

The Modelling and Analysis of Comet Assay Data

CHRISTMAS MEETING: SELINA DHINSEY

SUPERVISED BY PROF. TIM GREENSHAW, DR. JASON PARSONS & PROF. CARSTEN WELSCH



Proton Therapy

- Cancer can be treated therapeutically using photons and particles
- Aim: deposit energy into the tumour's cells, damage DNA, halt cell replication and kill the tumour
- Protons are favourable due to depositing max energy at end of track : Bragg peak
- Physical process well understood LET
- Biological response poorly characterised



Plot showing the varying energy depositions along the paths of protons, x-rays and carbon ions.

DNA

- Nuclear DNA is double stranded and alpha helical
- Sugar phosphate backbone outside negatively charged
- Base pairs inside joined by hydrogen bonding
- Bases pair up in a particular way -

complementary strands



DNA Damage

Main types of damage are strand breaks

SSBs - breaks on only one strand of DNA

Repaired using the intact complementary strand

Clustered/complex damage - many breaks of either type concentrated over a few nm



DSBs - breaks on both strands in very close proximity (few base pairs)

More difficult repair pathways required

How can we measure the damage?

Comet Assay: Measuring DNA Damage

Expose any DNA strand breaks

Type of damage can be determined

via assay conditions

Creates images to be analysed



Image Analysis



Image of control cells (no irradiation) after a comet assay



Image of irradiated cells that have formed comets following a comet assay. Comet heads are shown in red, and tails in green.

DNA Damage Measure - Thresholding

- Program to automatically measure comet properties in many images at once
- Uses thresholds to distinguish the whole comet body and comet head from background
- Thresholding not accurate enough-losing comet data
- Needs optimisation per image set not robust





Comet Identification using NN

- Built upon Microsoft's COCO model to identify * comets within images automatically as a first step
- Uses instance segmentation *
- Identifications based on structure rather than ** intensity - more robust across image sets
- Lack of data is an issue for testing and improving * accuracy

Classification





Semantic Segmentation



Instance Segmentation



Instance Segmentation



Monte Carlo Image Simulation

- Currently developing MC of electrophoresis process
- Simulating the transport of DNA fragments through agarose
- Better understand the physical kinematics of DNA movement during electrophoresis



Monte Carlo Image Simulation

Simulating DNA strands and strand breaks to create comet assay images – create training and ground truth data for NN

Model accurate for short strands but need better description of longer strands











STFC Industrial Placement

Currently on placement with Chanua – Science Park ic1

Using digital methods to create mental health and well

being products/services

Working on the data science and analysis aspects of

projects





HACKING HEALTH

Thank you, any questions?



Instance Segmentation Model

- Trained off manually created segmentation masks
- Transfer learning use already defined networks (Mask-RCNN/ COCO) pre-trained weights
- COCO trained on > 200,000 labelled images in 80 categories
- Based in Tensorflow popular, lots of documentation and examples etc.
- Successful model trained and verified off < 100 images</p>

MC DNA Simulation





Comet Assay Images

- Methods of analysis vary a lot
- By eye relative & carries lots of variability no numeric results
- Semi-automated user defines comet regions and comet properties are measured
- Full Automation quickest, but algorithms of how measurement are performed still vary





HeLa Cell Line

- HeLa Cells Henrietta Lacks
- Ovarian Cancer cells
- First immortal cell line continually reproduces
- Cells taken from Henrietta without her consent in 1951



18