant transcripts with low capture efficiency and high dropouts. This protocol is a custom-made workflow for scRNASeq analysis and hence, expectation of an exhaustive list of analysis tools is beyond the scope of this protocol.

TROUBLESHOOTING

Problem 1

Code is not running.

Potential solution

Check if:

- R is active.
- required packages are loaded.
- the input file format and location are proper.
- the code is typed keeping case sensitivity of R/Terminal in mind.
- the code is incomplete.

Problem 2

Could not install Seurat directly (Related to Step 6a- before you begin).

Potential solution

Update R to version $\geq$ 4.2.2 and then re-install the Seurat package.

Problem 3

None or few transformed cells obtained (Related to Step 10 - step-by-step method details).

Potential solution

Increase the initial number of cells to be treated with the carcinogenic agent.

Problem 4

If the packages give non-zero exit status.

Potential solution

- Check the error source and see if any dependencies are missing or have an error, rerun the installation of the package with the code:

```
>install.packages(''PackageName'', dependencies = TRUE)
```

- Update the library fftw3-dev with the code:

```
>sudo apt-get install fftw3-dev
```

Problem 5

Could not find function "outfile_name" (Related to Step 24- step-by-step method details).

Potential solution

Run the following set of code from the box prior to running the code for the step:

```
knitr::opts_chunk$set(echo = TRUE)

knitr::opts_chunk$set(cache = TRUE)

VERSION <- 20210204

outfile_name <- function (filename) {return (paste(VERSION,filename, sep = "-"))}

saveRDS.xz <- function(object,file) {con <- pipe(paste0("xz -T0>",file),"wb")

saveRDS(object, file = con)

close(con)}

EVAL <- 1
```

Problem 6

PFA fixation may lead to a high background. (Related to Step 27g- step-by-step method details).

Potential solution

Wash properly after fixation step and optimize based on the background observed.

Problem 7

Initial signal from immunostained cells is very low. (Related to Step 28i- step-by-step method details).

Potential solution

- Stabilize the lasers by allowing them to be ON for 10–15 min before starting the experiment
- Reoptimize the antibody dilution. A good starting point for optimization would be the recommendation in the product data sheet.
- Check if the cell number is more in sample with lower signal, which may arise due to stoichiometric differences (see Critical point of step-by-step method details of Part 2b, 1-e).

Problem 8

Non-specific signals. (Related to Step 28i- step-by-step method details).

Potential solution

- Control the cross-excitation and cross-reactivity by taking appropriate measures.
- Optimize the blocking step in immunocytochemistry procedure.

Problem 9

Few cells remain after processing for immunofluorescence.

Potential solution

Redo the experiment and add the reagents at each step very gently on the wall of the wells to avoid washing off the cells from the wells.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kasturi Mitra (kasturi.mitra@ashoka.edu.in).

Materials availability

None.

Data and code availability

Raw, analyzed, and meta data are available in Gene expression omnibus (GEO: GSE171772).

ACKNOWLEDGMENTS

The original work was supported by NIH award (#R33ES025662), UAB Flow Cytometry and Single-Cell Services Core (supported by P30 AR048311 and P30 AI027667), and Dr Mike Crowley from the Heflin Genomics Core for scRNASeq. K.M. is currently supported by DBT/Wellcome Trust India Alliance Senior Research Fellowship IA/S/20/2/505198 and S.A. by DBT-RA-2022/July/NE/3010. We sincerely acknowledge Dr. Brian Spurlock and Dr. Malay Kumar Basu who performed the original research that was followed by S.A. and M.S. in writing the current protocol. We acknowledge Yoshitha Sriramkumar who performed a test run for all the codes in the current protocol.

AUTHOR CONTRIBUTIONS

D.P. wrote the draft of part 1, S.A. wrote part 2a, M.S. wrote part 2b, and K.M. finalized the draft and coordinated between the authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Spurlock, B., Parker, D., Basu, M.K., Hjelmeland, A., Gc, S., Liu, S., Siegal, G.P., Gunter, A., Moran, A., and Mitra, K. (2021). Fine-tuned repression of Drp1-driven mitochondrial fission primes a "stem/progenitor-like state" to support neoplastic transformation. Elife *10*, e68394. https://doi.org/10.7554/ELIFE.68394.

2. Spurlock, B., Hanumanthu, V.S., and Mitra, K. (2021). Strategy of isolating "primed" tumor initiating cells based on mitochondrial transmembrane potential. Bio. Protoc. *11*, e3945. https://doi.org/10.21769/BIOPROTOC.3945.

3. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/J.CELL.2019.05.031.

4. Spurlock, B., and Mitra, K. (2021). Mito-SinCe^2 approach to analyze mitochondrial structure–function relationship in single cells. Methods Mol. Biol. *2275*, 415–432. https://doi.org/10.1007/978-1-0716-1262-0_27/COVER.

5. Mitra, K., and Lippincott-Schwartz, J. (2010). Analysis of mitochondrial dynamics and functions using imaging approaches. Curr. Protoc. Cell Biol. *46*. Unit 4.25.1-21. https://doi.org/10.1002/0471143030.CB0425S46.