Inter-laboratory automation of the in vitro micronucleus assay using imaging flow cytometry and deep learning

John W. Wills[1,2] · Jatin R. Verma[3] · Benjamin J. Rees[3] · Danielle S. G. Harte[3] · Qiellor Haxhiraj[3] · Claire M. Barnes[1] · Rachel Barnes[3] · Matthew A. Rodrigues[4] · Minh Doan[5] · Andrew Filby[6] · Rachel E. Hewitt[2] · Catherine A. Thornton[3] · James G. Cronin[3] · Julia D. Kenny[7] · Ruby Buckley[7] · Anthony M. Lynch[3,7] · Anne E. Carpenter[8] · Huw D. Summers[1] · George E. Johnson[3] · Paul Rees[1,8]

Abstract

The in vitro micronucleus assay is a globally significant method for DNA damage quantification used for regulatory compound safety testing in addition to inter-individual monitoring of environmental, lifestyle and occupational factors. However, it relies on time-consuming and user-subjective manual scoring. Here we show that imaging flow cytometry and deep learning image classification represents a capable platform for automated, inter-laboratory operation. Images were captured for the cytokinesis-block micronucleus (CBMN) assay across three laboratories using methyl methanesulphonate (1.25–5.0 µg/mL) and/or carbendazim (0.8–1.6 µg/mL) exposures to TK6 cells. Human-scored image sets were assembled and used to train and test the classification abilities of the "DeepFlow" neural network in both intra- and inter-laboratory contexts. Harnessing image diversity across laboratories yielded a network able to score unseen data from an entirely new laboratory without any user configuration. Image classification accuracies of 98%, 95%, 82% and 85% were achieved for 'mononucleates', 'binucleates', 'mononucleates with MN' and 'binucleates with MN', respectively. Successful classifications of 'trinucleates' (90%) and 'tetranucleates' (88%) in addition to 'other or unscorable' phenotypes (96%) were also achieved. Attempts to classify extremely rare, tri- and tetranucleated cells with micronuclei into their own categories were less successful ($\leq$57%). Benchmark dose analyses of human or automatically scored micronucleus frequency data yielded quantitation of the same equipotent concentration regardless of scoring method. We conclude that this automated approach offers significant potential to broaden the practical utility of the CBMN method across industry, research and clinical domains. We share our strategy using openly-accessible frameworks.

Keywords Micronucleus test · Genetic toxicology · Compound screening · Machine learning · High throughput · Image analysis.

Introduction

Across industry, government and academic research institutions the in vitro micronucleus test is one of the most widely used bioassays for the identification and quantification of chromosomal damage (Decordier and Kirsch-Volders 2006; Fenech 2000, 2020; Kirsch-Volders et al. 2011). Because DNA damage at the chromosome level is recognised as a key event in the initiation of carcinogenesis, the assay has become an essential component of genetic toxicity screening programmes worldwide (Fenech 2000). Harmonised assay protocols and scoring approaches have been detailed by Organisation for Economic Cooperation and Development (OECD)-Test Guideline 487 (OECD 2016). In addition to regulatory compound screening, the assay is also widely used for more specific research and clinical purposes including compound mode-of-action determinations, tumour radiosensitivity prediction and inter-individual monitoring of lifestyle, occupational and environmental factors including radiation biodosimetry assessments (Decordier and Kirsch-Volders 2006; Fenech 2000, 2020; Kirsch-Volders et al. 2011; Wang et al. 2019).

The micronucleus assay operates through the detection of whole chromosomes or chromosome fragments expressed by cells after nuclear division as satellite 'micronucleus' (MN) events. Because complete nuclear division is required to enable expression of these events, the 'cytokinesis-block' version of the assay was developed. This method inhibits cell division into daughter entities (cytokinesis) using the microfilament assembly inhibitor cytochalasin-B. This yields cells that have successfully undergone division easily identifiable by their binucleated appearance. In this way, the cytokinesis-block micronucleus (CBMN) assay allows scoring of micronucleus events in cells known to have undergone division during the treatment period. This avoids misleading results otherwise present due to pre-existing damage, sub-optimal cell culture conditions or from the selection of overly cytotoxic compound concentrations that retard or inhibit cell division and concomitant micronucleus expression (Decordier and Kirsch-Volders 2006; Fenech 2000; Kirsch-Volders et al. 2011).

Despite almost global utilisation, CBMN assay scoring still often relies upon manual observation and recording using light microscopy. Whilst manual scoring is the 'gold standard', even when slide identities are blinded, problems arise due to inter-scorer variability in addition to the process being time and labour intensive (Rodrigues 2014a, b, 2018). For these reasons, over the last two decades significant efforts have been directed towards automated approaches for both image collection and subsequent scoring. As recently reviewed (Rodrigues et al. , 2018), these largely involve slide and laser scanning microscopy systems that automate image collection in conjunction with traditional, threshold-based image classification techniques (Darzynkiewicz et al. 2011; Decordier et al. 2009, 2011; François et al. 2014; Maertens and White 2015; Rossnerova et al. 2011; Schunck et al. 2004; Seager et al. 2014; Smolewski et al. 2001; Varga et al. 2004; Verhaegen et al. 1994; Willems et al. 2010). Conventional flow cytometry methods have also been developed that aim to identify isolated micronuclei using fluorescence intensity measurements in the absence of image-based validation (Avlasevich et al. 2006; Bryce et al. 2008, 2010, 2013, 2007).

More recently, imaging flow cytometry unites the acquisition approach of flow cytometry with microscopical observation (Allemang et al. 2021; Rodrigues 2018, 2019; Rodrigues et al. 2014a, b, 2016a, b, 2018; Wang et al. 2019; Wilkins et al. 2017). This fluidics-based approach is well suited for processing cell suspension cultures (e.g., TK6 B-lymphocytes commonly used for the CBMN assay) enabling rapid collection of transmitted light brightfield, darkfield laser scatter and fluorescence images for populations of tens of thousands of single cells. Simple inclusion of a single nuclear fluorescent stain (e.g., Hoechst 33342, propidium iodide or DRAQ5 etc.) allows detection of parent nuclei and micronucleus events (Rodrigues 2018, 2019; Rodrigues et al. 2016b, 2018). Without need of further labels, the brightfield images provide essential context for detecting micronuclei associated with parent cells (Rodrigues et al. 2014a; Verma et al. 2018). The 'Amnis ImageStream^{X}' series cytometers (Luminex Corporation) further support unassisted data acquisition for multiple samples via a 96-well plate sampling attachment. Images are stored to sample-specific data files enabling archiving should human validation or reevaluation be required (Rodrigues et al. 2018). Traditional image classification approaches deployed within the manufacturer-supplied analysis software have shown utility for CBMN scoring automation (Rodrigues 2014a, 2014b, 2016a, 2016b, 2018, 2019; Wang et al. 2019; Wilkins et al. 2017). However, in our experience, these strategies require significant expertise to set up, in addition to frequent tuning to maintain acceptable performance, even within a single laboratory (Verma et al. 2018). Deviations of around 30% from the results obtained by manual microscopy scoring have also been reported in experiments utilising this approach to study irradiated peripheral blood lymphocytes (Rodrigues et al. 2016b). This outcome was in part attributed to the lack of flexibility of the implemented image analysis algorithms relative to the expertise of human judgement (Rodrigues et al. 2016b, 2018).

Building image classification strategies that *generalise* well enough to permit robust, entirely automated image classifications without need of human intervention or

configuration is a difficult task. This is because, even when protocols are harmonised, there will always be variability (e.g., illumination, focus and fluorescence staining heterogeneity etc.) in the input image data. This variation is even more extreme across laboratories due to the inevitable use of different imaging equipment, calibration settings, personnel, cell culture and bioassay regimens. Recently, artificial intelligence approaches have been achieving increasing success in providing generalised automation of image classification tasks (Caicedo et al. 2019; Moen et al. 2019). These approaches can use handcrafted features extracted from images in conjunction with machine learning algorithms, but increasingly, the availability of computational power is enabling the application of deep learning on image pixel data (Blasi et al. 2016; Eulenberg et al. 2017). This approach uses so-called deep convolutional neural networks in a manner inspired by neural connectivity in the brain. A typical image classification workflow involves assigning 'ground truth' class annotations to a large set of images before subdividing them into 'train' and 'test' data sets. The weights connecting the nodes of the neural network are then optimised during a training phase that attempts to match the input images to the annotated classifications. A potential issue due to the flexibility of neural networks as non-linear function approximators is that 'memorisation' due to over-fitting of training data can emerge (Zhang et al. 2017). For this reason, final network accuracy is assessed by cross validation against a test set that importantly was entirely 'unseen' during the training phase. Subsequently, the trained neural net can be deployed for the classification of new images.

In the context of the CBMN assay, deep learning approaches were recently used on imaging flow cytometry data using the cytometer manufacturer's 'Amnis Artificial Intelligence' software to identify binucleated cells in the 3-D reconstructed skin micronucleus assay. This binucleated cell population was then used as a refined start point from which to expedite manual identification of micronucleus events (Allemang et al. 2021). However, there would be considerable value in openly accessible frameworks for accessibility and for adaptability: the modular nature of modern, open source deep learning interfaces allows new network architectures to be easily switched or specifically tailored as they emerge. This flexibility provides complete ability to build bespoke solutions using the latest tools to pursue maximal accuracy and the accommodation of diverse research objectives.

Here, we used imaging flow cytometry to automate image capture for the CBMN assay across three laboratories using differing local protocols for cell culture, bioassay procedure, DNA staining, cytometer calibration and image collection. Given the inherent variability in the captured images, we investigate the ability of deep learning to enable robust, inter-laboratory scoring automation. To do this, we provide

an open framework that utilises the powerful, yet lightweight DeepFlow neural network architecture that has been previously optimised to achieve rapid training and classification of imaging flow cytometry data (Eulenberg et al. 2017).

Materials and methods

Multi-centre image collection

Image data was collected using three different Amnis ImageStream[X] imaging flow cytometers (Luminex Corporation, USA) across three locations: Central Biotechnology Services, Cardiff University School of Medicine (hereafter, Cardiff), the Department of Veterinary Medicine's Imaging Facility, University of Cambridge, UK (Cambridge) and at GlaxoSmithKline Research and Development, Stevenage, UK (GSK).

Chemicals

Methyl methanesulphonate (MMS) (#129925) (CAS registry number 66–27-3) and carbendazim (#378674) (CAS no. 10605–21-7) were purchased from Sigma-Aldrich (Merck), UK.

Cardiff and Cambridge: cell culture and cytokinesis-block micronucleus assay

P53 competent, virally transformed human B lymphoblastoid (TK6) cells were purchased from the Health Protection Agency Culture Collections (Wiltshire, UK). The cells were cultured in RPMI 1640 media (#A1049101, ThermoFisher) supplemented with 100 U/mL penicillin and 100 µg/mL streptomycin and containing 10% (v/v) heat-inactivated horse serum (#26050088, ThermoFisher). Cells were seeded at 2×10^5 cells/mL in 25 cm^2 flasks (ThermoFisher) and incubated at 37 °C for ~ 1.5 cell cycles (24–30 h) in the presence of MMS (0/1.25/2.5/5.0 µg/mL concentrations) or carbendazim (0/0.8/1.0/1.6 µg/mL concentrations) delivered using dimethyl sulphoxide (DMSO) as a vehicle, with co-exposed cytochalasin-B (#C6762, Sigma) added to a final concentration of 3 µg/mL as a cytokinesis block. Following exposure, cells were pelleted by centrifugation (200xg, 10 min) and washed once with 10 mL phosphate buffered saline (PBS). Cells were then pelleted and resuspended in 2 mL 1× BD FACS lysing solution (#349202, BD) for 12 min to achieve fixation and permeabilisation.

GSK: cell culture and cytokinesis-block micronucleus assay

TK6 (IVGT) cells (#13051501) purchased from ECACC, operated by Public Health England (Wiltshire, UK). The

cells were cultured in RPMI 1640 media with 2 mM glutamine (#52400025, ThermoFisher) supplemented with 100 U/mL penicillin and 100 µg/mL streptomycin (#15140-122, ThermoFisher), 1.8 mM sodium pyruvate (#11360-039, ThermoFisher) and containing 10% (v/v) heat-inactivated horse serum (#26050-088, BioSera, Labtech, UK). Cells were seeded at 2×10^5 cells/mL in 25 cm^2 flasks (ThermoFisher) and incubated at 37 °C for 24 h in the presence of carbendazim (0/0.8/1.2/1.6 µg/mL concentrations) delivered using dimethyl sulphoxide (DMSO) as a vehicle, with co-exposed cytochalasin-B (#C6762, Sigma) added to a final concentration of 6 µg/mL as a cytokinesis-block. Following exposure, cells were pelleted by centrifugation (200xg, 10 min) and washed once with 10 mL PBS (#10010-015, ThermoFisher). Cells were then pelleted and resuspended in 2 mL 1X BD FACS lysing solution (#349202, BD) for 12 min to achieve fixation and permeabilisation.

Nuclear labelling

Fixed, permeabilised cells were incubated with nuclear stains in PBS at room temperature. Nuclei and micronuclei were stained at the Cardiff and GSK laboratories by 30 min incubation with 0.05 mM DRAQ5 (peak excitation: 647 nm, peak emission: 681 nm) (#564902, BD). Samples at the Cambridge laboratory were stained with a 1:2500 dilution (8 µM) of Hoechst 33,342 (peak excitation: 351 nm, peak emission: 461 nm) (#62249, ThermoFisher) for 30 min. After labelling, cells were pelleted, resuspended and final cell concentrations adjusted through addition of PBS towards an optimal cell concentration for imaging flow cytometry (typically ~ 100 µL sample volumes at ~ 10^7 cells/mL).

Imaging flow cytometry

Brightfield and nuclear fluorescence images (20,000 images/sample) were collected using Amnis ImageStream^X (Luminex) flow cytometers using the 40× objective lens via the manufacturer's INSPIRE software at the Cardiff, Cambridge and GSK laboratories (described above). At Cardiff and GSK, DRAQ5-labelled cells were excited using 488 nm or 642 nm lasers (respectively) with the brightfield collected in channel 1 and DRAQ5 in channel 11. At Cambridge, Hoechst 33342-labelled cells were excited using a 405 nm laser with brightfield collection in channel 4 and nuclear fluorescence collection in channel 1. At all locations, a brightfield area range of 100–900 µm^2 was used to avoid debris, speed bead (i.e., the calibration beads that are run alongside cells to aid synchronisation of the camera and flow stream) and large aggregate image collection. Full details of image acquisition settings including the laser excitation powers the

exact cytometer models utilised at each location are provided in Supplementary Table S1.

Compensated image file generation using IDEAS

Prior to image extraction, raw image files (.rif) acquired by the INSPIRE software were converted to compensated image files (.cif) using identical settings via batch processing with a template using the IDEAS (version 6.2) software (Luminex). During the process, populations of cell images suitable for scoring were refined by gating out (brightfield area, 200–500 µm^2 versus aspect ratio, 0.75–1.0) debris and identifying a single cell population that was also suitably in focus. This was achieved by linescan gradient via the root mean square of the brightfield images ranging from 55 to 80.

Image data pre-processing: CIF to TIF extraction

Single, in-focus cell populations were exported from the IDEAS software in compensated image file format (.cif). The individual cell images within these files were then extracted to 16-bit grayscale, two-channel (nuclear fluorescence/brightfield) multipage TIF files using a custom script (code and example available for download from the BioStudies database (http://www.ebi.ac.uk/biostudies) in MATLAB and Python programming languages under accession number S-BSST641). During this TIF extraction process, each channel image was also max/min rescaled to normalise illumination. Images were also cropped and zero-padded (i.e., zeros added along image edges) enabling output at a constant 64×64 pixel-square size for input into the DeepFlow network.

Deep learning image classification

Automated scoring was achieved using a nine-class, feed-forward, image classification deep neural network built using our previously described "DeepFlow" architecture (Eulenberg et al. 2017). This network is optimised for the relatively small input dimensions of imaging flow cytometry data, and in itself utilises dual-path convolution/batch normalisation/nonlinearity subunits interspersed by max pooling from the popular "Inception" architecture (Szegedy et al. 2015). These subunit layers process and aggregate visual information at increasing scale before average pooling, the fully connected layer and softmax classification (full network architecture shown, Supplementary Figure 1). Images were passed to the network with an input size of $64 \times 64 \times 2$ (x, y, channels), with augmentation by random x/y reflection, rotation, translation, 90–110% image scaling and zero-center batch normalisation. Training lasted for 30 epochs using a batch size of 88 with optimisation under ADAM using cross-entropy loss. The initial learn rate was 5×10^{-3}, dropping

every five epochs by 0.9, with L2 regularisation 1×10^{-4} and epsilon 1×10^{-8}. Images were shuffled every epoch. The final pre-trained network alongside test images and all code detailing training hyper-parameters and final layer weightings are available for download in MATLAB (using the Deep Learning Toolbox) or Python (using TensorFlow/keras) languages at the BioStudies database (http://www.ebi.ac.uk/biostudies) under accession number S-BSST641.

Ground truth curation by human scoring

For the Cardiff/Cambridge analyses, cell image data across compounds (carbendazim and MMS) and exposure concentrations (0–5 µg/mL) were merged to create diverse ground truth training sets that contained the wide representation of different cell phenotypes essential for effective network training. Ground truth classifications for each image were assigned by biologists with extensive experience manually scoring the in vitro micronucleus assay, with phenotypes assigned through consideration of both the nuclear fluorescence and the brightfield image (i.e., ensuring nuclear events belonged to one cell etc.). As per micronucleus assay test guidance, the aim was to only score cells positive for micronucleus events, where the micronuclei were fluorescently-labelled, were circular/oval in shape, were within the size range of 1/3–1/16th that of the parent nuclei, and that were clearly inside the cell boundary of the parent cell (Fenech 2000; OECD 2016). At the GSK laboratory, TK6 cells were exposed to just the carbendazim compound (0/0.8/1.2/1.6 µg/mL concentrations) with the experiment conducted in triplicate. For the initial network cross validation with the GSK data, five thousand human-scored cell images were used with these events equally accumulated from across all carbendazim exposures. For the concentration–response analysis, cell populations of two thousand events were scored per concentration in triplicate by either human-scoring or by the neural network.

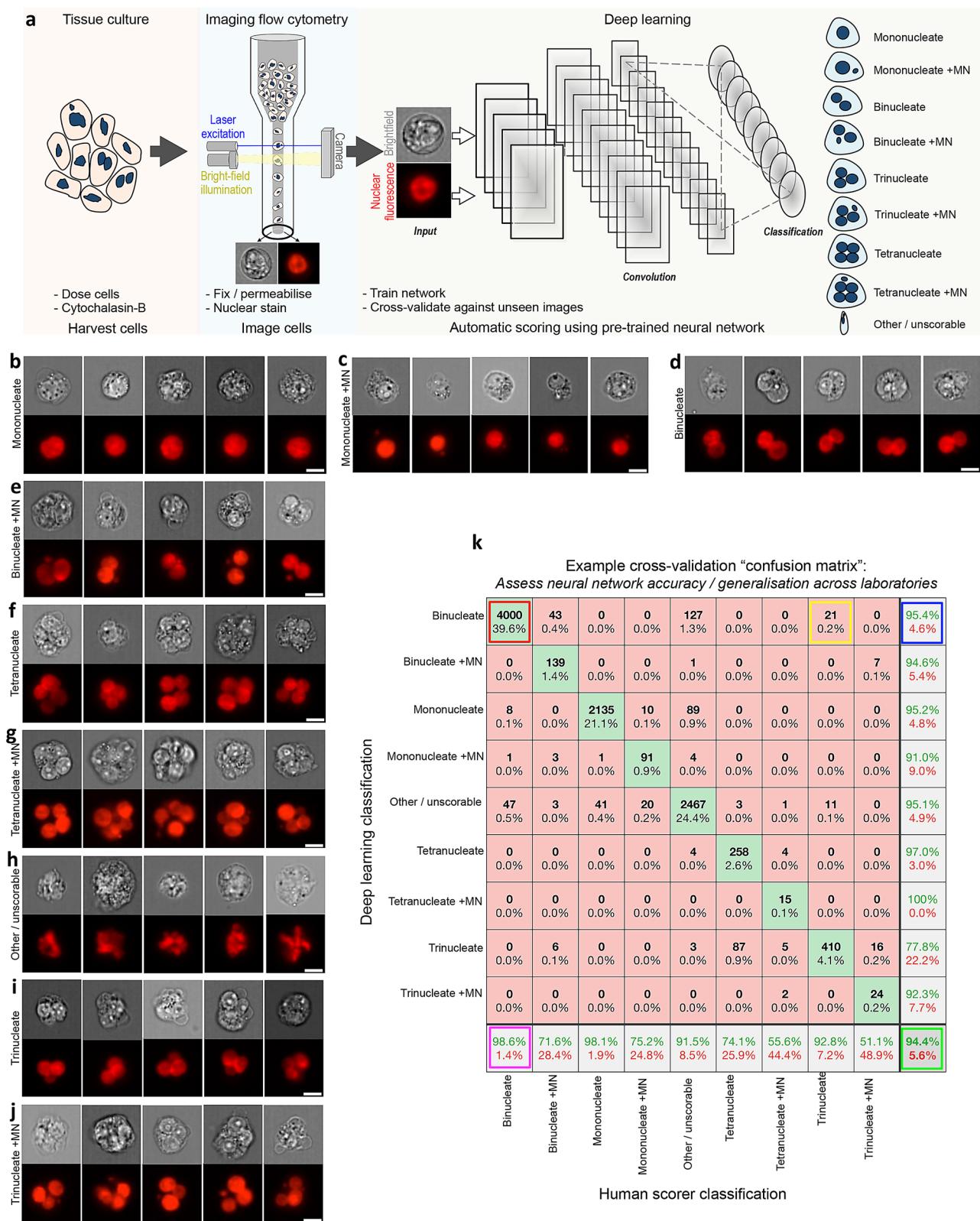### Statistical significance of micronucleus responses relative to control

Assessment of micronucleus response significance was conducted according to the framework described in Johnson et al. (2014). Response data was $\log_{10}$ transformed and assessed for normality and variance homogeneity by Shapiro–Wilk and Bartlett tests, respectively. Where the transformed data passed these tests ($p > 0.05$), comparisons of micronucleus responses relative to untreated negative controls employed one sided post hoc Dunnett's test with alpha 0.05. Data sets that failed these tests ($p < 0.05$) were analysed using the non-parametric post hoc Dunn's test.

Benchmark dose analysis

To compare the concentration–response relationships obtained from human expert scoring relative to those obtained from automatic scoring using the trained neural network, nonlinear regression analysis using the Benchmark Dose (BMD) framework was used. Using the freely available PROAST software, concentration–response data were analysed using both the exponential and the Hill model family recommended for the assessment of continuous toxicity data by the European Food Safety Authority (EFSA) (Hardy et al. 2017). In each analysis, combined data sets (i.e., across scoring methods) were analysed together with 'scoring method' specified as a potential covariate (Wills et al. 2016). More complex models with additional parameters were accepted if the fit significantly ($p < 0.05$; log-likelihood) improved. Here, as in previous work, we found that the log-steepness (*parameter d*) and maximum response (*parameter c*) could reasonably be held equal across concentration–response curves, whereas the parameters for background response (*parameter a*), potency (*parameter b*), and within-group variance (*var*) were found to be covariate-dependent (Slob and Setzer 2014). The BMD output describes the 'equipotent concentration' of the modelled concentration–response relationships in addition to the bounding, two-sided 90% confidence interval for each level of the covariate. The benchmark response (BMR) size (also termed the critical effect size) used was 50%, which represents a 50% increase in response relative to the background established in the vehicle (zero-concentration) control.

Results

Here, we investigate the ability of deep learning to provide generalised automation of CBMN assay scoring using imaging flow cytometry data acquired according to local protocols across three different laboratories (Cardiff, Cambridge and GSK). Figure 1a demonstrates our workflow. At the end of the assay, cells were fixed and permeabilised before fluorescent nuclear staining. The choice of nuclear stain varied across the different laboratories according to compatibility with the laser configuration of the local imaging cytometer. At Cambridge, cells were labelled with the blue-fluorescent dye Hoechst 33342 which was stimulated by a 405 nm laser with image capture using an ImageStream[X] cytometer. At Cardiff and GSK, ImageStream[X] MKII cytometers were used in conjunction with the red-emitting DRAQ5 nuclear stain and excitation by either a 488 nm or 642 nm laser (respectively). Full details of image acquisition settings at each laboratory are shown in Supplementary Table 1. Image acquisition speeds depended on cell concentrations, in

addition to the time taken to purge the flow stream and load each new sample; approximately ~2000–5000 cell images/min was typical.

After image collection, a template file created in the cytometer manufacturer's IDEAS software was used to automatically batch-save populations of single cells that

◄**Fig. 1** Automating the in vitro micronucleus assay using imaging flow cytometry and deep learning image classification. **a** Workflow: harvested cells were fixed and permeabilised before counterstaining the nuclei with a fluorescent DNA stain. Transmitted light brightfield (grey) and nuclear fluorescence (red) images were then automatically captured by high-throughput imaging flow cytometry. After initial training using a human-annotated image set, single cell images from the cytometer can be automatically classified using the neural network image classification algorithm. **b–j** Example image classifications according to a nine-class network developed to score the cytokinesis-block in vitro micronucleus assay in human lymphoblastoid TK6 cells. **k** Example cross-validation 'confusion matrix' obtained during preliminary network optimisations and presented here to demonstrate confusion matrix interpretation. The matrix represents an image set scored by humans that is 'unseen' during network training. The horizontal direction represents the human scorer classifications, whilst the vertical direction shows the automated output classifications from the network. The green diagonal represents correct, matching classifications: for example (indicated, red box) 4000 'binucleate' images, representing 39.6% of the total test image set, were classified correctly. Away from this diagonal, misclassifications are shown, e.g., (yellow box) 21 images (0.2%) labelled as 'trinucleates' by human scoring were incorrectly classified as 'binucleates' by the network. In the bottom-right corner (green box) the overall network accuracy and overall misclassification rate are shown for all nine classes (94.4% and 5.6%, respectively). In the white squares down the right-hand side of the matrix, the network precision, i.e., true positive/(true positive plus false positive) (green percentages) and the false discovery rate, i.e., 100-precision (red percentages) are shown for each classification. The horizontal bottom white row shows the network sensitivity, i.e., true positive/(true positive plus false negatives) (green percentages) and false negative rates (red percentages), respectively. Therefore, for example, 95.4% of the images classified as binucleates by the network were binucleates by human-scoring (blue box), whereas the trained model can be expected to correctly assign the binucleate class 98.6% of the time (magenta box). *Scale bars equal 5 microns* (color figure online)

additionally met acceptable focus criteria (see Methods). These cell populations served as the input into the deep learning scoring pipeline. This workflow is provided for download in both MATLAB and Python programming languages at the Biostudies database (accession no. S-BSST641). In brief, the download demonstrates initial image pre-processing to normalise image illumination across cytometers in addition to how to build and train the DeepFlow neural network using a human-scored training image set. After successful training, the saved network can subsequently be used to automate the scoring of new images. For example, Fig. 1b–j shows typical events classified by a pretrained, nine-class network with cell classes for mononucleates, binucleates, trinucleates and quadranucleates with or without micronucleus events in addition to a final class for 'other or unscorable' phenotypes.

As introduced above, an essential component of network testing involves cross validation with human-scored test images unseen during the training phase. We display this evaluation as a confusion matrix, which compares network outputs to the human scores for every image in the test set (explained, Fig. 1k). In the subsequently presented results,

we use this strategy to rigorously test the ability of a range of trained networks to enable automated CBMN assay scoring in both intra- and inter-laboratory contexts. In each instance, human-scored image sets were built from cell events pooled across the available compounds and exposures. This strategy was chosen to maximise the diversity of cellular phenotypes present, as well as to ensure that the rarer, micronucleated phenotypes that predominately manifested at higher exposures were well represented.

First, we tested the ability of a network trained on one laboratory's data to work well for unseen data from that same laboratory (i.e., 'single-laboratory testing') using imaging flow cytometry data collected at either Cardiff or Cambridge (Fig. 2). In this single laboratory context, images were randomly assigned to training (60%) and unseen testing (40%) groups. In both instances, the overall accuracies within this single-laboratory context were very high (91.3% and 90.5% for Cardiff and Cambridge, respectively). However, the compiled test sets were quite imbalanced in terms of the numbers of images per class, with network performance with some of the sparser classifications less well represented by the metric of overall accuracy.

For Cardiff (Fig. 2a), whereas accuracy in classification of the common parent nuclei classes (i.e., mononucleates, binucleates, trinucleates) was generally very good (> 97%), 20 out of a total of 78 events (~ 25%) human-scored as 'binucleate + MN' were misclassified as 'binucleates' by the network. Similarly, around 35% of the human-scored 'mononucleate + MN' events were outputted into the 'mononucleate' or 'other/unscorable' classes, with a further ~ 20% of 'tetranucleated' test images misclassified as 'trinucleates'. Despite scoring ~ 10,000 total events from the Cardiff cytometer, the very rarest cell phenotypes represented by the 'tetranucleate with MN' and 'trinucleate with MN' classes presented at very low frequency (~ 0.27% and 0.47%, respectively). This led to sparsity in the training set which appeared associated with the network missing micronucleus events, as the 'trinucleate + MN' images were often misclassified into the 'trinucleate' or 'tetranucleate' classes. In a similar manner, 'tetranucleate + MN' images were often misclassified into the 'trinucleate' or 'binucleate + MN' categories.

Similar results were observed within the Cambridge laboratory (Fig. 2b). Whereas accuracies with the 'mononucleate plus MN' and 'binucleate plus MN' classes showed slight improvement when compared against Cardiff, accuracies with the sparser, micronucleated tri- and tetranucleated cells again suffered (~ 44 and ~ 33% error rates, respectively).

We next considered the ability of the networks trained on single-laboratory data to generalise to the task of scoring the image data collected from the opposite Centre (Fig. 3). This was expected to be a difficult task given that the networks had been trained initially with fairly small numbers of images, because the two laboratories
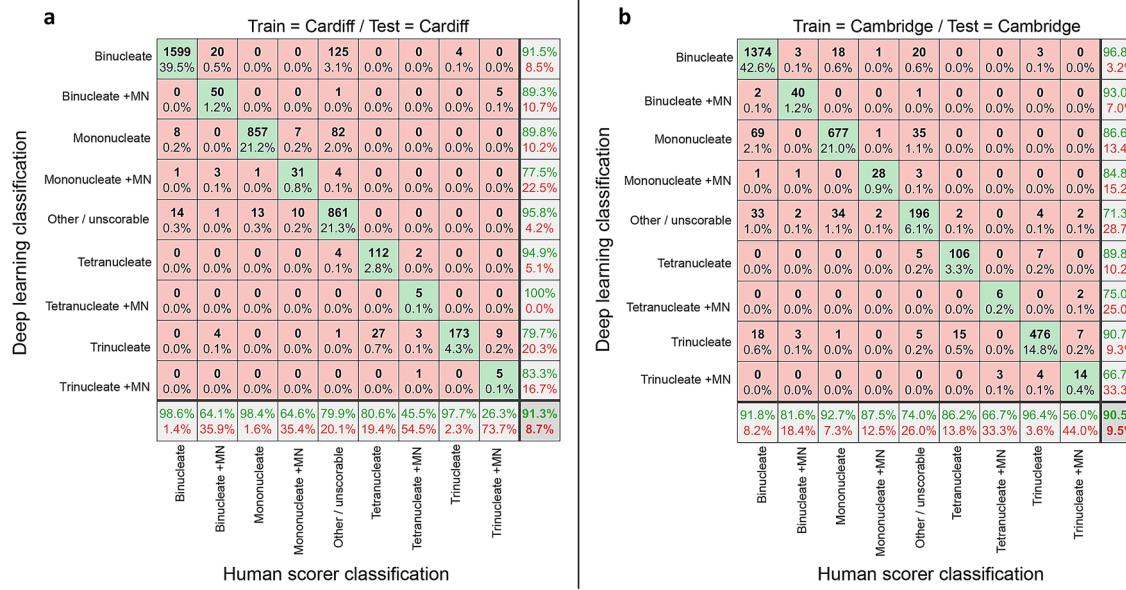
Fig. 2 Assessing automated scoring accuracies using intra-laboratory train and test data. **a**, **b** Confusion matrices comparing human scoring versus deep learning image classifications for test image sets of approximately four 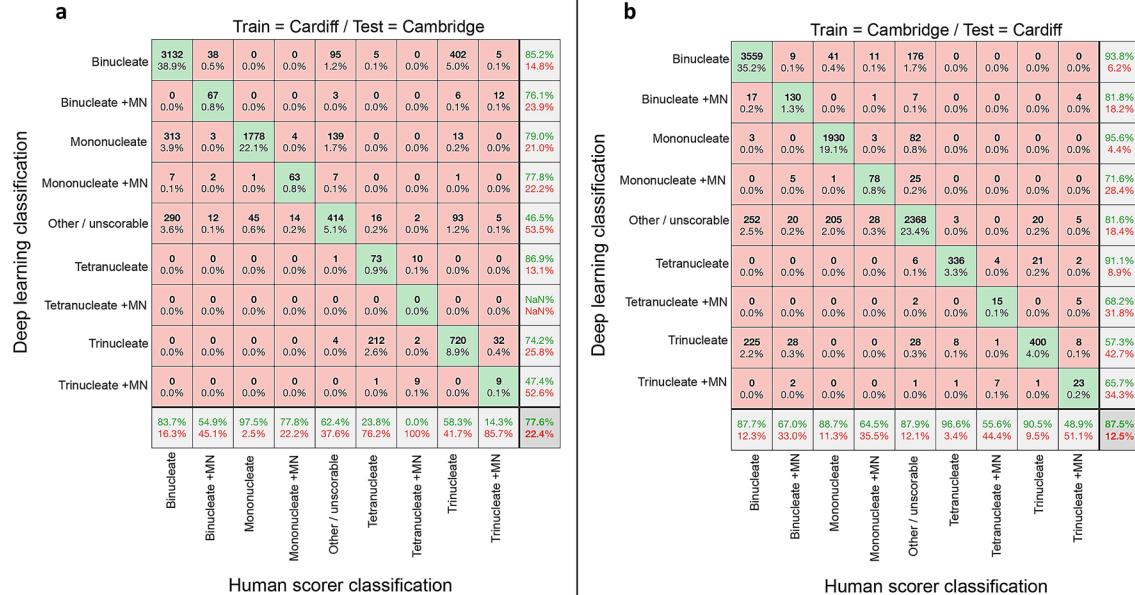thousand unseen images. In each instance, the results reflect the outputs from nine-class networks trained and tested exclusively on image-data from one imaging cytometer at either the **a** Cardiff or **b** Cambridge laboratories

Fig. 3 Assessment of automated network scoring accuracies using inter-laboratory test data. **a**, **b** Confusion matrices comparing human scoring versus deep learning image classifications for test image sets of approximately ten thousand unseen images. In each instance, the results reflect the outputs from nine-class networks trained exclusively on image data from one laboratory's imaging cytometer before cross-validation testing against image data collected at a different laboratory. **a** Network accuracies after training using Cardiff data before testing on unseen Cambridge data. **b** Network accuracies after training on Cambridge data then testing on unseen Cardiff data

had utilised different cytometer models (IS^X versus IS^X Mk II) and nuclear stains (Hoechst at Cambridge or DRAQ5 at Cardiff). This presented the likelihood of overfitting during training—yielding networks highly adapted to the task of scoring data from that particular laboratory.

Despite these factors, at first glance the overall accuracies appeared quite encouraging at 77.6% for the Cardiff-trained network classifying the Cambridge images (Fig. 3a) and 87.5% for the Cambridge network classifying Cardiff images (Fig. 3b). Comparing across the individual classes, it was apparent that the Cambridge-trained model generalised slightly better to the task of scoring the Cardiff data than was observed vice versa. Closer examination, however, showed that the metric of overall accuracy was weighted by the prevalence of the easily identified 'mononucleate' and 'binucleate' phenotypes, which masked assessment of the ability of the networks to identify the micronucleated classes representing DNA-damage events (Fig. 3a, b). In this regard, in almost all instances, the accuracy of micronucleated event detection suffered considerably compared to the results achieved with laboratory-matched test data (Fig. 2).

With these single-laboratory results established, the images from Cambridge and Cardiff were combined together. This increased the diversity of training exemplifications considerably given the use of two different nuclear stains, two compounds, different imaging cytometers and no 'hold out' requirement for cross validation testing. Training a new DeepFlow neural network on this combined training set (~ 19,000 images) took approximately 1 h using modest hardware (single RTX 2080 GPU). The resulting network was then cross validated using a test set where both the bioassay and imaging cytometry were conducted at an entirely new, third laboratory (GSK). Scoring ~ 5,000 test-images took around 6 s on the RTX 2080 hardware or ~ 82 s on a single CPU. This time, the network showed much better ability to generalise to the task of successfully scoring the images from the new laboratory (Fig. 4a). Across the four core classes central to utilisation of CBMN assay (i.e., 'mononucleate', 'mononucleate plus MN', 'binucleate' and 'binucleate plus MN'), and with no user input or configuration required, the network achieved 98%, 82%, 94%, and 85% accuracies, respectively.
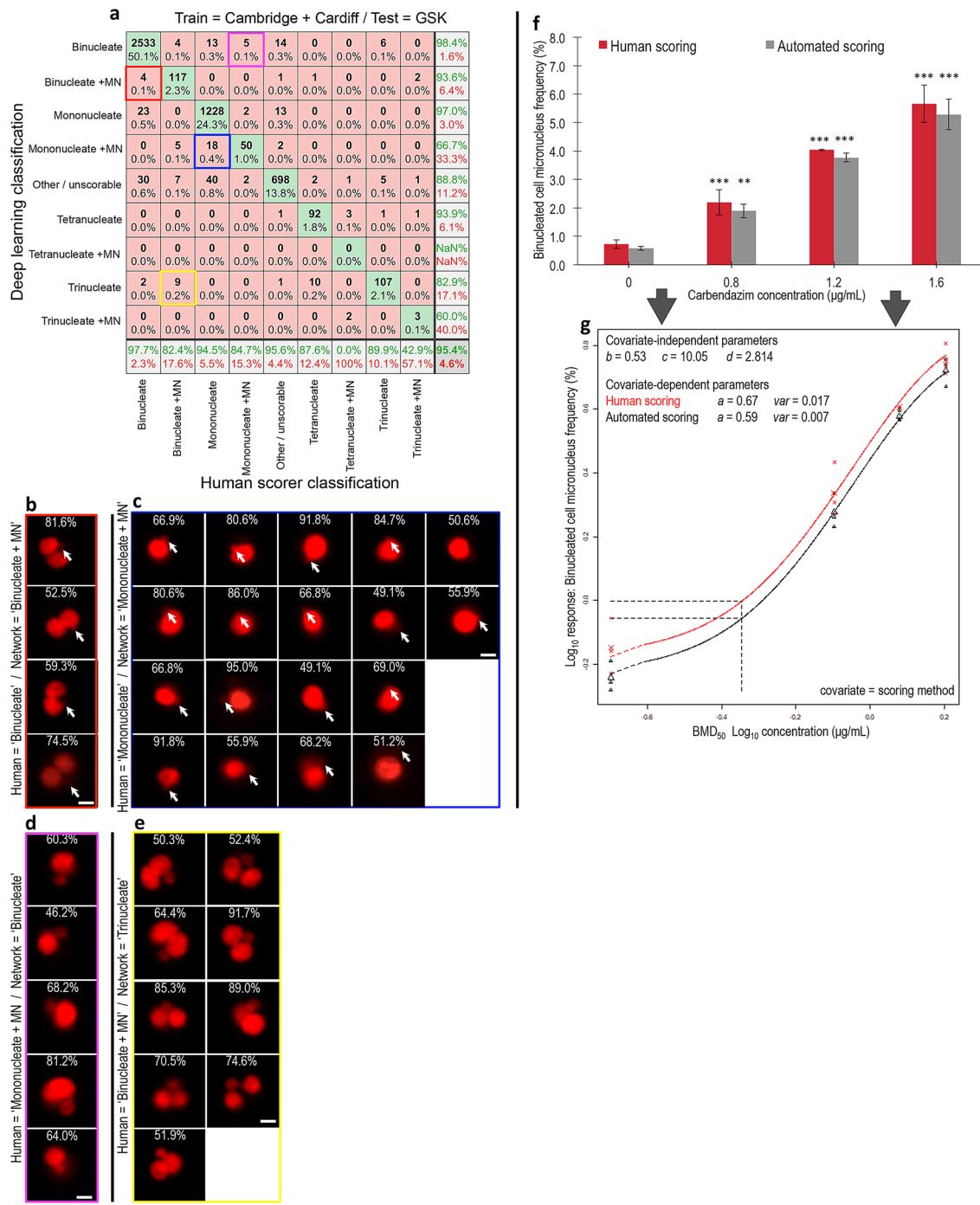
We then examined failure cases, starting with 22 instances where the network detected micronucleus events in cells scored by humans as just mono- or binucleated (Fig. 4a). Surprisingly, many did, in fact, appear to have faint or partially occluded potential micronucleus or nuclear bud events that would have been extremely difficult for the human scorer to detect (Fig. 4b, c). Similarly, visualisation of cell events scored by humans as either 'mononucleate with MN' or 'binucleate with MN', but outputted by the network as 'binucleate' or 'trinucleate' showed that these images often contained very large micronucleus events (Fig. 4d, e). Indeed, some of these likely exceeded the upper size limitation typically imposed on micronucleus classifications (i.e., ≤ 1/3 diameter of the parent nuclei) suggesting additional validity to the network's outputs.

Progressing towards the less frequent cell phenotypes, the accuracies achieved with the 'trinucleate' and 'tetranucleate' cell classes were also good at 90% and 88%, respectively. However, detection of these cell types with micronucleus events was either quite poor or failed entirely. Again, this outcome was likely related to extreme sparsity in occurrence (< 0.25% frequency in the training data). In an attempt to improve accuracies with these classes, we tried both class weighting the classification layer and combining tri- and tetranucleated events with and without micronucleus events into a single, 'polynucleated' class (Supplementary Figure 2). Whereas both strategies somewhat improved the classification accuracies with these rare events, they were also found to compromise the accuracies achieved with one or more of the four core phenotypes more central to successful CBMN assay scoring.

Given that the frequency of binucleated cells with or without micronucleus events represents the core readout for successful DNA damage assessment by the CBMN assay, after validating the network we proceeded to assess the binucleated-cell micronucleus frequency for a three concentration plus control experiment conducted in triplicate with carbendazim at the GSK laboratory. For each concentration and replicate, 2000 cell images were scored both manually and automatically. Visually, the resultant concentration–response relationships appeared similar across the human and neural network scoring approaches, with the human scores consistently fractionally higher for each concentration group (Fig. 4f). To better understand the consequences of this using a recognised, quantitative framework for genotoxic potency estimation, the concentration–response relationships were fitted using both the exponential and the Hill model families recommended for the assessment of continuous toxicity data using Benchmark Dose (BMD) analysis (Hardy et al. 2017). With scoring method specified as a potential covariate, model fitting with the PROAST package resulted in covariate-dependent parameterisation for the background response (parameter *a*) and for within-group variation (*var*). For both model families, this parameterisation subsequently allowed rejection of scoring method as covariate, yielding the *same* estimation for the equipotent, benchmark concentration from both manual and automated methods (Fig. 4g). Model fits to the data are presented in Supplementary Figure 3.

Discussion

The CBMN assay represents a globally significant method for the identification and quantification of chromosomal damage (Fenech 2000, 2020; OECD 2016). Its utility reaches beyond regulatory compound screening to encompass inter-individual monitoring of wide-ranging lifestyle, occupational and environmental factors (Fenech 2020;

Kirsch-Volders et al. 2011; Wang et al. 2019). Despite this, continued reliance upon time-consuming and user-subjective manual scoring represents a bottleneck to broadening practical utilisation (Seager et al. 2014; Verma et al. 2017, 2018). In this pilot study, we show that rapid image acquisition by imaging flow cytometry in conjunction with deep learning image classification represents a capable platform for automated, inter-laboratory operation. We share our strategy via openly accessible frameworks.

As an image acquisition method, imaging flow cytometry is now well established as a means for high-throughput CBMN data capture with concomitant image archiving potential (Rodrigues et al. 2014a, 2016a, 2018). Moreover, this is achieved with simple sample preparation involving a single nuclear stain and brightfield to provide the context that events lie inside parent cells (Rodrigues et al. 2018). Comparison studies have shown that the captured images contain concentration–response information that aligns to

◀**Fig. 4** Network accuracy and concentration–response assessment using unseen test data from a new laboratory. **a** Confusion matrix showing human versus deep learning image classifications for a test image set of approximately five thousand unseen images. Here, the neural network was trained using image data from both the Cambridge and Cardiff laboratories before testing on new, unseen imaging cytometry data acquired at a third laboratory (GSK). **b** Cell events human scored as 'binucleates' but classified as 'binucleate plus MN' by the neural network (i.e., red square in **a**). **c** Cell events human scored as 'mononucleates' but classified as 'mononucleate with MN' by the neural network (i.e., blue square in **a**). **b**, **c** Close examination of the purportedly misclassified cells shows that many display indistinct events that might be micronucleus or nuclear buds missed by the human scorer (indicated, white arrows). **d** Cell events human scored as 'mononucleate with MN' but classified as 'binucleate' by the neural network (i.e., magenta square in **a**). **e** Events human scored as 'binucleate with MN' classified as 'trinucleate' by the neural network (i.e., yellow square in **a**). **d**, **e** In both instances, some of the human-scored micronucleus events encroach upon the 1/3 parent nuclei upper-size limitation typically imposed on micronucleus classifications. **b**–**e** For each event, the white percentages represent neural network confidence in the outputted classification. **f** Binucleated-cell micronucleus frequencies for a three concentration plus control concentration–response experiment performed in triplicate for carbendazim exposure to TK6 cells. Scores were established from image sets of 2,000 events per replicate by human scoring or by the cross-validated network established in (**a**). (*) (**) (***) indicate statistical significance relative to control at $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively. **g** Covariate benchmark dose (BMD) modelling using concentration–response data from either the human (black) or automated neural network (red) scores established in (**f**). The horizontal and vertical dashed lines represent interpolation to determine the equipotent, benchmark concentration for a benchmark response size of 50%. Regardless of human or automated scoring, the model predicts the same benchmark concentration. *Scale bars equal 5 microns* (color figure online)

results obtained from 'gold standard' manual microscopy scoring (Verma et al. 2018). Whereas conventional flow cytometry offers faster throughput, it lacks this image-based validation and archiving capability whilst additionally requiring cell lysis to operate. This prevents utilisation of the cytokinesis-block version of the assay, complicating quantitation of mononucleated, binucleated and different classes of multinucleated cells in addition to necessitating means to reliably exclude DNA fragments arising from apoptotic and necrotic cells from micronucleus count data (Bryce et al. 2007; Lukamowicz et al. 2011; Rodrigues et al. 2018).
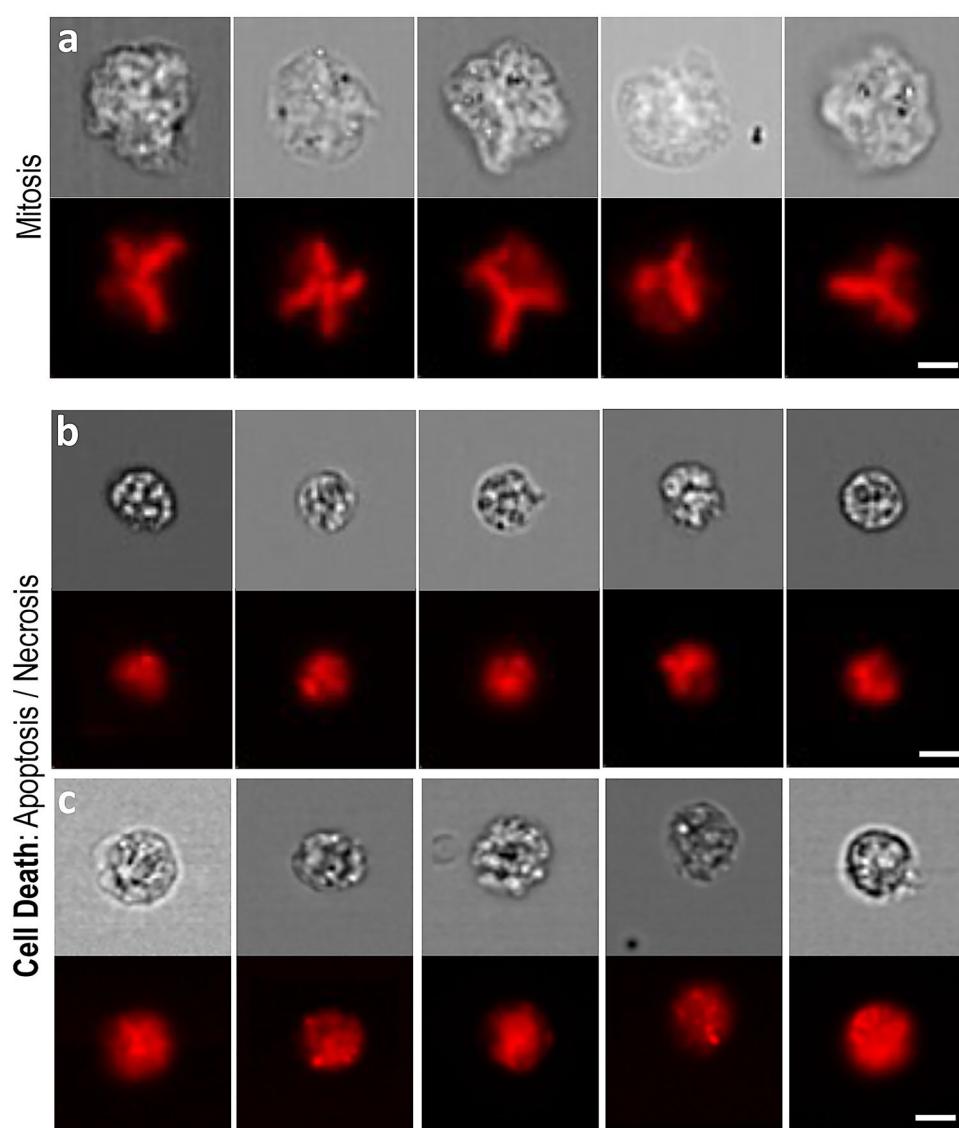
Beyond image collection, automated scoring of imaging flow cytometry data—as with other automated microscopy strategies—has thus far largely relied upon traditional, threshold-based image classification techniques. These require image analysis expertise to implement, alongside user-configuration and tuning to maintain performance (Rodrigues et al. 2018; Seager et al. 2014; Verma et al. 2017). Unfortunately, much as with traditional manual scoring, this is time-consuming and subjective.

In contrast, once successfully trained, the results achieved here suggest that deep learning image classification has the potential to eliminate these expertise and user-input requirements, dramatically reducing the time to results. This comes from encompassing image diversity during network training and harnessing it to improve the consistency and robustness of subsequent classifications. To this end, here we show that utilisation of diverse training data curated across two laboratories utilising different nuclear stains, multiple compounds and two different cytometer models yielded a capable neural network for scoring automation. Without user configuration, the network was able to classify data collected from an entirely new laboratory with > 82% accuracy for each of the four cell phenotypes central to CBMN performance (i.e., mononucleate and binucleate cells with or without micronucleus events) in addition to successfully classifying tri- and tetranucleated cells (> 88% accuracy) and unscorable events (96% accuracy). Importantly, these seven classes encompassed virtually all of the cell images encountered (> 99%). Success at micronucleus detection in both mononucleate and binucleate cell classes further suggests that this single network could be used to automate scoring of both mononuclear and cytokinesis-block versions of the assay.

Despite this success with the assay classes central to CBMN scoring, the scarce, tri- and tetranucleated phenotypes with micronucleus events proved more challenging. Commonly employed methods such as class weighting or class combination offered little in the way of accuracy improvements, and often compromised accuracy with the other classes. These findings suggest that significant increases in the representation of these sparse events during training will likely be required to improve success. In this context, imaging flow cytometry is well suited to examine whether an improved image bank leads to enhanced accuracy in scoring given the high rates of image capture achievable. Our results also suggests that class reduction does not necessarily simplify the classification problem and may instead cause ambiguities. In this way, future expansions to the number of classes to encompass all distinctive cellular phenotypes may represent a route to improving overall network performance.

In this regard, we identified additional, potentially scorable cell phenotypes (Fig. 5). In particular, cell death events (i.e., due to apoptosis and necrosis) were visually apparent, but we were unable to determine apoptotic from necrotic events using just the brightfield and nuclear fluorescence images alone. Cells caught during mitosis also represented distinctive events. At the same time, we were less convinced that more subtle phenotypes relevant to the expanded, CBMN cytome assay such as nuclear buds and bridges could reliably and consistently be detected—given the relatively low resolution of the image data (Fenech 2007). However, it is important to note that previous studies demonstrating

Fig. 5 Other scorable cell phenotypes captured by imaging flow cytometry. **a** Cells undergoing mitosis were visually apparent according to metaphase spread-type nuclear fluorescence imagery (red) alongside large, brightfield-delineated cell sizes (grey). **b**, **c** Cell death events displayed shrunken cell sizes in conjunction with granular brightfield and fluorescence imagery. In the case of cell death, two distinctive cell phenotypes appeared visually separable according to cell size and the number, size and extent of nuclear foci formation (**b** versus **c**). Whether these observations represented distinct apoptotic versus necrotic events was unclear from the nuclear fluorescence and brightfield information alone. *Scale bars equal 5 microns* (color figure online)

capture of these phenotypes by imaging flow cytometry have utilised both the 60× ImageStream objective lens in addition to hypotonic treatments to swell cell volumes prior to imaging (Rodrigues et al. 2018; Rodrigues 2019). Hypotonic treatments were not used here but may improve image capture of these more subtle phenotypes. With regards to network class expansion to encompass these events—or, indeed for simultaneous measurement of other endpoints—the ImageStream platform is capable of multiplexed imaging. Additional channels might, therefore, be used to simultaneously measure other DNA-damage pathways [e.g., ϒH2AX for DNA double-strand breaks (Smart et al. 2011)], or to improve the reliability of ground truth image curations through use of additional fluorescent markers to differentiate events such as apoptotic from necrotic cells.

Manual scoring of the images for this experiment was more challenging than the exemplar images shown

might suggest. Fundamentally, the acquired images are relatively low resolution (i.e., cells occupy ~ 64 × 64 pixels) and further image degradation is always present as a result of the capture of moving objects by time delay integration. The acquired images also represent a central, 2-D projection of a 3-D cell object. This means that nuclei and micronucleus events may overlap each other, or they may lie outside of the plane of optimal focus (Rodrigues et al. 2018). These factors all served to make ground truth assignments more complicated, even for experienced CBMN scorers. Whereas network accuracy assessments by confusion matrix provided a more representative breakdown of outputs when compared to simplistic overall accuracy measures, it is a relatively stringent success measure, because any ambiguity in human score assignment is not captured. A potential advantage of automated network classification approach

is, therefore, likely greater consistency—even in error—than arises from manual scoring.

Regarding image focussing, the ImageStream platform offers 'extended depth of field' (EDF) technology, whereby image deconvolution is used to improve the utility of out of focus events through projection onto a single plane (Ortyn et al. 2007). Whereas previous studies have shown this technique can improve accuracy in 'spot counting' applications, the strategy has been reported less helpful for the provision of improved CBMN data (Parris et al. 2015; Rodrigues 2018; Rodrigues et al. 2014a). This was attributed to a slight degradation in overall image resolution, compromising differentiation of micronucleus events from parent nuclei (Rodrigues 2018). On a similar theme, the ImageStream platform is also configurable with 20×, 40× or 60× objective lenses. Here, image collection was via the 'standard', 40× objective across all laboratories. This approach was chosen as previous work has shown that whilst greater resolution is achievable with the 60× objective, focus depth also decreases, reinforcing the out of plane difficulties described above (Rodrigues et al. 2018).

Whilst considering the nature and utility of imaging flow cytometry data, a relevant comparison is to that provided by other automated imaging methods, such as slide scanning platforms. In addition to the potential for higher resolution imaging, here an overlooked advantage comes from the ability to use slide-based preparations created by cytocentrifugation. This technique causes the flattening and spreading of cellular content, presenting nuclear objects on a more two dimensional plane (Fitzgerald and Hosking 1982; Shanholtzer et al. 1982). From a practical perspective, however, this also necessitates the consistent preparation of high-quality slides with optimal cell densities (Rodrigues et al. 2018). Meanwhile, a major advantage of the imaging flow cytometry approach is that single cell image data is inherently acquired by the fluidics-based processing of individualised cells.

Conclusions

As a platform for the CBMN assay, imaging flow cytometry combines the high throughput and multiplexing potential of flow cytometry with the image-based validation and archiving attributes of automated microscopy. Here we demonstrate accurate, automated assay scoring using a neural network for data collected in a laboratory wholly separate to that in which the algorithm was trained. This proves that without any human configuration, the machine is able to correctly anticipate the decisions of the expert human on unseen images in a new setting. For the first time, this suggests the possibility for generalised scoring automation through dissemination of a pretrained network for the ImageStream platform established from ground truth agreed by a single, expert group. Such an approach would provide the ultimate in terms of standardisation and result reliability, but more importantly could enable adoption of the assay beyond current practitioners as local expertise in scoring and/or image analysis would no longer be required. For these reasons, we believe that full development of this automated, accessible, inter-laboratory approach would represent a truly twenty-first century method with significant potential to transform CBMN utility across industry, research and clinical domains.

Declarations

References

Allemang A, Thacker R, DeMarco RA, Rodrigues MA, Pfuhler S (2021) The 3D reconstructed skin micronucleus assay using imaging flow cytometry and deep learning: a proof-of-principle investigation. Mutat Res Genet Toxicol Environ Mutagen 865:503314. https://doi.org/10.1016/j.mrgentox.2021.503314

Avlasevich SL, Bryce SM, Cairns SE, Dertinger SD (2006) In vitro micronucleus scoring by flow cytometry: differential staining of micronuclei versus apoptotic and necrotic chromatin enhances assay reliability. Environ Mol Mutagen 47(1):56–66. https://doi.org/10.1002/em.20170

Blasi T, Hennig H, Summers HD et al (2016) Label-free cell cycle analysis for high-throughput imaging flow cytometry. Nat Commun 7(1):10256. https://doi.org/10.1038/ncomms10256

Bryce SM, Bemis JC, Avlasevich SL, Dertinger SD (2007) In vitro micronucleus assay scored by flow cytometry provides a comprehensive evaluation of cytogenetic damage and cytotoxicity. Mutat Res 630(1–2):78–91. https://doi.org/10.1016/j.mrgentox.2007.03.002

Bryce SM, Avlasevich SL, Bemis JC et al (2008) Interlaboratory evaluation of a flow cytometric, high content in vitro micronucleus assay. Mutat Res 650(2):181–195. https://doi.org/10.1016/j.mrgentox.2007.11.006

Bryce SM, Avlasevich SL, Bemis JC, Phonethepswath S, Dertinger SD (2010) Miniaturized flow cytometric in vitro micronucleus assay represents an efficient tool for comprehensively characterizing genotoxicity dose-response relationships. Mutat Res 703(2):191–199. https://doi.org/10.1016/j.mrgentox.2010.08.020

Bryce SM, Avlasevich SL, Bemis JC et al (2013) Flow cytometric 96-well microplate-based in vitro micronucleus assay with human TK6 cells: protocol optimization and transferability assessment. Environ Mol Mutagen 54(3):180–194. https://doi.org/10.1002/em.21760

Caicedo JC, Goodman A, Karhohs KW et al (2019) Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nat Methods 16(12):1247–1253. https://doi.org/10.1038/s41592-019-0612-7

Darzynkiewicz Z, Smolewski P, Holden E et al (2011) Laser scanning cytometry for automation of the micronucleus assay. Mutagenesis 26(1):153–161. https://doi.org/10.1093/mutage/geq069

Decordier I, Kirsch-Volders M (2006) The in vitro micronucleus test: from past to future. Mutat Res 607(1):2–4. https://doi.org/10.1016/j.mrgentox.2006.04.008

Decordier I, Papine A, Plas G et al (2009) Automated image analysis of cytokinesis-blocked micronuclei: an adapted protocol and a validated scoring procedure for biomonitoring. Mutagenesis 24(1):85–93. https://doi.org/10.1093/mutage/gen057

Decordier I, Papine A, Vande Loock K, Plas G, Soussaline F, Kirsch-Volders M (2011) Automated image analysis of micronuclei by IMSTAR for biomonitoring. Mutagenesis 26(1):163–168. https://doi.org/10.1093/mutage/geq063

Eulenberg P, Köhler N, Blasi T et al (2017) Reconstructing cell cycle and disease progression using deep learning. Nat Commun 8(1):463. https://doi.org/10.1038/s41467-017-00623-3

Fenech M (2000) The in vitro micronucleus technique. Mutat Res 455(1–2):81–95. https://doi.org/10.1016/s0027-5107(00)00065-8

Fenech M (2007) Cytokinesis-block micronucleus cytome assay. Nat Protoc 2(5):1084–1104. https://doi.org/10.1038/nprot.2007.77

Fenech M (2020) Cytokinesis-block micronucleus cytome assay evolution into a more comprehensive method to measure chromosomal instability. Genes 11(10):1203. https://doi.org/10.3390/genes11101203

Fitzgerald MG, Hosking CS (1982) Cell structure and percent viability by a slide centrifuge technique. J Clin Pathol 35(2):191–194. https://doi.org/10.1136/jcp.35.2.191

François M, Hochstenbach K, Leifert W, Fenech MF (2014) Automation of the cytokinesis-block micronucleus cytome assay by laser scanning cytometry and its potential application in radiation biodosimetry. BioTechn 57(6):309–312. https://doi.org/10.2144/000114239

Hardy A, Benford D, Halldorsson T et al (2017) Update: use of the benchmark dose approach in risk assessment. EFSA J 15(1):4658. https://doi.org/10.2903/j.efsa.2017.4658

Johnson GE, Soeteman-Hernández LG, Gollapudi BB et al (2014) Derivation of point of departure (PoD) estimates in genetic toxicology studies and their potential applications in risk assessment. Environ Mol Mutagen 55(8):609–623. https://doi.org/10.1002/em.21870

Kirsch-Volders M, Plas G, Elhajouji A et al (2011) The in vitro MN assay in 2011: origin and fate, biological significance, protocols, high throughput methodologies and toxicological relevance. Arch Toxicol 85(8):873–899. https://doi.org/10.1007/s00204-011-0691-4

Lukamowicz M, Kirsch-Volders M, Suter W, Elhajouji A (2011) In vitro primary human lymphocyte flow cytometry based micronucleus assay: simultaneous assessment of cell proliferation, apoptosis and MN frequency. Mutagenesis 26(6):763–770. https://doi.org/10.1093/mutage/ger044

Maertens RM, White PA (2015) RE: Recommendations, evaluation and validation of a semi-automated, fluorescent-based scoring protocol for micronucleus testing in human cells. Mutagenesis 30(2):311–312. https://doi.org/10.1093/mutage/geu066

Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D (2019) Deep learning for cellular image analysis. Nat Methods 16(12):1233–1246. https://doi.org/10.1038/s41592-019-0403-1

OECD (2016) Test guidline 487 guideline for the testing of chemicals, in vitro mammalian cell micronucleus test. Organ Econ Cooper. https://doi.org/10.1787/9789264264861-en

Ortyn WE, Perry DJ, Venkatachalam V, Liang L, Hall BE, Frost K, Basiji DA (2007) Extended depth of field imaging for high speed cell analysis. Cytometry A 71(4):215–231. https://doi.org/10.1002/cyto.a.20370

Parris CN, Adam Zahir S, Al-Ali H, Bourton EC, Plowman C, Plowman PN (2015) Enhanced γ-H2AX DNA damage foci detection using multimagnification and extended depth of field in imaging flow cytometry. Cytometry A 87(8):717–723. https://doi.org/10.1002/cyto.a.22697

Rodrigues MA (2018) Automation of the in vitro micronucleus assay using the Imagestream imaging flow cytometer. Cytometry A 93(7):706–726. https://doi.org/10.1002/cyto.a.23493

Rodrigues MA (2019) An automated method to perform the in vitro micronucleus assay using multispectral imaging flow cytometry. JoVE 147:59324. https://doi.org/10.3791/59324

Rodrigues MA, Beaton-Green LA, Kutzner BC, Wilkins RC (2014a) Automated analysis of the cytokinesis-block micronucleus assay for radiation biodosimetry using imaging flow cytometry. Radiat Environ Biophys 53(2):273–282. https://doi.org/10.1007/s00411-014-0525-x

Rodrigues MA, Beaton-Green LA, Kutzner BC, Wilkins RC (2014b) Multi-parameter dose estimations in radiation biodosimetry using the automated cytokinesis-block micronucleus assay with imaging flow cytometry. Cytometry A 85(10):883–893. https://doi.org/10.1002/cyto.a.22511

Rodrigues MA, Beaton-Green LA, Wilkins RC (2016a) Validation of the cytokinesis-block micronucleus assay using imaging flow cytometry for high throughput radiation biodosimetry. Health Phys 110(1):29–36. https://doi.org/10.1097/hp.0000000000000371

Rodrigues MA, Probst CE, Beaton-Green LA, Wilkins RC (2016b) Optimized automated data analysis for the cytokinesis-block micronucleus assay using imaging flow cytometry for high throughput radiation biodosimetry. Cytometry A 89(7):653–662. https://doi.org/10.1002/cyto.a.22887

Rodrigues MA, Beaton-Green LA, Wilkins RC, Fenech MF (2018) The potential for complete automated scoring of the cytokinesis block micronucleus cytome assay using imaging flow cytometry. Mutat Res Genet Toxicol Environ Mutagen 836:53–64. https://doi.org/10.1016/j.mrgentox.2018.05.003

Rossnerova A, Spatova M, Schunck C, Sram RJ (2011) Automated scoring of lymphocyte micronuclei by the MetaSystems Metafer image cytometry system and its application in studies of human mutagen sensitivity and biodosimetry of genotoxin exposure. Mutagenesis 26(1):169–175. https://doi.org/10.1093/mutage/geq057

Schunck C, Johannes T, Varga D, Lörch T, Plesch A (2004) New developments in automated cytogenetic imaging: unattended scoring of dicentric chromosomes, micronuclei, single cell gel electrophoresis, and fluorescence signals. Cytogenet Genome Res 104:383–389. https://doi.org/10.1159/000077520

Seager AL, Shah UK, Brüsehafer K et al (2014) Recommendations, evaluation and validation of a semi-automated, fluorescent-based scoring protocol for micronucleus testing in human cells. Mutagenesis 29(3):155–164. https://doi.org/10.1093/mutage/geu008

Shanholtzer CJ, Schaper PJ, Peterson LR (1982) Concentrated gram stain smears prepared with a cytospin centrifuge. J Clin Microbiol 16(6):1052

Slob W, Setzer RW (2014) Shape and steepness of toxicological dose-response relationships of continuous endpoints. Crit Rev Toxicol 44(3):270–297. https://doi.org/10.3109/10408444.2013.853726

Smart DJ, Ahmedi KP, Harvey JS, Lynch AM (2011) Genotoxicity screening via the γH2AX by flow assay. Mutat Res Genet Toxicol Environ Mutagen 715(1):25–31. https://doi.org/10.1016/j.mrfmmm.2011.07.001

Smolewski P, Ruan Q, Vellon L, Darzynkiewicz Z (2001) Micronuclei assay by laser scanning cytometry. Cytometry 45(1):19–26. https://doi.org/10.1002/1097-0320(20010901)45

Szegedy C, Wei L, Yangqing J et al (2015) Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) https://arxiv.org/abs/1409.4842.

Varga D, Johannes T, Jainta S et al (2004) An automated scoring procedure for the micronucleus test by image analysis. Mutagenesis 19(5):391–397. https://doi.org/10.1093/mutage/geh047

Verhaegen F, Vral A, Seuntjens J, Schipper NW, de Ridder L, Thierens H (1994) Scoring of radiation-induced micronuclei in cytokinesis-blocked human lymphocytes by automated image analysis. Cytometry 17(2):119–127. https://doi.org/10.1002/cyto.990170203

Verma JR, Rees BJ, Wilde EC et al (2017) Evaluation of the automated microflow and metafer platforms for high-throughput micronucleus scoring and dose response analysis in human lymphoblastoid TK6 cells. Arch Toxicol 91(7):2689–2698. https://doi.org/10.1007/s00204-016-1903-8

Verma JR, Harte DSG, Shah UK et al (2018) Investigating FlowSight imaging flow cytometry as a platform to assess chemically induced micronuclei using human lymphoblastoid cells in vitro. Mutagenesis 33(4):283–289. https://doi.org/10.1093/mutage/gey021

Wang Q, Rodrigues MA, Repin M et al (2019) Automated triage radiation biodosimetry: integrating imaging flow cytometry with high-throughput robotics to perform the cytokinesis-block micronucleus assay. Radiat Res 191(4):342–351. https://doi.org/10.1667/rr15243.1

Wilkins RC, Rodrigues MA, Beaton-Green LA (2017) The application of imaging flow cytometry to high-throughput biodosimetry. Genome Integr 8:7. https://doi.org/10.4103/2041-9414.198912

Willems P, August L, Slabbert J et al (2010) Automated micronucleus (MN) scoring for population triage in case of large scale radiation events. Int J Radiat Biol 86(1):2–11. https://doi.org/10.3109/09553000903264481

Wills JW, Johnson GE, Doak SH, Soeteman-Hernández LG, Slob W, White PA (2016) Empirical analysis of BMD metrics in genetic toxicology part I: in vitro analyses to provide robust potency rankings and support MOA determinations. Mutagenesis 31(3):255–263. https://doi.org/10.1093/mutage/gev085

Zhang C, Bengio S, Hardt M, Recht B, Vinyals O (2017) Understanding deep learning requires rethinking generalization. ICLR. https://arxiv.org/abs/1611.03530