Exploiting AI To Improve DNA Damage Detection and Quantification.

Elliot Corless[1], Norah A Owiti,[1] Debra A MacKenzie,[2] Karen L Cooper,[2] Laurie G Hudson,[2] Cheng Zheng[3], Johnnye Lewis,[2] Peter So[3], and Bevin P. Engelward[1],

[1]Massachusetts Institute of Technology, Department of Biological Engineering
[2]University of New Mexico, Department of Health Sciences
[3]Massachusetts Institute of Technology, Department of Biological Engineering

The Engelward Lab

ABSTRACT

Rational: The CometChip assay is a well established, high throughput, in-vitro method for the detection and quantification of DNA damage in cells based on the Comet assay, developed in 1984. Current commercially available analysis software is expensive, proprietary, and requires manual input from the user. Manipulation of data analysis by eye increases the amount of time necessary to analyze each image dataset and introduces bias that can affect DNA damage quantification. This projects aims to use machine learning to train an image analysis algorithm that can ultimately be used to fully automate image analysis. Fluorescent images of electrophoresed comets must be analyzed to calculate the extent of DNA damage. Expensive commercially available analysis software is used to determine the main parameter which is the cutoff between the "head" (intact DNA with low mobility) and the "tail" (damaged DNA with higher mobility) of the comet. The percentage of total DNA in the tail represents the amount of damage. Current software cannot determine the quality of individual comets, and often the software places the cutoff line far from the correct location forcing users to painstakingly manually sift through hundreds (a very small dataset) to hundreds of thousands (a large dataset) of images "improve" their data by manually eliminating debris/non analyzable comets, and setting the head/tail cutoff by eye. This will eliminate the possibility of manual input from users, greatly increasing throughput (by orders of magnitude, in some cases) and providing objective and consistent analysis parameters.

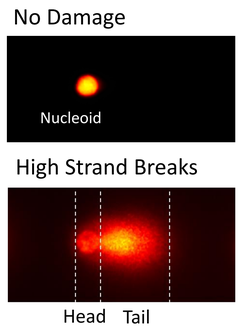
Comet vs CometChip Assay: the Comet assay relied on randomly patterned cells on an agarose coated glass slide, this led to significant overlapping when overloaded, and few cells on the plate when underloaded, creating high degrees of variability in data quality between slides, as well as a low number of images per slide even under ideal loading conditions. The CometChip assay exploits micro-patterned micro-wells in agarose that can be separated into the "macro" wells of a typical 96 well plate.

Current progress
machine learning has been applied to both automation steps, leading to a 99.7% (9997/1000)screening of debris or ununiform comets in the first step, and an over 95% correct cutoff determination in the second step. We will soon be incorporating these into a free to use program that will be compared with existing software.

FUNDING

This work was supported by the National Institute of Environmental Health Sciences Superfund Basic Research Program, National Institute of Health, P42 ES027707.

The Comet Assay



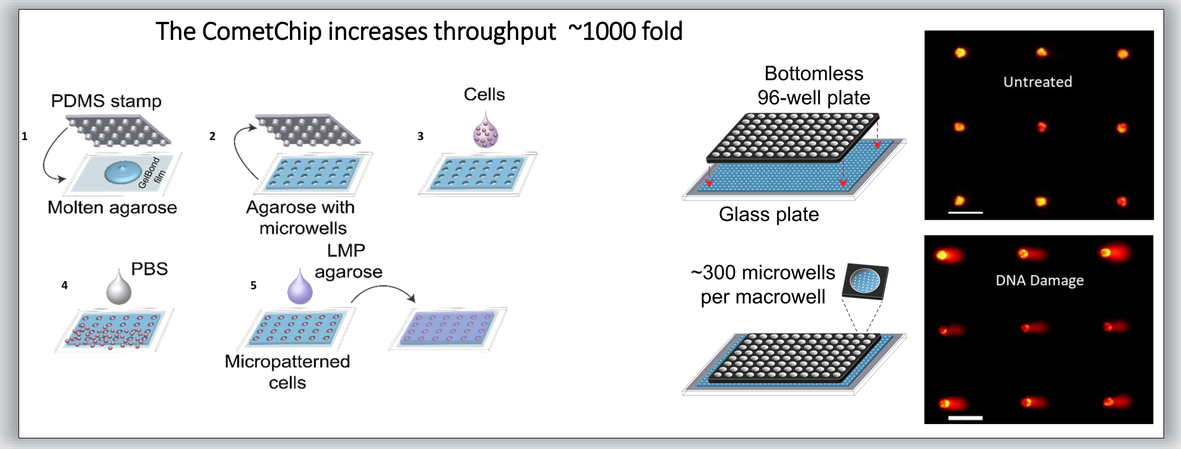
In vivo comet assay is adopted in regulatory guidelines
- Genotoxicity testing (OECD, 2014)[1]
- Standard test battery (FDA, 2012)[2]

In vitro comet assay adapted in Japan and EU, not the US
- Lack of sensitivity in detection of bulky adducts

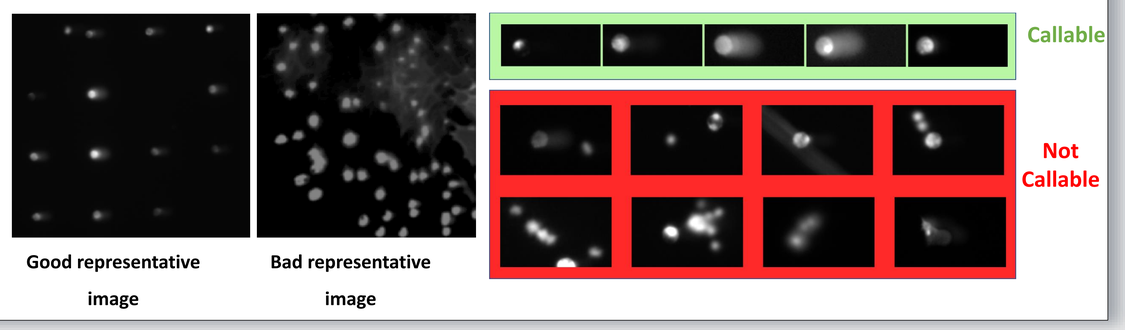
Traditional comet assay is laborious and low-throughput

1. OECD/OCDE TG 489. 2014. OECD guideline for the testing of chemicals.
2. U.S. Dept of Health and Human Services, Food and Drug Administration. 2012. Guidance for industry.

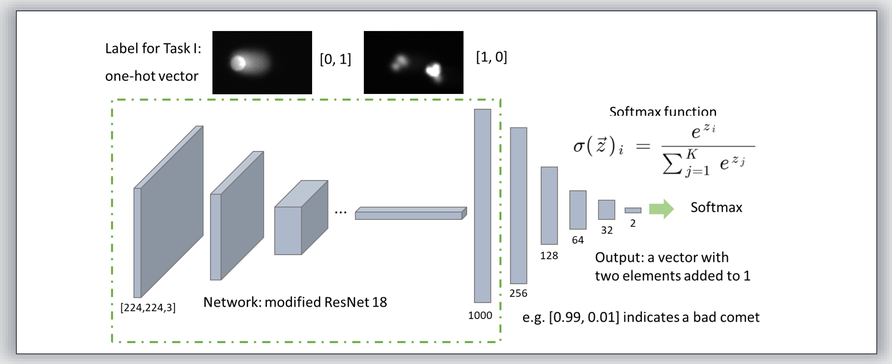
The CometChip Platform



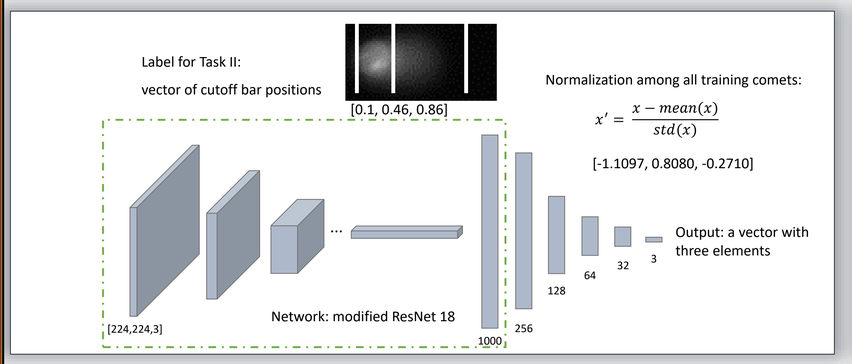
CometChip Images Have Variable Quality



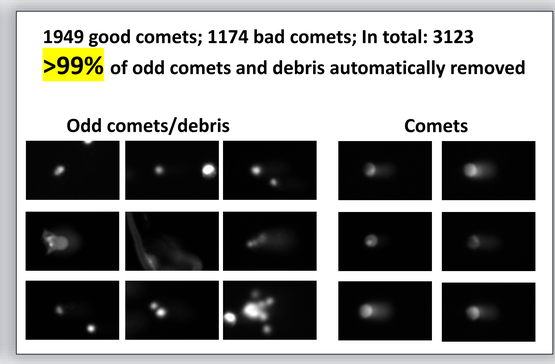
Task I: Identify Analyzable Comets



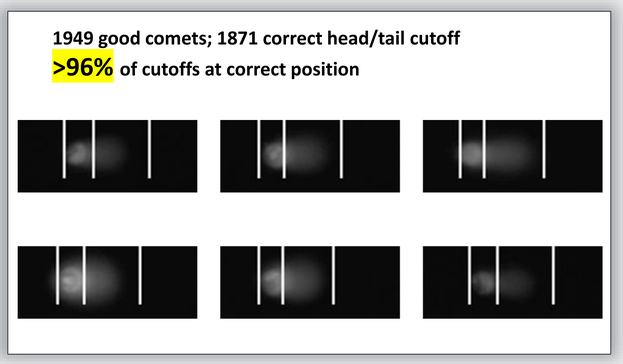
Task II: Detect Cutoffs For Head and Tail



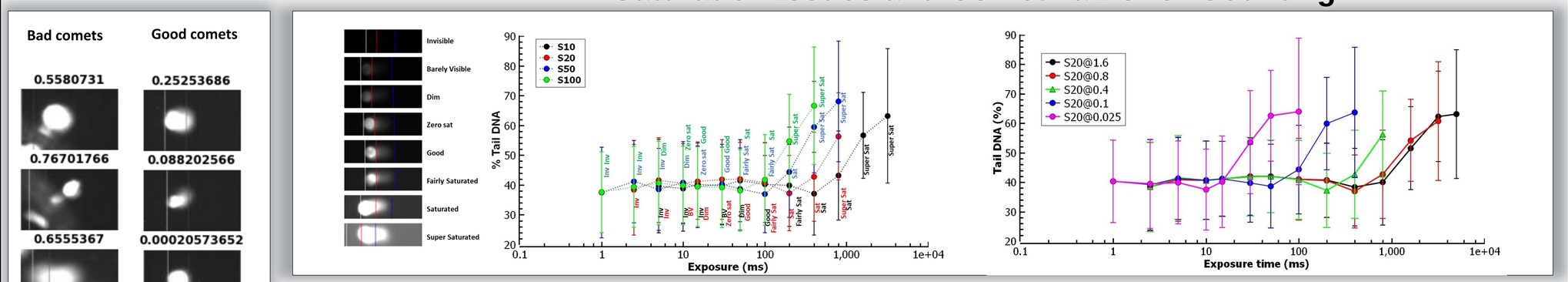
Result from Task I



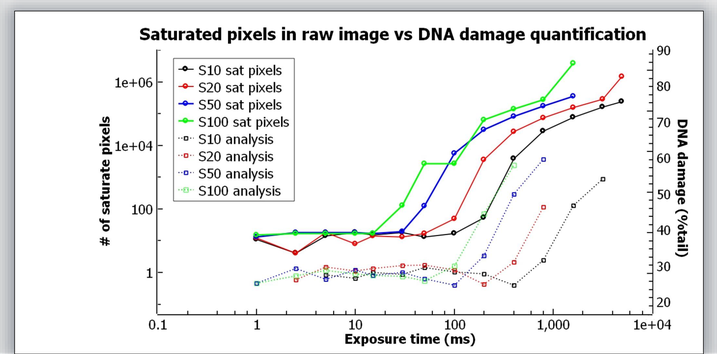
Result from Task II



Saturation Issues and Comet Tail Over-Counting



Determining guidelines for Pixel Saturation



Conclusions and Significance

- **The CometChip platform is robust and high-throughput, will soon have a free and equally high throughput analysis software**
- **Machine learning is capable of producing very high accuracy in results without input from user.**
- **Ideal exposure times have been empirically determined and guideline set for the effects of image pixel saturation which will be accounted for in finished product.**
- **Once completed, this open-source, free software will remove any fiscal hurdles for CometChip image analysis, and drastically increase throughput and fidelity of DNA damage quantification using CometChip.**