Modular deep learning enables automated identification of monoclonal cell lines

Brodie Fischbacher✉, Sarita Hedaya✉, Brigham J. Hartley, Zhongwei Wang, Gregory Lallos, Dillion Hutson, Matthew Zimmer, Jacob Brammer, The NYSCF Global Stem Cell Array Team, and Daniel Paull✉

Monoclonalization refers to the isolation and expansion of a single cell derived from a cultured population. This is a valuable step in cell culture that serves to minimize a cell line’s technical variability downstream of cell-altering events, such as reprogramming or gene editing, as well as for processes such as monoclonal antibody development. However, traditional methods for verifying clonality do not scale well, posing a critical obstacle to studies involving large cohorts. Without automated, standardized methods for assessing clonality post hoc, methods involving monoclonalization cannot be reliably upscaled without exacerbating the technical variability of cell lines. Here, we report the design of a deep learning workflow that automatically detects colony presence and identifies clonality from cellular imaging. The workflow, termed Monoqlo, integrates multiple convolutional neural networks and, critically, leverages the chronological directionality of the cell-culturing process. Our algorithm design provides a fully scalable, highly interpretable framework that is capable of analysing industrial data volumes in under an hour using commodity hardware. We focus here on monoclonalization of human induced pluripotent stem cells, but our method is generalizable. Monoqlo standardizes the monoclonalization process, enabling colony selection protocols to be infinitely upscaled while minimizing technical variability.

The isolation and subsequent expansion of a single cell derived from a cultured population establishes monoclonality and is often considered an essential step in developing high-quality cell lines. This procedure is intended to minimize or eliminate genomic and phenotypic heterogeneity in an attempt to maximize the uniformity of cell lines. For example, a newly genome-engineered cell population may comprise an admixture of cells with divergent alleles, zygosity and epigenetic characteristics. A homogenous cell line can thus only be re-established by ensuring all cells in the population are descendent from a single ancestral cell that was isolated downstream of any event with a high proclivity to introduce variation. This step is referred to as monoclonalization.

An example of a cell-culturing process in which monoclonalization is often considered critical is in that of human induced pluripotent stem cells (iPSCs). Because of their capacity for unlimited self-renewal and ability to differentiate via any lineage, this cell type offers immense promise for modelling disease states in vitro, enabling non-invasive genetic association studies, particularly as they relate to drug responses. Such efforts necessarily entail large, population-level cohorts. Cell-line derivation throughput is therefore the paramount limiting factor in unlocking the vast promise that iPSC technology holds in relation to fields such as functional genomics and precision medicine. The iPSC reprogramming process exerts a large amount of stress on cells, resulting in a population that is highly heterogeneous with regard to variables such as residual load of viral reprogramming vector and introduced chromosomal aberrations, eliciting the potential need to monoclonalize. As this step has historically incurred a critical bottleneck during automated and high-throughput derivation of iPSCs, we focus on this cell type as a case example for investigating monoclonalization methodologies.

Single-cell isolation is typically achieved via fluorescence-activated cell sorting (FACS), a form of flow cytometry. This process enables rapid sorting of individual cells, but there are a number of means by which it can result in undesirable outcomes. Sorted cells may not survive, leaving an empty well. Alternatively, faults in the sorting process may erroneously transfer more than one cell to the destination well, resulting in polyclonality. Furthermore, for any given cell type, there may be variety of morphological or physiological changes that can occur during development that alter the quality of the cell line. In the case of stem cells, for instance, there are a number of known morphological markers that indicate loss of pluripotency, a common defect in newly reprogrammed iPSCs. As a result of these factors, the presence, clonality and quality of cell aggregations in putatively monoclonalized wells must be validated post hoc.

At present, the only method for validating monoclonality is through manual inspection of microscopic imaging, which is performed at regular intervals to track the growth of colonies after sorting. Doing so is highly time-consuming, with technicians often spending several hours per day classifying wells according to colony presence, clonality and morphology. More critically, however, the reliance on human judgement introduces key sources of bias and technical variability, particularly when such protocols are distributed among multiple investigators and research groups. As a result of this lack of standardization, monoclonalization protocols cannot be reliably upscaled without exacerbating the technical variability of cell lines. All of these factors make monoclonalization a highly desirable target for automation, which would enable colony selection protocols to be infinitely expanded and distributed at scale, while minimizing technical variability.

Deep learning, based on the use of convolutional neural networks (CNNs), has enabled enormous advances in computer vision over the past several years and has become an invaluable tool in automating the analysis of biomedical images of various types.

The New York Stem Cell Foundation Research Institute, New York, NY, USA. ✉✉ e-mail: bfischbacher@nyscf.org; dpaull@nyscf.org
These techniques have already been applied to numerous processes in stem cell research, including for the automated inference of differentiation\textsuperscript{9,16} and prediction of function in iPSC-derived cell types\textsuperscript{17}. To our knowledge, however, CNNs have never been employed in automatically identifying clonality during monoclonalization protocols for any cell type.

In domain-specific tasks, deep learning models frequently match or surpass the image-analysing performance of human investigators\textsuperscript{9,16}. Dedicated neural network architectures exist for specific tasks such as image classification\textsuperscript{20} and segmentation\textsuperscript{21}. Specifically, detection networks, which are trained to detect and localize each instance of a given object class in images, clearly offer a promising opportunity for the automated verification of monoclonality, which ultimately relies on the counting of individual cells. Implementations of detection networks in other scientific endeavours have previously proven highly successful\textsuperscript{22}. These typically adhere to standardized procedures for training and inference, involving annotating images with object bounding boxes for training, followed by fitting the labelled data via defined network architectures such as R-CNN\textsuperscript{23} and YOLO\textsuperscript{4,14}.

A number of key nuances inherent to monoclonalization make the task resistant to automation through standardized, widely adopted deep learning practices. For example, confirming a monoclonal well requires the enumeration of individual starting cells. These typically occupy <0.01% of the well’s field of view and are too small to be visible to human investigators without manually magnifying the image at the precise location of the cell. Greyscale imaging exacerbates this difficulty, typically exhibiting a large amount of noise. Debris particles very often appear subjectively indistinguishable from starting cells, and investigators frequently rely on information in later images, such as growth, to confirm whether a specific particle is a cell or an abiotic artefact.

Irrespective of the above, verifying clonality necessarily depends on the interaction between images taken at different time points. For instance, enumerating individual cells in a day 0 image to validate that the sorting process was successful in isolating exactly one starting cell provides no information about the cell’s subsequent survival, expansion or retention of desirable morphological traits. Conversely, validating that only a single colony is visible at the time of inspection does not suffice to confirm monoclonality, given multiple starting cells may give rise to a single, polyclonal mass of cells that superficially resemble monoclonal colonies. In short, insofar as human investigators are able to assess, there are no cases in which a single image may contain all the information necessary to infer the clonality of a well. For this reason, it is not feasible to construct a conventional training set consisting simply of images and their corresponding semantic labels.

In this Article, we report an algorithm design that overcomes these difficulties by leveraging the chronological directionality inherent to the cell culturing process. Our computational workflow, termed Monoqlo, integrates multiple CNNs, each having its own ‘modular’ functionality. Monoqlo provides a highly scalable framework, capable of analysing datasets numbering in the tens of thousands of images in under an hour using commodity hardware. Through the combination of automated stem cell culture and deep learning, this work demonstrates an example of machine learning being applied to the identification of monoclonal cell lines from bright-field microscopy.

**Results**

**Neural network modularity.** We modularize the task of automatically assigning clonality into four distinct deep-learning-enabled functionalities (Fig. 1 and Extended Data Fig. 1). The decision to modularize was based on empirical inferences made during preliminary investigations. Namely, consistent with the principles of transfer learning\textsuperscript{25}, we initially suspected that a CNN’s feature-extracting capacity would be best optimized by consolidating all image types into a single training set. However, we found that networks trained in this manner performed poorly, often failing to distinguish between object classes. In particular, they often reported object types that could not feasibly occur in the image in question, for instance, detecting fully developed colonies in images generated immediately after seeding (Extended Data Fig. 2). This indicated that a single model would not perform well across the diversity of image magnifications and object classes employed during monoclonalization. So, we instead stratify our training set based on chronological timestamps as well as magnification and crop level, and train four separate neural networks, each having its own ‘modular’ functionality. First, we assign the term ‘global detection’ to the task of detecting the presence or absence of colonies in a full-well image. Second, we refer to the task of detecting colonies in cropped images of various well regions at a variety of zoom magnifications as ‘local detection’. Third, we term the task of enumerating individual cells in a fully magnified, cropped image ‘single-cell detection’. We sought

---

**Fig. 1 | Summary of the four CNN ‘modules’ used in Monoqlo.** **a**, Simple schematic representations of the two neural network architectures used for the tasks of detection and classification, adapted from ref. \textsuperscript{20}. FPN, feature pyramid network. FC indicates fully connected layers. **b**, The respective functionalities of each of the three detection modules with representative target data and outputs shown. Percentages in the top right of each image represent the magnitude of the zoom relative to a full-well image. **c**, Illustration of the use of four morphological classes in Monoqlo’s classification step. See Supplementary Fig. 1 for further information on the definition of each morphological class.
to achieve all three of the aforementioned tasks through the use of the RetinaNet detection architecture\(^2\). Finally, in the only entirely classification-based task in this effort, we aim to train a model to categorize images cropped around colony regions into morphological classes, here referred to as ‘morphological classification’ (as summarized in Extended Data Fig. 1). Modularizing in this manner enables us to capitalize on the temporal directionality of the cell culturing process, for example restricting detectable object classes to those that may realistically exist in an image based on its scan date.

**Workflow design overview.** We designed the Monoqlo computational workflow to integrate each of our trained neural networks. The laboratory automation workflow that generates data for use with Monoqlo, an illustrative example of a monoclonal colony growing in culture, and the design of Monoqlo itself are summarized in Figs. 2–4, respectively. Our algorithm processes images on a per-well basis in a reverse chronological fashion. That is, for each physical well, the algorithm begins by analysing the most recently generated scan. In our case, this is an image that has been cropped only to remove the black borders of the image, preserving the entire field of the physical well. These images are passed to the global detection model, the output of which is a coordinate vector demarcating the bounding boxes of any detected colonies. Our algorithm then expands these coordinates until each dimension of the bounding box is twice that of the predicted colony, loads the next most recent image for that well. Alternatively, if the workflow continues to detect exactly one colony until reaching the day 0 scan, the resulting image will be magnified and cropped exactly around the ancestral cell or cells. This image can then be passed to the single-cell detection model, providing a count of the number of starting cells. On this basis, the well may then finally be declared either monoclonal or polyclonal.

**Neural networks learn to detect colonies and classify morphology.** We benchmarked the efficacy of Monoqlo as a unified, modular workflow, first by testing its accuracy on a manually curated, class-balanced validation set, and subsequently by evaluating its clonality identification performance (irrespective of morphology) post hoc on a raw, unfiltered dataset from real-world monoclonalization runs. Our curated test set included 100 wells from each of three classes—empty, monoclonal and polyclonal—randomly selected from historical records of manually classified wells. The imaging date at which processing was initiated for each well was randomly generated from the range of days 8–18. Real-world scenario validation was performed on a monoclonalization run (DMR0001) that comprised 768 wells in total, spanning a time frame of 19 days, thus yielding a data volume of 18,240 images. A manual image review found 561 of these wells to be empty; that is, they contained no cell or cells. This image can then be passed to the single-cell detection model, providing a count of the number of starting cells. On this basis, the well may then finally be declared either monoclonal or polyclonal.

Deep-learning workflow with modularization identifies clonality. We benchmarked the efficacy of Monoqlo as a unified, modular workflow, first by testing its accuracy on a manually curated, class-balanced validation set, and subsequently by evaluating its clonality identification performance. We evaluated the performance of each CNN in its respective task by the Monoqlo-predicted bounding box X dimension (Pearson's \(r = 0.917\); \(P < 2.2 \times 10^{-16}\), Fisher transformation; Extended Data Fig. 3).
wells were reported as being polyclonal, of which 57 (93.4%) were confirmed by ground-truthing and four were found to be monoclonal. The results of both validations are summarized in Fig. 5.

Discussion
This work presents the automated identification of clonality using a deep learning object detection approach. We believe this has the potential to remove a critical restriction on scalability in a number of cell-culturing domains. This includes the present case of iPSC derivation, where monoclonalization is considered essential for two reasons. First, in cases of viral reprogramming, there is a large amount of cell-to-cell variance in the residual load of the Sendai viral vector used to deliver transcription factors to the inner cell during reprogramming. Second, the reprogramming process often leads to severe chromosomal abnormalities, presumably due to stress-induced mitotic disruptions. Both of these factors cause profound phenotypic variation, resulting in unpredictable, highly heterogeneous cell lines. Our team has previously suggested that the monoclonalization process could exert further physiological stress on cells, but single cell cloning remains critical in a number of cases. Given the extent to which cohort size dictates the viability of population studies, the removal of this bottleneck, as demonstrated in the present work, represents a significant step in fully unlocking the immense research potential of iPSCs.

Perhaps more importantly, in addition to initial derivation, huge efforts are being made towards optimizing CRISPR–Cas9 editing efficiency and other forms of genome engineering in iPSCs, which hold enormous potential for functionally annotating gene variants, disease modelling and validating polymorphisms identified in genetic association studies. Owing to the genomic heterogeneity the editing process introduces, newly edited populations must be monoclonalized to ensure that all cells carry the same genotype. This holds true for gene editing not only in iPSCs, but all cell types. We therefore view the genome engineering pipeline to be another critical case in which the Monoqlo framework alleviates a major bottleneck in disease research and therapeutic development.

We suspect that our algorithm could be adapted to any cell type, provided the cells are capable of being imaged and form discrete clonal masses. As an important example, antibody development is one of the most common use cases for monoclonalization due to the epitope specificity of monoclonal antibodies. Many of the most frequently used cell types in antibody development have been successfully detected in microscopy imaging with CNNs. As monoclonal antibodies form the central component of many drug discovery efforts, the Monoqlo framework may have the potential to offer a valuable tool to the pharmaceutical industry at large.

The present study adds to previous instances of deep learning applications in iPSC process automation. In particular, there is a
great deal of interest in optimizing CNNs for use with bright-field microscopy, because immunostaining and fluorescence microscopy imaging\(^1\) have much larger costs in terms of financial investment and investigation time. Christiansen et al.\(^4\) have successfully trained deep learning models to predict fluorescent labels from bright-field images alone. This work further demonstrates the predictive power

---

**Fig. 4 | Overview of Monoqlo’s design and algorithmic logic.** The arrows represent the processing order in the algorithm’s reverse chronological analysis, beginning with the most recent scan. If a colony is detected, the region around the colony is cropped in the previous day’s scan and the image is passed to the local detection model. The process is repeated, progressively reducing the field of view being analysed. If multiple colonies are detected in any scan, the well is declared polyclonal and no further scans are analysed. Upon reaching the earliest ‘day 0’ scan, the resulting image is passed to the local detection model. Based on the number of cells detected, a clonality for the well is finally declared. Solid and dotted lines indicate image processing and information processing steps, respectively.
of deep learning in various analysis tasks using simple microscopy images without the requirement of fluorescent labelling.

Waisman et al.\(^9\) showed that standard CNN architectures, such as ResNet50, may be trained to distinguish differentiated and undifferentiated stem cells in culture, even at early onset. Our classification CNN differs from theirs in that we stratify our training classes to a greater extent, as opposed to a binary ‘differentiated versus undifferentiated’ approach. Doing so served to increase the robustness of our algorithm when applied in real-world cell-culturing scenarios, in which there is a high degree of variability in iPSC colony morphology due to factors other than pluripotency status. Additionally, our network is trained on images cropped around distinct, singular colonies as opposed to field-of-view images containing numerous, randomly seeded cell aggregations. In this sense, our training data are more akin to those employed by Kavitha et al.\(^4\), in which a vector-based CNN is used to distinguish ‘healthy’ from ‘unhealthy’ colonies. Furthermore, they highlight the benefits of using segmented colony images and outline a number of key difficulties

---

**Fig. 5 | Results of Monoqlo framework validations.**

- **a.** Well-level clonality identification performance of the Monoqlo framework on real-world production run data. The outer colours represent the ground-truthed clonality of the well (colour meanings are indicated in the legend). The inner colours represent the clonality identified by Monoqlo, with dual-colour wells thus indicating Monoqlo errors.
- **b.** Class-specific clonality identification performance of Monoqlo on a manually curated, class-balanced test dataset.
- **c.** Summary of Monoqlo clonality performance, with analyses restricted to monoclonal, morphologically healthy wells that were selected for further passaging by biologists. Bars indicate the number of ground-truthed healthy, monoclonal wells that were predicted to belong to each class by Monoqlo.
involved in doing so in an automated capacity, without which classification models using segmented colony images cannot be integrated into real-world automated production workflows. By using our classification network in conjunction with colony detection models from the wider Monoqlo framework, we automate the segmentation step, enabling fully autonomous deployment in laboratory automation scenarios.

We recognize a number of limitations of our approach. For example, in cases where two or more starting cells are displayed precisely adjoining another one in the earliest available scan, the well's clonality status must be considered ambiguous. This is because it cannot be determined whether the cells were sorted independently from the source plate or if a single cell was successfully sorted in isolation and subsequently divided. Certain other efforts have attempted to address this ambiguity through fluorescence microscopy applied to nuclear-stained images, which allows nuclear segmentation and helps to resolve the spatial distribution of individual cells. However, this does not entirely eliminate ambiguity, because physically adjacent cells, even if clearly distinct, could certainly still have a monoclonal origin. We submit that there are a limited number of feasible approaches to handling this ambiguity. As a result, investigators may wish to simply assume that any well containing multiple cells at the time of earliest scan is polyclonal.

Although our algorithm represents the state of the art in automated clonality inference, our results indicate that its performance remains imperfect. Therefore, the exact manner of its application and the extent to which it can be relied upon as a fully autonomous monitoring system will depend on the aims and priorities of the process to which it is applied. If the acceptable margin of error for misclassifying polyclonal wells as monoclonal is 0%, for example, then post hoc validation by human investigators will be required. Such instances will nonetheless benefit from enormously reduced workloads for manual image review. If, alternatively, the paramount priority is that non-empty wells do not go undetected and small error rates in clonality assignment are deemed acceptable, then the critical variable will be the performance of the global detection model. In the present study, the 100% detection rate for colonies of sufficient size for passaging suggests Monoqlo's suitability for deployment as a dependable, fully autonomous system. In both of these cases, however, the defining variable will be the performance of the deep learning models in isolation. CNNs trained by different investigators and tailored to different cell/image types could show substantially inferior or superior performance to those described in the present study.

As for future directions, we believe Monoqlo could help facilitate investigations into a number of key questions that remain to be answered with regard to the predictive potential of deep neural networks in iPSC research. A number of studies have demonstrated that deep learning approaches can sometimes discriminate between biological groups in images where a morphological phenotype was not previously known to exist, or where it was suspected to exist but was not visible to even a trained human investigator. For example, Poplin et al. showed that CNNs can predict factors such as cardiovascular disease risk, gender and smoking status from individual retinal images, none of which was previously thought to manifest morphologically in the retina. Furthermore, in the case of iPSCs, deep neural networks have been successfully trained to predict donor identity from imaging of clinical-grade iPSC-derived retinal pigment epithelium. With these discoveries in mind, we suggest the likely existence of thus far unidentified predictive markers in iPSC colony morphology. For instance, it may be possible to predict with better-than-random accuracy at an early stage whether a presently undifferentiated colony will spontaneously differentiate. Successfully training such a model would confer an enormous benefit to iPSC derivation, given the substantial costs associated with continuing to culture cells that may ultimately become unusable. Other candidate targets for CNN classification- or regression-based prediction include Sendai virus load, future quality control pass/fail status and relative differentiation affinity for specific germ layers.

Training such models will invariably require large training volumes. The Monoqlo framework allows colonies to be algorithmically segmented and cropped from raw datasets, in addition to automatically filtering out images of empty wells, which typically represent the vast majority of images. In many cases, investigators may also be able to label images in batch on the basis of the classification they assign to the most recent image of a given colony or well. Applying our classification network, which identifies differentiation, allows Monoqlo to retroactively assign labels such as 'will differentiate' or 'won't differentiate' to earlier instantiations of the colony. This may mitigate the need for extensively laborious, manual reviews and labelling of unfiltered image sets, enabling partially or fully autonomous generation of large training volumes for future models. As such, our algorithm provides an invaluable tool for generating custom datasets for future investigations of the utility of deep learning in iPSC research.

In summary, we have demonstrated a framework in which deep learning algorithms with a modular design can automate the verification of monoclonality in bright-field microscopy, requiring relatively little labelling. We have further expanded the functionality of our workflow to the classification of colony morphology, demonstrating the potential for autonomous monitoring of monoclonal cell line development and clonal selection in automation workflows. Monoqlo represents a crucial step in enabling widespread distribution of high-throughput cell-line production and editing workflows. This may eliminate a critical bottleneck in the specific case of iPSC derivation and genome editing, moving current technology closer to the goal of unrestricted upscaling and distribution of pluripotent stem cells for biomedical research applications. Finally, in contrast to depending solely on machine learning models to contend with all aspects of a given task, we view this work as a useful example to highlight the benefit of combining the well-recognized, immense capabilities of CNNs with human-designed algorithmic solutions.

### Methods

All research detailed in this study complies with all relevant ethical regulations and our institute has received Federalwide Assurance 100 (FW A00021304) for the Protection of Human Subjects (FWA00021304). All cell derivation, handling and processing protocols are IRB-approved (WCG IRB), and consent was obtained from all human participants. For the purposes of this study, all cell lines were fully de-identified.

### Monoclonalization of hiPSCs

All cell lines used in the present study were obtained from the NYSRC repository (nysrc.org/repository) and were subjected to rigorous authentication and quality control protocols. Destinatin plates (PerkinElmer #6005182) were pre-coated with 17 μg of Geltrax lactate dehydrogenase-elevating virus-free (LDEV-free), hiESC-qualified, reduced growth factor basement membrane matrix (Thermo Fisher #A1413002) diluted in 50 μl DMEM/F12 (Thermo Fisher #A1413002) for 1 h in a 37 °C incubator. Following incubation, 150 μl of d0 medium, 1 μl of SYTOX AADvanced Ready Flow reagent (Molecular Probes #P3566) and 1.5× pluripotent stem cell Freedom supplement (Thermo Fisher #A25736SA) were added to the 50 μl of Geltrax + DMEM/F12 present in the well and incubated for 1 h in a 37 °C incubator. hiPSC colonies maintained on Geltrax in Freedom PSC medium (FBD; both Thermo Fisher) were dissociated with Accutase (Thermo Fisher #A110501) for 5–10 min at 37 °C. Accutase was quenched with sort buffer (magnetic-activated cell sorting buffer, Milltenyi, containing 10% CloneR) and the cell suspension was pelleted by centrifugation at 150 x g for 30 s.

For samples sorted using the WOLFSorter, the sort buffer was supplemented with SYTOX AADvanced Ready Flow reagent (Thermo Fisher #R37173) instead of PI.
Image acquisition and labelling. All images were sourced from repositories of historical data from the monoclonalization step employed during the iPSC production process of the NYSCF Global Stem Cell Array. These images, previously used for manually verifying clonality, are generated automatically once per 24-hour period from seeding until plates are discarded. All scans, which were generated by Nexcelom Celigo cytometers, are bright-field images acquired at a resolution of 1 μm per pixel, providing an image dimension of 7,544 × 7,544 pixels after stitching from 16 individual fields. We annotated a total of 3,139 images with bounding boxes and object classes. An additional 2,224 unannotated images of empty wells were included in the training set as background-only images. During preliminary investigations, we found doing so to be pivotal in reducing the rate of false detections. All annotations were generated in Pascal VOC format using the Labelling software (https://github.com/rtutalin/labelling). We augmented our dataset by applying random flip and rotation transforms to the images (as per Perez and Wang16), for enforcing the morphological criteria required for categorizing each object class were designated by PhD-level biologists specializing in iPSC culture. Annotations were made by technicians of PhD, MS and BS level, with all annotations being independently corroborated by an additional investigator. During initial investigations, we found that models sometimes incorrectly classified aggregations of dead cells, most of which had emanated from living colonies, as colony object detections. We corrected for this by labelling these objects as a separate class, as opposed to treating them as background. We termed this object type ‘overspill’ (Extended Data Fig. 6).

CNN architectures. For all object detection tasks in the present work, we use the RetinaNet CNN framework for one-shot detection, first introduced in ref. 49. The defining advancement proposed in this work was the use of the novel focal loss function, which adjusts the per-sample cross-entropy loss to more heavily penalize misclassifying difficult examples than easy examples, thereby resolving the issues imposed upon object detection tasks by class imbalance. Our final detection CNN architecture consists of a ResNet50 convolutional backbone with dynamic input dimensions, which performs feature extraction and passes the learned representations to a feature pyramid network. The RetinaNet50 architecture consists of 50 convolutional blocks, each consisting of a two-dimensional convolutional layer with rectified linear unit activation and batch normalization, and uses residual (skip) connections between several convolutional blocks. The overall network incorporates the use of anchor boxes that represent predefined candidate object locations, as described in ref. 49, imaging the object detection task to be trained in end-to-end fashion. The outputs from the feature pyramid network are passed to two submodels, one of which performs regression to refine the localization of the object bounding boxes, and the other performing object classification. Finally, the network's output layer produces a filtered vector denoting the posterior probability of each anchor box containing an object, including the object's class being indicated as a one-hot vector, as well as the refined pixel coordinates of the predicted bounding boxes. Overall, our detection networks each have more than 3.6 × 10^7 parameters. For morphological classification, we use the ResNet34 architecture. Our anchor box definitions had sizes 32, 64, 128, 256 and 512 pixels for each of the 15 levels of the feature pyramid, the 7 × 7 levels 3–7. For each feature map location, we define anchor boxes using ratios of 0.5, 1 and 2. For full details on the neural network architectures, see Supplementary Information.

Training of deep learning models. RetinaNet detection models were trained using a Keras RetinaNet implementation (https://github.com/lzyrx/keras-retinanet) with the ResNet50 convolutional backbone having weights pretrained on the ImageNet dataset. Preprocessing involved subtracting ImageNet means from images and normalizing the pixel intensity values to range between 0 and 1. We also implemented a hand-crafted algorithm for cropping the thick black borders around the well from the image; this removes the outermost line on each edge of the image and repeats until the maximum, raw pixel intensity value for the given line exceeds 70. Each CNN model was trained for 60 epochs, with weights being saved after each epoch, allowing the checkpoint with the smallest validation loss to be selected as the final model for use in the Monoqlo workflow. In each case, we used the Adam optimizer, selecting an initial learning rate of 10^-3, then reducing the learning rate by a factor of 10 every 10 epochs after epoch 40. These hyperparameters were selected based on learning trajectories from cursory training attempts after experimenting with multiple hyperparameter values. Training curves and a confusion matrix from the morphological classification model are presented in Extended Data Fig. 7.

Design of hand-crafted programmatic solutions. We identified a number of circumstances in which the out-of-the-box CNNS, while not otherwise appearing to have led to erroneous results, could be robustly corrected for using simple programmatic logic. Perhaps most prominently, we found that detection 

References
1. Kwakkenbos, M. J. et al. Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming. Nat. Med. 16, 123–128 (2010).
2. Wang, G. et al. Efficient, footprint-free human iPSC genome editing by consolidation of Cas9/CRISPR and piggyBac technologies. Nat. Protoc. 12, 88–103 (2017).
3. Vischer, P. M. et al. 10 years of GWAS discovery: biology, function and translation. Am. J. Human Genet. 101, 1–22 (2017).
4. Seki, T., Yuasa, S. & Fukuda, K. Generation of induced pluripotent stem cells from a small amount of human peripheral blood using a combination of activated T cells and Sendai virus. Nat. Protoc. 7, 718–728 (2012).
5. Chen, Y. H. & Pruett-Miller, S. M. Improving single-cell cloning workflow for gene editing in human pluripotent stem cells. Stem Cell Res. 31, 186–192 (2018).
6. Pauli, D. et al. Automated, high-throughput derivation, characterization and differentiation of induced pluripotent stem cells. Nat. Methods 12, 885–892 (2015).
7. Tsai, C.-C. et al. Screening method for rapid identification of hybridomas. US patent 9,797,838 (2017).
8. Ellis, J. et al. Alternative induced pluripotent stem cell characterization criteria for in vitro applications. Cell Stem Cell 4, 198–199 (2009).
9. Waisman, A. et al. Deep learning neural networks highly predict very early onset of pluripotent stem cell differentiation. Stem Cell Rep. 12, 845–859 (2019).
10. Kyttälä, A. et al. Genetic variability overrides the impact of parental cell type and determines iPSC differentiation potential. Stem Cell Rep. 6, 200–212 (2016).
11. Miller, J. D. et al. Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell Stem Cell* **13**, 691–705 (2013).
12. Križhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* 1097–1105 (NIPS, 2012).
13. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
14. Wainberg, M., Merico, D., Delong, A. & Frey, B. J. Deep learning in biomedicine. *Nat. Biotechnol.* **36**, 829–838 (2018).
15. Caiicedo, J. C., McQuin, C., Goodman, A., Singh, S. & Carpenter, A. E. Weakly supervised learning of single-cell feature embeddings. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition Pattern Recognition* 9309–9318 (IEEE, 2018).
16. Kusumoto, D. et al. Automated deep learning-based system to identify endothelial cells derived from induced pluripotent stem cells. *Stem Cell Rep.* **10**, 1687–1695 (2018).
17. Schaub, N. J. et al. Deep learning predicts function of live retinal pigment epithelium from quantitative microscopy. *J. Clin. Invest.* **130**, 1010–1023 (2019).
18. Esteva, A. et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).
19. Pereira, S., Pinto, A., Alves, V. & Silva, C. A. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **35**, 1240–1251 (2016).
20. Courdaly, N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018).
21. Havaei, M. et al. Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **35**, 18–31 (2017).
22. Caiicedo, J. C. et al. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nat. Methods* **16**, 1247–1253 (2019).
23. Girshick, R. East R-CNN. In *Proc. IEEE International Conference on Computer Vision* 1440–1448 (IEEE, 2015).
24. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: unified, real-time object detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 779–788 (IEEE, 2016).
25. Oquab, M., Bottou, L., Laptev, I. & Sivic, J. Learning and transferring mid-level image representations using convolutional neural networks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 1717–1724 (IEEE, 2014).
26. Lin, T. Y., Goyal, P., Girshick, R., He, K. & Dollar, P. 2017. Focal loss for dense object detection. In *Proc. IEEE International Conference on Computer Vision* 2980–2988 (IEEE, 2017).
27. Agu, C. A. et al. Successful generation of human induced pluripotent stem cell lines from blood samples held at room temperature for up to 48 hr. *Stem Cell Rep.* **5**, 660–671 (2015).
28. Garg, P. et al. Genome editing of induced pluripotent stem cells to decipher cardiac channelopathy variant. *J. Am. College Cardiol.* **72**, 62–71 (2018).
29. Rahman, S. H. et al. Rescue of DNA-PK signaling and T-cell differentiation by CRISPR-Cas9-induced precise gene editing in mammalian cells. *PLoS Genet.* **11**, e1005239 (2015).
30. Warren, C. R. et al. Induced pluripotent stem cell differentiation enables functional validation of GWAS variants in metabolic disease. *Cell Stem Cell* **20**, 547–557 (2017).
31. Wang, Y. et al. Genome editing of isogenic human induced pluripotent stem cells recapitulates long QT phenotype for drug testing. *J. Am. College Cardiol.* **64**, 451–459 (2014).
32. Chu, V. T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **33**, 543–548 (2015).
33. Smurnyy, Y. et al. DNA sequencing and CRISPR-Cas9 gene editing for target validation in mammalian cells. *Nat. Chem. Biol.* **10**, 623–625 (2014).
34. Khazaie, M. B., Conry, R. M. & LoBuglio, A. F. Human immune response to mononuclear cells. *J. Immunother.: Emphasis Tumor Immunol.* **15**, 42–50 (1994).
35. Vojtěšek, B., Bartek, J., Midgley, C. A. & Lane, D. P. An immunochemical approach for automatic recognition of colonies of induced pluripotent stem cells. *PLoS ONE* **12**, e0189974 (2017).
36. Christiansen, E. M. et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**, 792–803 (2018).
37. Kavitha, M. S. et al. Deep vector-based convolutional neural network approach for automatic recognition of colonies of induced pluripotent stem cells. *PLoS ONE* **12**, e0189974 (2017).
38. Poplin, R. et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat. Biomed. Eng.* **2**, 158–164 (2018).
39. Perez, L. & Wang, J. The effectiveness of data augmentation in image classification using deep learning. Preprint at https://arxiv.org/pdf/1712.04621.pdf (2017).
40. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
41. Lin, T. Y. et al. Feature pyramid networks for object detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 2117–2125 (IEEE, 2017).
42. Dahl, G. E., Sainath, T. N. & Hinton, G. E. Improving deep neural networks for LVC/SR using rectified linear units and dropout. In *Proc. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* 8609–8613 (IEEE, 2013).
43. Joffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Preprint at https://arxiv.org/pdf/1502.03167.pdf (2015).
44. Ren, S., He, K., Girshick, R. & Sun, J. Faster R-CNN: towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems* 91–99 (NIPS, 2015).
45. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/pdf/1412.6980.pdf (2014).
46. Jeong, H. J., Park, K. S. & Ha, Y. G. Image preprocessing for efficient training of YOLO deep learning networks. In *Proc. 2018 IEEE International Conference on Big Data and Smart Computing* (BigComp) 635–637 (IEEE, 2018).

**Acknowledgements**

This work was supported by The New York Stem Cell Foundation (NYSCF). We thank the members of the NYSCF leadership team, specifically R. Monsma, S. Noggle, R. Aiyar, C. Anzel, L. Schwarzbach, J. Wallerstein and S. Solomon, for their support throughout this work. We also thank L. Mehran and M. Berliss for their guidance on reporting biological research protocols. We thank C. Richardson for his hugely helpful guidance on the release of Menonlo.

**Author contributions**

B.F. and Z.W. conceptualized the Menonlo framework, including the use of reverse chronological analysis for the assessment of coloniality. B.F. trained and validated RetinaNet detection models and wrote the Python software for the execution and automated deployment of Menonlo, including data-handling logic, image processing and integration of deep learning models. B.F. conceptualized the use of classification networks in automatically assigning morphological classifications to the most recent colony images. B.F., S.H., C.P. and J.B. conceptualized the labeling system for classifications of colony morphology. S.H. labelled training data and trained and validated all morphology classification models. B.F., S.H., C.P. and J.B. conceptualized the labeling system for classifications of colony morphology. S.H. labelled training data and trained and validated morphology classification models. G.L. and D.P. developed NYSCF’s iPSC monolayer-plating laboratory-automation and colony-selection protocols. B.F., B.H., J.B. and D.P. and NYSCF Global Stem Cell Array Team performed image annotations for training the RetinaNet models. D.H., B.H., M.Z., J.B. and NYSCF Global Stem Cell Array Team performed manual monolocalizations, validation of the Menonlo framework and subsequent cell culture and imaging using robotic systems.

**Competing interests**

B.E. and Z.W. are co-inventors on a pending patent regarding an image system and method of use (pub. no. WO2021067977/1A1). The authors declare no other competing interests.

**Additional information**

Extended data is available for this paper at https://doi.org/10.1038/s42256-021-00354-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42256-021-00354-7.

Correspondence and requests for materials should be addressed to B.E. or D.P.

Peer review information *Nature Machine Intelligence* thanks Santiago Mirniku, Lassi Paavolainen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021
Extended Data Fig. 1 | Examples of each morphological class used in training Monoqlo’s classification CNN module. M1 is the desired morphology, indicating a healthy, pluripotent stem cell colony, and is defined as having a clearly defined, tight perimeter, round shape, no evidence of differentiation and a core with a smooth, transparent appearance. M2 is defined as a colony with the morphology of M1 but with a differentiated fringe. In the displayed example, differentiation and thus loss of pluripotency is clearly shown by the spindle-shaped cell formations and round core with a dark coloration in the bottom left of the tile. M3 is defined as a colony with a poorly-defined shape and often a predominantly dark coloration, which can indicate either differentiation or a dense aggregation of dead cells. M4 is a fully differentiated colony, composed entirely of sprawling, spindle-shaped cell aggregations, and displaying none of the desired morphological markers of pluripotency or iPSC health status.
Extended Data Fig. 2 | Example of poor performance by a generalized model trained across all functionalities. In this instance, the colony detection is correct. However, the cell detection, in addition to being incorrect, is impossible at the given image magnification and time point.
Extended Data Fig. 3 | Predicted Colony Width vs Ground Truth. Relationship between width of colony bounding box predicted by Monqlo's global detection model and the true width measured by biologists with a scale bar image overlay, plotted using 268 measurement-prediction pairs.
Extended Data Fig. 4 | Example of abiotic artifacts causing false colony detections by Monoqlo’s global detection model. a) and b) represent the same image report by Monoqlo, full view and zoomed, respectively.
Extended Data Fig. 5 | Example gating strategy. Representative gating strategy employed during FACS-sort monoclonalization of iPSCs.
Extended Data Fig. 6 | Overspill labelling example. Labelling example in which an additional object class, ‘overspill’ (indicated by blue bounding boxes,) is annotated to improve model performance and mitigate erroneous detections of the ‘colony’ (green bounding box) object class.
Extended Data Fig. 7 | Model training and selection. a, Training and validation accuracy trajectories of the classification CNN, plotted against epoch. Red and green dots signify training and validation accuracies, respectively. b, Confusion matrix of fully trained classification CNN when validated on held-out test set. Scale bar indicates color shading key, indicating the number of examples classified for respective classes as a proportion of total number of examples for the given class. c, Example training and validation accuracy over train time of the RetinaNet detection CNN.
Extended Data Fig. 8 | Overlapping detections. Example of overlapping reports of colonies by Monoqlo’s local detection model where only a single colony exists after ground-truthing.
Extended Data Fig. 9 | Colony splitting example. Illustration of the concept of “colony splitting”, where an apparent single colony is revealed, during reverse-chronological analysis, to have originated in multiple colonies which ultimately merged.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a □ Confirmed

□ □ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

□ □ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

□ □ The statistical test(s) used AND whether they are one- or two-sided

Only common tests should be described solely by name; describe more complex techniques in the Methods section.

□ □ A description of all covariates tested

□ □ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

□ □ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

□ □ For null hypothesis testing, the test statistic (e.g. f, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

Give P values as exact values whenever suitable.

□ □ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

□ □ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

□ □ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection  Nexcelom Bioscience Satellite, version 5.2.0.0; labelling, version 1.4.4

Data analysis  Python version 3.6.9; NVIDIA CUDA toolkit version 10.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All of the data used in training models for and validating the Monoqio framework, as described in the manuscript text, are readily available upon request. All images will be supplied de-identified with regard to cell line donor and NYSCF internal logistical identifiers such as run number or project code.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- [ ] Life sciences
- [ ] Behavioural & social sciences
- [ ] Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
- Sample size (number of wells/images) was determined by data availability at time of analysis

Data exclusions
- No data were excluded from analyses

Replication
- Monooqlo is constructed using publicly available software, and the code is made available to those wishing to validate the algorithm. The framework is described in detail, in order to allow the same approach to be applied to other data sets than those used in this study, as trained machine learning models are very sensitive to subtle batch effects, such as may be present when using a different cell type, camera or being investigated by a different research group.

Randomization
- During validation on the manually curated validation set, the wells which were selected for analysis were selected using random indexing (Python’s “Random” library) from all wells which had previously been labeled as either empty, monoclonal or polyclonal and which had not been employed during training or train-time validation of machine learning models.

Blinding
- Investigators performing manual inspection of the validation run were blinded to the results of Monooqlo’s analysis of the same run.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a                             | n/a     |
| ☒ Antibodies                    | ☒ Involved in the study |
| ☒ Eukaryotic cell lines         | ☒ ChIP-seq |
| ☒ Palaeontology and archaeology | ☒ Flow cytometry |
| ☒ Animals and other organisms   | ☒ MRI-based neuroimaging |
| ☒ Human research participants   |         |
| ☒ Clinical data                 |         |
| ☒ Dual use research of concern  |         |

Antibodies

Antibodies used
- Used during FACS sorting - SSEA4-647: 1:100 BD #560219; Tra-1-60-488: 1:100 BD #560173; CD56-V450: 1:100 BD #560360; CD13-PE: 1:100 BD #555394

Validation
- Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer’s website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.

Eukaryotic cell lines

Policy information about: cell lines

Cell line source(s)
- All iPSC lines used were reprogrammed from fibroblasts or peripheral blood mononuclear cells obtained with participant consent under an IRB-approved protocol.

Authentication
- All cell lines were authenticated using the Fluidigm SNP Trace Assay which confirms their identity with their putative donors

Mycoplasma contamination
- All cell lines tested negative for mycoplasma contamination

Commonly misidentified lines
- No commonly misidentified lines were used in this work

(See ICLAC register)
Flow Cytometry

Plots

- The axis labels state the marker and fluorochrome used (e.g. CD4-PerCP).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation: Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument: FACSARIA-iiL™ Cell Sorter (BD Biosciences), WOLFSorter

Software: n/a

Cell population abundance: hiPSC colonies maintained on Geltrex in Freedom PSC media (FRD1) (both ThermoFisher) were dissociated with Accutase (ThermoFisher #A1110501) for 5-10 min at 37°C. Accutase was quenched with Sort buffer (MACS Buffer Miltenyi, containing 10% CloneR) and the cell suspension pelleted by centrifugation at 130 RCF. Cells were stained with antibodies: SSEA4-647: 1:100; BD #560219, Tra-1-60-488: 1:100; BD #560173, CD56-V450: 1:100; BD #560360, CD13-PE: 1:100; BD #555394 before being rinsed with a second centrifugation and resuspended in Sort Buffer + Propidium Iodide (PI), 1:5000, Thermofisher #P9560).

Gating strategy: Gating strategy summarized in Supplementary Figure 6

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.