# **Project Plan: Full Text Screening - AI model**

Timeline in the method project:

Training sets 2 & 3 ready in December

Full abstract screening done around January?

Timeline of AI full text screening

What is the desired schedule, just so I know what to prioritize as I have another project as well.

The currently relevant topics are highlighted with orange.

## **Scope**

The goal is to assist screeners with inclusion/exclusion decisions based on full data.   
The project can be divided to two sub tasks:

1. Tagging the data
2. Use tagged data to exclude/include, automatically or with human in the loop.

Tagging can mean either word counts from pre-defined dictionary, or question answering/data answering with NLP. I will first focus on the possibilities with NLP.

**Input:**  Full pdf

**Input format for the model**: Pdf file with text. Image like pdfs are not considered yet.

How big proportion of the papers are image like pdfs where some off-shelf OCR should be applied? This can cause additional errors in the predictions. And if we take text like input, this means that Figures and Tables are not included.

**Desired output:**   
Tagging information with NLP  
PICO:

* Subject: [human, animal, in-vivo]
  + If animal, what species
  + If in-vivo, what cells
* Intervention / Exposure:
  + Radiation: [gamma, x, alpha,…]
  + Other interventions?
* Comparison
* Outcome

Decision Tree:

* Is literature review: y/n
* Includes ionizing radiation: y/n
* Only UV: y/n
* …

One possible approach is also to use some separate classifier to detect human/animal and in-vivo studies, and based on that apply the corresponding model for tagging.

**Output format:** JSON? .txt? .csv? and the structure of these outputs?

## **Data collection**

(abstract screening done manually)

**SD:**157 screened papers  
88 excluded papers  
Questions: (The numbers notate the number of papers that were annotated with this flag at least by one screener.)

* 'This article concerns '
  + Human (100)
  + Cells (36)
  + Animal (45)
  + More than one study type (5)
* 'Are you able to screen this article?'
  + Yes
  + No, there is something wrong with the pdf link
  + No, for another reason
  + No, the article is written in a language that I cannot read (5)
* 'Are you including or excluding this study?'
  + Include (92)
  + Exclude (88)
* 'Is the data reported separately by sex? '
  + FALSE
  + TRUE
* 'If excluding, which exclusion criteria applies to this study?'
  + '1. Study Design. Not a controlled study.' (43)
  + '5. Outcome. ' (8)
  + '4. Population. Age/Gender/Disease status etc. not matching our PICOS'(30)
  + '2. Radiation. No ionising radiation in intervention group.'(6)
  + '3. Co-intervention. Accompanying exposure (surgery accepted).' (12)
* 'Please give details'
  + 'Human and Animal', (This is sometimes annotated together with “Human” label)
  + 'cells and animal',
  + 'in vitro, in vivo and ex vivo studies',
  + 'Animal (frogs) and in vitro experiments both described. Only in vitro study is valid for data extraction.'
* 'Please provide details'
  + 'No PDF available'

**CNS:**Proposed question structure:

1. Does this article describe primary study? (reviews?)
2. 'This article concerns '
   * Human
   * Cells
   * Animal
   * More than one study type (can we make it multichoice?)
3. Does this article consider intervention or exposure with radiation?
4. Are the subjects not exposed to any other chemical … ? Are all the subjects exposed to other chemical, biological or physical agents then radiation? We do not consider surgeries as such agents.
5. Is there evidence for absence of a control group?
6. Is it only about neoplastic or cancer related effects?
7. 'This article concerns '
   * Human
   * Cells
   * Animal
   * More than one study type (can we make it multichoice?)

* Is the study literature review?
  + y/n
* Is there evidence that the study does not include ionising radiation health/biological effets?
  + y/n
* Is it only about UV radiation? / What type of radiation is considered?
  + ??
* Is the study about the effects on the CNS? / What effects does the study address? Is one of the outcomes investigated in the CNS?
  + y/n
* Is it only about neoplastic or cancer related effects?
  + y/n
* Is it only about anxiety levels/emotional distress?
  + y/n
* Is there evidence for absence of control group?
  + y/n
* Are all the subjects exposed to other chemical, biological or physical agents then radiation? We do not consider surgeries as such agents.
  + y/n

If human study

* How old are the study subjects?
* How many subjects are there?
* Is the study conducted on unborn/pregnant/deceased subjects?

If animal study

1. Which animals are considered?
2. Are the animals adults?
3. Are they healthy animals?
4. Is it comparing exposure to non exposure?
5. IS there a single group pre- vs. post-intervention?

If in-vitro

1. Which cells are considered?
2. Are the the cells from healhy subjects?
3. Are the cells from healthy animals?
4. Is it comparing exposure to non exposure?
5. IS there a single group pre- vs. post-intervention?

How were the pdfs extracted from Zotero? R Code provided by Edinburgh, run where and when?

## **DaTA Preprocessing**

#### Read paper

A close-up of a paper

Description automatically generated

If Image based pdf:

Use some OCR tool and proceed similarly.

A close-up of a paper

Description automatically generatedIf regular pdf with text:

1. Use pdfMinerSix to find bounding boxes of chapters   
   ([Welcome to pdfminer.six’s documentation! — pdfminer.six \_\_VERSION\_\_ documentation (pdfminersix.readthedocs.io)](https://pdfminersix.readthedocs.io/en/latest/))
2. Read the text using pyPDF2.
3. Find the most used formatting (font and size) in the document.
4. Only include the chapters where the most common formatting is used, and there is minimum of 5 letters (to exclude tables with only values)

* GROBIT tool could be used as well: [kermitt2/grobid: A machine learning software for extracting information from scholarly documents (github.com)](https://github.com/kermitt2/grobid)

#### Embed the paper

Embed all the included chapters and store the embeddings for later use.  
ChatGPT: $0.0001 / 1K tokens

#### Query

1. Select an appropriate prompt template
2. Find the most relevant chapters with K-nearest algorithm on the embedding space
3. Give these chapters as the context for the query. (Or make a summary based on the data)
4. Use e.g. ChatGPT API: $0.0030 / 1K tokens for the input

#### Sanity checking etc?

What to do if the model is unsure? Flag?

## **Possible methods**

Pre-Pico   
Abstract text 🡪 PICO labels  
 [Species, Intervention, Comparator, Strain, Induction, Outcome]  
([hemmlin/pre-pico-test (github.com)](https://github.com/hemmlin/pre-pico-test))  
From my initial exploration on the code it seems like data extraction from abstract  
Waiting for further Documentation  
BIRD based solutions might not be the best approach for long context: [2203.11258.pdf (arxiv.org)](https://arxiv.org/pdf/2203.11258.pdf)

Donut 🍩 Image 🡪 question answering, or data extraction to json[arxiv.org/pdf/2111.15664.pdf](https://arxiv.org/pdf/2111.15664.pdf)Nice open source pre trained model with easy api through HuggingFaceAlready working somehowCould be limited in the size of context?Similar approach to Microsoft DUBLIN [2305.14218.pdf (arxiv.org)](https://arxiv.org/pdf/2305.14218.pdf)

General purpose VDUse.g. LayoutLM, LayoutLMv2, BERT  
OCR 🡪 Transformer 🡪 Encoder

Spade  
[Spatial Dependency Parsing for Semi-Structured Document Information Extraction (aclanthology.org)](https://aclanthology.org/2021.findings-acl.28.pdf)similar to WYVERN  
graph based method that utilizes the OCR bounding boxes to obtain better performance  
Maybe not so useful with articles where the hierarchy of the text is quite simple.

PromptingText context 🡪 question answering.  
[lamini-ai/llm-classifier: Classify data instantly using an LLM (github.com)](https://github.com/lamini-ai/llm-classifier)  
- both zero- and few-shot methods available  
- requires docker 🡪 not possible to test in my work laptop.[2303.05352.pdf (arxiv.org)](https://arxiv.org/pdf/2303.05352.pdf)[2305.19835.pdf (arxiv.org)](https://arxiv.org/pdf/2305.19835.pdf)Would require paid version of ChatGPT API, but probably easy to implement and can be used for highlighting. Fully zero-shot method.Or using Llama or other open-source option.ChatGPT API pricing: [Pricing (openai.com)](https://openai.com/pricing)  
0.003$/1K tokens  
🡪 approx. 0.03$/paper + answer tokens.

Classifiers for long context  
[applsci-12-04554-v2.pdf](file:///C:\Users\linda.hemmann\Downloads\applsci-12-04554-v2.pdf)  
[2203.11258.pdf (arxiv.org)](https://arxiv.org/pdf/2203.11258.pdf)  
- bench mark  
- truncated BERT is quite good 🡪 is it wise to do this with AI?  
- should we just reuse the abstract classifier with random or ranked sentences from the paper?  
[Building a classification service with Llama2 in Python (mikulskibartosz.name)](https://mikulskibartosz.name/building-classification-service-with-llama2-in-python)  
[sh0416/llama-classification: Text classification with Foundation Language Model LLaMA (github.com)](https://github.com/sh0416/llama-classification/tree/master)  
-Llama2 based classifiers seem to be quite popular  
- unfortunately, the maximum context size is **2048 tokens**

### Other papers:

[Screening for in vitro systematic reviews: a comparison of screening methods and training of a machine learning classifier | Clinical Science | Portland Press](https://portlandpress.com/clinsci/article/137/2/181/232436/Screening-for-in-vitro-systematic-reviews-a)

**LLM for interpreting tables:**[2305.14336.pdf (arxiv.org)](https://arxiv.org/pdf/2305.14336.pdf)  
[2001.01469.pdf (arxiv.org)](https://arxiv.org/pdf/2001.01469.pdf)

**OCR libraries:**Pytesseract: [pytesseract · PyPI](https://pypi.org/project/pytesseract/)

easyOCR: [JaidedAI/EasyOCR: Ready-to-use OCR with 80+ supported languages and all popular writing scripts including Latin, Chinese, Arabic, Devanagari, Cyrillic and etc. (github.com)](https://github.com/JaidedAI/EasyOCR)

PyPDF: [pypdf · PyPI](https://pypi.org/project/pypdf/)

LangChain: [Introduction | 🦜️🔗 Langchain](https://python.langchain.com/docs/get_started/introduction.html)  
- works both with image and text pdfs

trOCR: [TrOCR (huggingface.co)](https://huggingface.co/docs/transformers/model_doc/trocr)

docTR: [docTR documentation (mindee.github.io)](https://mindee.github.io/doctr/)  
- Gives the spatial information and the text, can be used to recreate the document for highlighting etc.

PaddleOCR: [paddleocr · PyPI](https://pypi.org/project/paddleocr/)  
- gives bounding box, for highlighting  
- English and Chinese

Clova OCR   
- commercial

## **Results**

Clearly define the desired evaluation statistics and benchmarking metrics based on the project's objectives. What will be considered a successful outcome for this project?

Without any tuning the human/animal/cell classification performs as:

Confusion Matrix:

Animal Cells Human Unsure

Animal 10 1 0 2

Cells 2 2 4 1

Human 2 1 14 8

Unsure 0 0 0 0

Animal Cells Human Unsure

Animal 5 0 0 2

Cells 2 0 2 0

Human 2 1 9 6

Unsure 0 0 0 0

Animal Cells Human Unsure

Animal 18 3 1 4

Cells 3 11 2 0

Human 1 7 20 27

If the classification is separated to two phases, 11 miss annotated papers in the data (10%).

## **Documentation and Resources**

Github Project: [ESA-RadLab/FullTextTagging (github.com)](https://github.com/ESA-RadLab/FullTextTagging)  
Front end?

## **Paper outline**:

Format?

How to differentiate from paper-qa and other similar if taking more machine learning oriented approach? ([whitead/paper-qa: LLM Chain for answering questions from documents with citations (github.com)](https://github.com/whitead/paper-qa))

## **Glossary**

VDU – Visual Document Understanding  
OCR – Optical Character Recognition  
IE – Information Extraction  
DocVQA – Document Visual Question Answering

## **Other notes**

The performance of general-purpose VDUs relies on the performance of OCR (Donut)

A graph of different colored bars

Description automatically generated with medium confidence

# Questions

* What to do “papers” like *../data/test\_SD\_files//2017\_Abstracts.pdf* and what does the screening annotations for such summaries even mean?
* How to handle studies with several classes? Can you choose several options in SyRF?
* Why are not all the “relativePfdPaths” not found in the Zotero?
* Where is the database of all the pdf files?

Wrongly annotated?:

**Annotation : Cells**

**Prediction: Human**

../data/test\_SD\_files//Yoon et al\_2009\_Expression of Activated Checkpoint Kinase 2 and Histone 2AX in Exfoliative Oral.pdf

CONTEXT:

The oral cell samples were collected from 100 subjects before and after the exposure to 2.34

cGy radiation. From each participant, an average of 100,000 cells was collected (range:

10,000 to 1,000,000). Approximately 40% of each sample were oral epithelial cells with

small ovoid nuclei and abundant cytoplasm and were deemed adequate for analysis. The

remaining 60% were of the surface keratin, anucleated squames, inflammatory cells and

amorphous debris. Of the 200 samples collected from 100 subjects, four samples from three

subjects had insufficient numbers of cells for analysis and were excluded from the study.

Subsequently, the analysis was performed for 97 subjects (196 samples).

We recognize several limitations of our study. The data are limited by the relatively small

sample size. Moreover, the samples were collected after exposure to a set dose of 2.34 cGy,

and the dose response was not assessed. Thus the ranges of doses that will induce detectable

pChk2 and γ-H2AX expression in exfoliative oral cells are unknown. Furthermore, the time

window for reliable detection of marker expression has not been determined. The high rate

of false negatives and the discrepancy rate between the two markers that are intimately

linked also make this modality too unpredictable at present to serve as an indicator of low-

dose radiation exposure for individuals. Our data nevertheless demonstrate that it is feasible

to measure biomarker expression in exfoliated oral cells to assess radiation exposure status.

Based on this proof-of-concept study, in a future investigation, we can assess the dose

response and the time course for the biomarkers and define a dose and time range that yield

highest sensitivity.

Exfoliated oral epithelial cells were collected from consenting participants before and after

exposure to ionizing radiation and with approval from the Columbia University Institutional

Review Board. We recruited 100 individuals, 55 males and 45 female, from the Columbia

University Dental Radiology clinic who were to receive routine dental radiographs (full

mouth series). The participants’ ages ranged from 20 to 77, the median being 47 years. The

ethnic background consisted of Hispanic (n = 67), white (n = 27), black (n = 5) and Asian (n

= 1). The dental radiographs consist of 18 F-speed intraoral films, equivalent to a radiation

dose of approximately 2.34 cGy. We collected exfoliative oral cells from the consenting

participants before and 20 min after exposure to ionizing radiation from the radiographs. As

a comparison, oral cell samples were collected sequentially from 10 nonirradiated subjects

20 min apart.

**Annotation: Cells**

**Prediction: Animal**

../data/test\_SD\_files//Flegal et al\_2013\_The Lack of Cytotoxic Effect and Radioadaptive Response in Splenocytes of Mice.pdf

CONTEXT:

Adult male or female C57BL/6J mice (Jackson laboratory, Bar Harbor, ME, USA) aged 7–8 weeks

were used in this study. The animals were acclimatized for a minimum of one week after their arrival

and then randomly assigned to experimental groups (5 mice per group). All animals were housed in a

specific-pathogen-free environment in the Biological Research Facility at Chalk River Nuclear

Laboratories (Chalk River, ON, Canada). Mice were maintained in cages inside ventilated cage racks

equipped with an automatic ventilation and watering system. In tritium exposure experiments, drinking

water with tritium was provided from bottles ad libitum. Animals were fed ad libitum and their health

status was examined daily. The facility was equipped with automatic computer-controlled temperature

(23 °C), air ventilation and 12-h light/dark cycle. Tests for infections were performed routinely and all

mice tested were negative. All protocols were performed in accordance with the guidelines of the

Canadian Council on Animal Care [21] with the approval of the local Animal Care Committee.

For low dose γ-irradiation, mice in plastic cages were irradiated with low doses of 20 or 100 mGy

using an open beam 60Co-γ-source (γBeam-150, Atomic Energy of Canada Limited, Chalk River) at a

dose rate of 1 mGy/min. Control mice were sham-irradiated. Twenty three hours after low dose in vivo

γ-irradiation or after 1-month HTO exposure, the mice were sacrificed by cervical dislocation and

spleens were removed and rinsed in phosphate buffered saline. Splenocytes were isolated and cultured

in RPMI media supplemented with 10% FBS, 2 mM L-glutamine in T25 flasks at 37 °C and

a 5% CO2 and 95% air atmosphere. Twenty four hours after low dose γ-irradiation of mice

(approximately 30 min after culture initiation in a CO2-incubator), flasks with cell suspensions were

irradiated with a challenging dose of 2 Gy using a GammaCell-220 device (Atomic Energy Canada

Limited, Chalk River) at a dose rate of 6.6 Gy/min (60Co γ-ray source) at room temperature.

Immediately after irradiation the cell cultures were returned to the CO2-incubator and incubated for

various times before sampling. At time points 0 (without 2 Gy dose), 1 and 24 h after challenging

irradiation, cell aliquots were collected and processed for flow cytometry.

**Annotation: Human**

**Prediction: Cells**

../data/test\_SD\_files//Wiencke et al\_1992\_Heterogeneity in the clastogenic response to X-rays in lymphocytes from.pdf

CONTEXT:

Cell culture and cytogenetics

Venous blood was drawn into 15 ml heparinized vac-

utainers. Whole blood (0.5 ml) was added to 4.5 ml of

RPMI 1640 medium containing 10 percent fetal calf

serum, 2 mM glutamine, 100 units/ml penicillin, 100

wg/ml streptomycin, and two percent phytohaemag-

glutinin M (PHA-Gibco). The blood was cultured at

37°C in one-ounce glass prescription bottles. The same

lots of vacutainers, serum and prescription bottles were

used for all experiments. For studies of X-ray-induced

chromatid breaks in the late G2 portion of the cell cycle,

cultures were irradiated 48 hours after incubation with

PHA. Colcemid (final concentration = 2 × 10 .7 M) was

added immediately after radiation treatments and cells

harvested 3 h later. Fixation was performed by stan-

dard cytologic procedures, i.e., the cells were exposed

to a 0.075 M KC1 solution for 8 min to spread chromo-

somes and then fixed in methanol-acetic acid solution

(3:1). Cytologic preparations were made by placing

cells on wet slides and staining with Giemsa. The slides

were scored for chromatid and isochromatid breaks,

gaps and chromatid exchanges. Chromatid gaps were

defined as achromatic lesions less than the width of the

chromatid. The blood samples in all experiments were

received, coded, and the cells were scored blindly.

Co-cultivation experiments

To study the relative progression of cells into meta-

phase following X-ray treatments, PHA-stimulated

blood cultures from a male A-T heterozygote were

mixed in equal volumes with parallel blood cultures

from a normal female blood-donor immediately prior

to irradiation (i.e., 48 h). Mixed male and female cul-

tures, along with the appropriate controls, were treated

with 60 cGy of X-rays and the cells harvested 0.5, 1.0,

2.0, 3.0, and 6.0 h later. In these experiments, colcemid

was added 0.5 h prior to harvest. The effect of

irradiation on the mitotic index of each population was

determined by examining the proportion of male (XY)

and female (XX) metaphases observed on the slides.

Only cells with 46 chromosomes were analyzed.

Mitotic indices are based on counts of at least 3,000

interphase cells from the same cultures used for male:

female ratios.

Materials and methods

The study group consisted of obligate A-T hetero-

zygotes (i.e., parents of patients diagnosed with A-T)

and unrelated individuals with no family history of the

disease. Nineteen A-T heterozygotes and 45 control

subjects were studied. Data on participants' age, gen-

der, and smoking history were collected. Repeat

studies were carried out on seven subjects (five con-

trols and two A-T heterozygotes), approximately one

year apart. A-T heterozygotes and A-T probands who

were studied were drawn from cases diagnosed or fol-

lowed at Moffitt Hospital, San Francisco, California or

at the Brigham and Women's Hospital, Boston, Mas-

sachusetts over the time period, May 1986 to June 1988.

Controls, who were roughly matched for age and gen-

der, consisted of friends of A-T heterozygotes and lab-

oratory or support staff. All blood samples were coded

prior to laboratory analyses; in all cases, samples were

**Annotation: Human**

**Prediction: Cells**

../data/test\_SD\_files//Wu et al\_2012\_Identification of ERp29 as a biomarker for predicting nasopharyngeal carcinoma.pdf

CONTEXT:

Cell viability assay. 2-(2-methoxy-4-nitrophenyl)-3-(4-

nitrophenyl)-5-(2,4-disulfophenyl)-2H-tetrazolium, mono sodium

salt (CCK-8, Beyotime) assay was used to detect cell viability

in response to irradiation. Briefly, cells were seeded in 96-well

culture plates at 3x103 for CNE-1 cells or 2x103 for 6-10B cells per

well. After incubation for 8 h, the cells were exposed to 8 Gy X-ray

irradiation. Cell viability was determined by CCK-8 at various

time intervals according to the manufacturer's instructions.

Optical densities were determined on a microtiter plate reader

(Peskin and Winterbourn 2000) at 450 nm. Three independent

experiments were done in triplicate.

Clonogenic survival assay. Cells were plated in triplicate at

cell population of 102, 2x102, 4x102, 103, 104, 105 per dish, and

then were exposed to a range of radiation doses (0-8 Gy). After

irradiation, the cells were cultured for no less than 12 days

and the number of surviving colonies (defined as a colony

with >50 cells) was counted and the data normalized to the

appropriate sham-irradiated control group. Survival para-

meters D0 and N were fitted according to the linear quadratic

equation [S=1-(1-e-D/Do)N] using SigmaPlot 9.0 software

(Systat Software Inc., USA). Three independent experiments

were done.

**Annotation: Human**

**Prediction: Cells**

../data/test\_SD\_files//Akita et al\_2010\_MESENCHYMAL STEM CELL THERAPY FOR CUTANEOUS RADIATION SYNDROME.pdf

CONTEXT:

adipose cells are recommended. In order to elucidate efficacy of these stem cells,

both in vitro and in vivo experiments are undertaken.

An in vitro stem cell biology and analysis

In order to investigate human mesenchymal stem cell proliferation, sub-

confluent cultured hMSCs were used and irradiated by an X-ray radiation

generator. The cells were immediately transferred to the incubators after

irradiation. For control cells of different species of origins are used. Both human

neuroblastoma cellls (NG1087-15) and rat pheochromocytoma cells (PC-12) are

used. Cell proliferation was consistent in three cell groups in the normal

condition (no radiation and normal medium), however, 20Gy irradiation caused

cell death in groups of NG1087-15 and PC-12 in 48 hours, in contrast, the

hMSCs survived up to next 96 hours.

In an electron microscopy, irradiated hMSCs demonstrated surface microvilli all

over the cells, however the hMSCs still survived after 60 Gy irradiation, of which

dose is considered medium and induce the significant intestinal bleeding.

An in vivo model and whole body irradiation by an X-ray generator

Animals were aged 10 weeks and weighing 300–350 g, and were used. Animals

were obtained from CLEA JAPAN (Tokyo, Japan), housed in the laboratory

animal centre for biomedical research, Nagasaki University School of Medicine

(Nagasaki, Japan), and the protocol of the animal experiment was approved by

the Institutional Animal Care and Use Committee of Nagasaki University, no.

0204080111. They were handled according to the guidelines established for

animal care at the centre. Each rat had free access to both sterile water and

standard rodent soft chow ad libitum.

4 Gy or 20 Gy whole body irradiation to 10 nude rats (F344/NJCl-rnu), which are

deleted T-cell function and thus acute immune rejection to human derived cells is

minimized, were performed at Atomic bomb Disease Institute, Nagasaki

University, by a X-ray radiation generator (EXS-300-5, Toshiba, 200kV, 15 mA,

0.405 Gy/min). Animals were divided into two groups of five each, control group

and hMSCs- with bFGF-treated group and surgical procedures were performed

immediately after irradiation.

**Annotation: Animal**

**Animal (frogs) and in vitro experiments both described. Only in vitro study is valid for data extraction.**

**Prediction: Cells**

../data/test\_SD\_files//Audette-Stuart et al\_2011\_Adaptive response in frogs chronically exposed to low doses of ionizing.pdf

CONTEXT:

For the preliminary ﬁeld experiments, the liver cells from six

leopard frogs, exposed to a 4 Gy “challenge” dose in vivo, were used

(Table 2). One frog, a female, was collected from Dew Drop Lake.

For the laboratory (in vitro) experiments, the cells were main-

tained in culture for approximately 24 h prior to being irradiated.

Some samples were subjected to an “adapting” irradiation dose

(100 mGy in 2003 and 50 mGy in 2008) delivered at a dose rate of

about 5 mGy/min in 2003 and 10 mGy/min in 2008 (GammaBeam

150 C 60Co irradiator, Atomic Energy of Canada Limited). The

cultures were incubated at room temperature for 3 h, after which

some samples received a 4 Gy “challenge” dose (at a dose rate of

about 16 Gy/min in 2003 and 8 Gy/min in 2008, Gammacell 220,

Atomic Energy of Canada Limited (AECL)). Apart from the actual

exposures to ionizing radiation, care was taken for all samples to be

subjected to the same treatments (including transport and

handling). In preparation for the micronucleus assay, 10 mL of

cytochalasin B (0.2 mg/mL) (SigmaeAldrich Canada Ltd.) was added

per 1.0 mL of culture suspension to arrest cells at cytokenesis, 48 h

before the cells were harvested and ﬁxed onto slides.

**Annotation: Animal**

**in vitro, in vivo and ex vivo studies**

**Prediction: Cells**

../data/test\_SD\_files//Klokov et al\_2013\_Low dose IR-induced IGF-1-sCLU expression.pdf

CONTEXT:

Cell lines, treatments and survival assays

Human MCF-7 breast cancer cells and a stably transfected 1403 MCF-7

clone (MCF-7 cells containing stably integrated 1403bp CLU promoter

fused to ﬁreﬂy luciferase) were cultured in Dulbecco’s modiﬁed Eagle

medium (BioWhittacker; Walkersville, MA, USA) containing 10% fetal

bovine serum (HyClone; Utah, USA). Human HCT116 colorectal carcinoma

parental (DNA mismatch repair (MMR)-defective, TbRII(cid:2)), a corrected clone

(HCT116 3--6 cells, MMRþ, TbRIIþ) and p53(cid:2)/(cid:2) cell lines were conﬁrmed

for p53, MMR and TbRII expression statuses as described,44,45 and grown in

10% fetal bovine serum-Dulbecco’s modiﬁed Eagle medium. Growth arrest

responses of these cells to TGFb1 were monitored by changes in 3H-

Thymidine incorporation into DNA, using standard protocols. HMECs, life-

extended using hTERT and CD4 overexpression, were kindly obtained from

Dr David Euhus (UT Southwestern). All cells and their stable derivatives

were maintained at 371C at 5% CO2-95% air. For TGFb1 or IGF-1

treatments, cells were serum-starved (0.5% fetal bovine serum) overnight,

and exposed to IGF-1 or TGFb1 at the indicated doses in normal serum-

Dulbecco’s modiﬁed Eagle medium. Cell irradiations were performed using

a JL Shepherd Assoc. (An Fernando, CA, USA) 137Cs Mark I-68 irradiator

(3.87Gy/min), with appropriate shielding to lower dose rate for accurate

whole body irradiations at low doses. Mock-irradiated or dimethyl

sulfoxide-treated (UT) cells were treated identically to IR- or TGFb1-

exposed cells without cytotoxin treatments. All cells were routinely tested

and found free of mycoplasma. Survival was assessed by colony forming

ability assays.

Western blot analyses of irradiated tissues

In separate experiments, mice (three per group) were treated with whole

body IR at doses of 0.02 or 1Gy. Tissues (colon/small intestine, bone

marrow, muscle, spleen and lung) were then extracted 24 or 72h later as

indicated and immediately homogenized as described.43 Tissue homo-

genates were incubated on ice for 30min, centrifuged (5000(cid:3)g, 15min,

41C) and protein samples diluted to a ﬁnal concentration of 2mg/ml.

Extracts were combined from each of three mice per group. Proteins were

then separated by 8--12% SDS-polyacrylamide gel electrophoresis, and

steady state levels analyzed by western blotting. Westerns shown are

representative of three experiments with similar results.

**Annotation: Cells**

**Prediction: Animal**

../data/test\_SD\_files//Liu et al\_2019\_Reduction of Delayed Homologous Recombination by Induction of Radioadaptive.pdf

CONTEXT:

Pure Chemical Industries) at 4(cid:3)C. Before analysis with flow

cytometry, the cells were washed once with PBS ((cid:2)), resus-

pended in Opti-MEM (Life Technologies, Carlsbad, Califor-

nia), and filtered through a 35-mm cell strainer (Corning Inc).

The cell suspensions were gated using forward and side scatter

and analyzed on a BD FACScalibur flow cytometer (Becton

Dickinson, Frankloin Lakes, New Jersey) with an excitation

laser at 48 nm, emission filters to measure green fluorescence

(530 nm), and autofluorescence (580 nm). The cells from

C57BL/6-Tg (CAG-EGFP) mice constitutively expressing

EGFP and wild-type C57BL/6J mice with no EGFP expression

were used, respectively, as a positive control and a negative

control for flow cytometry. Cells with significantly higher lev-

els of fluorescence at 530 nm than 580 nm were judged as

recombinant cells, namely, RaDR-GFP-positive cells with

“green fluorescence.” Six to 12 animals were used per experi-

mental point, and for each sample at least 1 million cells were

analyzed. The frequency of recombinant cells was expressed as

the number of GFP-positive cell parts per million of nucleated

bone marrow cells or splenocytes.

Statistical Analysis

Statistical evaluation of the data was done using the w2 test for

the 30-day survival and Student t test for the recombinant cell

frequency. The statistical significance was assigned to P < .05.

Enumeration of Recombinant Cells

To comparatively evaluate the homologous recombination fre-

quency in the nucleated cells in the bone marrow and spleen in

mice under experimental conditions capable of inducing RAR

(receiving both the priming dose of 0.50 Gy X-rays and the

challenge dose of 4.00 Gy X-rays) or incapable of inducing

RAR (receiving only the challenge dose of 4.00 Gy X-rays),

the recombinant cells (GFP-positive cells) in the nucleated

bone marrow cells and splenocytes were analyzed by flow

cytometry. In brief, the mice were euthanized by CO2 asphyx-

iation the following day after the 30-day survival test, and the

femurs and spleens were collected. Single-cell suspensions of

dissociated bone marrow tissues and spleens in phosphate-

buffered saline free from calcium and magnesium ions (PBS

((cid:2)); Wako Pure Chemical Industries, Ltd, Japan) were filtered

through a 40-mm cell strainer (Corning Inc, Corning, New

York) after treating with Tris-buffered ammonium chloride for

the lysis of red blood cells and washing with PBS ((cid:2)), and then

the nucleated bone marrow cells and splenocytes were fixed

with 1% paraformaldehyde phosphate buffer solution (Wako

**Annotation: Animal**

**Prediction: Cells**

../data/test\_SD\_files//Martinel Lamas et al\_2013\_Protection of Radiation-Induced Damage to the Hematopoietic System, Small.pdf

CONTEXT:

Cell culture and radiation dose–response curves

The human breast cancer cell lines MDA-MB-231 and

MCF-7 (American Type Tissue Culture Collection, VA, USA)

were cultured in supplemented RPMI 1640 as it was previously

described [8]. For the radiosensitivity studies, MCF-7 and

MDA-MB-231 cells were seeded in 6-well plates (1,200 cells/

well) and were treated with 10 µM JNJ7777120 or remained

untreated. The radiobiological parameters (SF 2Gy: fraction of

surviving cells after exposure to 2 Gy dose; Dose 0.01: dose

that reduces survival to 1%; Dose 0.10: dose that reduces

survival to 10%) were calculated from the clonogenic surviving

curves as it was described [8].

Histopathological studies

SMG and small intestine were removed and were fixed with

10% neutral buffered formalin while femur bone marrows were

fixed with Bouin’s solution. Tissue samples were embedded in

paraffin and cut into serial sections of 4 µm thick. The

histopathological characteristics were examined on tissue

sections after hematoxylin-eosin (H&E) staining as it was

previously described [3,6,7]. Also, specimens were stained with

periodic acid schiff (PAS) staining and counterstained with

hematoxylin to evaluate glycogen deposits.

**Annotation: Cells**

**cells and animal**

**Prediction: Animal**

../data/test\_SD\_files//Qu et al\_2016\_Protein expression of nucleophosmin, annexin A3 and nm23-H1 correlates with.pdf

CONTEXT:

Xenograft tumorigenicity assays. BALB/c athymic nude mice

(males and females, aged 4‑6 weeks; n=36) were obtained from

the Animal Laboratory Center of Guangxi Medical University

[Guangxi, China; license number, SCXK (Gui) 2009-002].

The mice were housed five per cage, and maintained under

specific pathogen‑free conditions. The mice were randomly

divided into three groups of 12 animals each: Group A,

treated with CNE-2R cells; group B, treated with CNE-2

cells; and group C, treated with saline. In group A, CNE-2R

cells (1x107 in a total volume of 0.2 ml) were subcutaneously

injected into the right hind legs of 6 mice, and both the right

and left hind legs of 6 mice. In group B, CNE-2 cells (1x107

in a total volume of 0.2 ml) were subcutaneously injected

into the right hind legs of 6 mice, and both the right and left

Cell lines and cell culture. Poorly-differentiated human

NPC CNE-2 cells were purchased from the Cancer Hospital

of Fudan University (Shangha, China). CNE-2R cells were

induced by treating the parental CNE-2 cells with frac-

tioned cobalt-60 γ‑ray irradiation (total dose, 6,400 cGy;

Theratron 780; Theratronics International Ltd., Kanata,

Canada) (3). The CNE-2R and CNE-2 cells were cultured

separately in RPMI 1640 medium (Hyclone; GE Healthcare

Life Sciences, Logan UT, USA) supplemented with 10% fetal

bovine serum (Hangzhou Sijiqing Biological Engineering

Materials Co., Ltd., Hangzhou, China), penicillin G (100 kU/l;

(North China Pharmaceutical Co., Ltd., Shijiazhuang, China)

and streptomycin (100 mg/l; Qilu Pharmaceutical Co., Ltd.,

Jinan, China). The cells were maintained at 37˚C in a 5% CO2

incubator until use.

Hematoxylin and eosin (H&E) staining. The mice were sacri-

ficed 2 weeks after irradiation, and autopsies were performed

on all injected mice. Xenograft tumors were excised, imme-

diately placed in 10% neutral‑buffered formalin and fixed for

24‑48 h. Subsequent to fixation, samples were dehydrated

and embedded in paraffin. A series of 4‑µm sections were

prepared from each specimen, mounted on poly-lysine-coated

glass slides and dried for 4‑5 h at 37˚C to promote adhesion.

H&E staining was performed on one section from each

**Annotation: Human**

**Human and Animal**

**Prediction:Animal**

../data/test\_SD\_files//Delp et al\_2016\_Apollo Lunar Astronauts Show Higher Cardiovascular Disease Mortality.pdf

CONTEXT:

Forty-four male C57BL/6 mice (Jackson Laboratory, Bar Harbor, ME), 16 weeks of age, were individually

housed at the Brookhaven National Laboratory animal facility at Long Island, New York. Animals were maintained

in a controlled environment (12:12 hour light-dark cycle, 24 ± 2 °C) and provided food and water ad libitum.

Mice were randomized by body mass to one of four groups: control (Con, n = 11), hindlimb unloaded (HU,

n = 11), total body irradiated (TBI, n = 11), and the combined TBI and HU (TBI+ HU, n = 11). One week after

the conclusion of the unloading treatment for the HU and TBI+ HU groups, the mice in all four groups were

shipped to Florida State University, individually housed in the animal vivarium under controlled environmental

conditions (12:12 hour light-dark cycle, 24 ± 2 °C) and provided food and water ad libitum.

For the animal studies, a one-way ANOVA and Fisher LSD post-hoc tests were used to detect differences in

body, tissue and vascular characteristics. Vasomotor responses were evaluated using repeated-measures ANOVAs

to detect differences between experimental groups and drug doses or pressure changes. All values are presented

as means ± SE. A value of P ≤ 0.05 was considered statistically significant.

References

1. Boerma, M. et al. Space radiation and cardiovascular disease risk. World J Cardiol 7, 882–888, doi: 10.4330/wjc.v7.i12.882 (2015).

2. Pietsch, J. et al. The effects of weightlessness on the human organism and mammalian cells. Curr Mol Med 11, 350–364 (2011).

3. Ball, J., Evans, C. & eds. Safe passage: Astronaut care for exploration missions. (National Academies Press, 2001).

4. Zeitlin, C. et al. Measurements of energetic particle radiation in transit to Mars on the Mars Science Laboratory. Science 340,

The second unique feature of the present study is that it is the first to examine the long-term mortality risks

of spaceflight in LEO and deep space. Exposure to charged particles comprising the galactic cosmic rays in deep

space has the potential to elicit a number of complications in biological tissue. Recent work in rodents and cell

culture has highlighted the potentially harmful effects of such exposures on the cardiovascular system20,27–30,

which may translate to astronauts engaged in deep space expeditions.