Mitochondrial Phenotypes Distinguish Pathogenic MFN2 Mutations by Pooled Functional Genomics Screen

Authors
Alex L. Yenkin, John C. Bramley, Colin L. Kremitzki, Jason E. Waligorski, Mariel J. Liebeskind, Xinyuan E. Xu, Maria A. Vakaki, Vinay D. Chandrasekaran, Robi D. Mitra, Jeffrey D. Milbrandt, William J. Buchser

Affiliations
Department of Genetics, Washington University School of Medicine, St Louis MO 63110
Functional Imaging for Variant Elucidation at the McDonnell Genome Institute

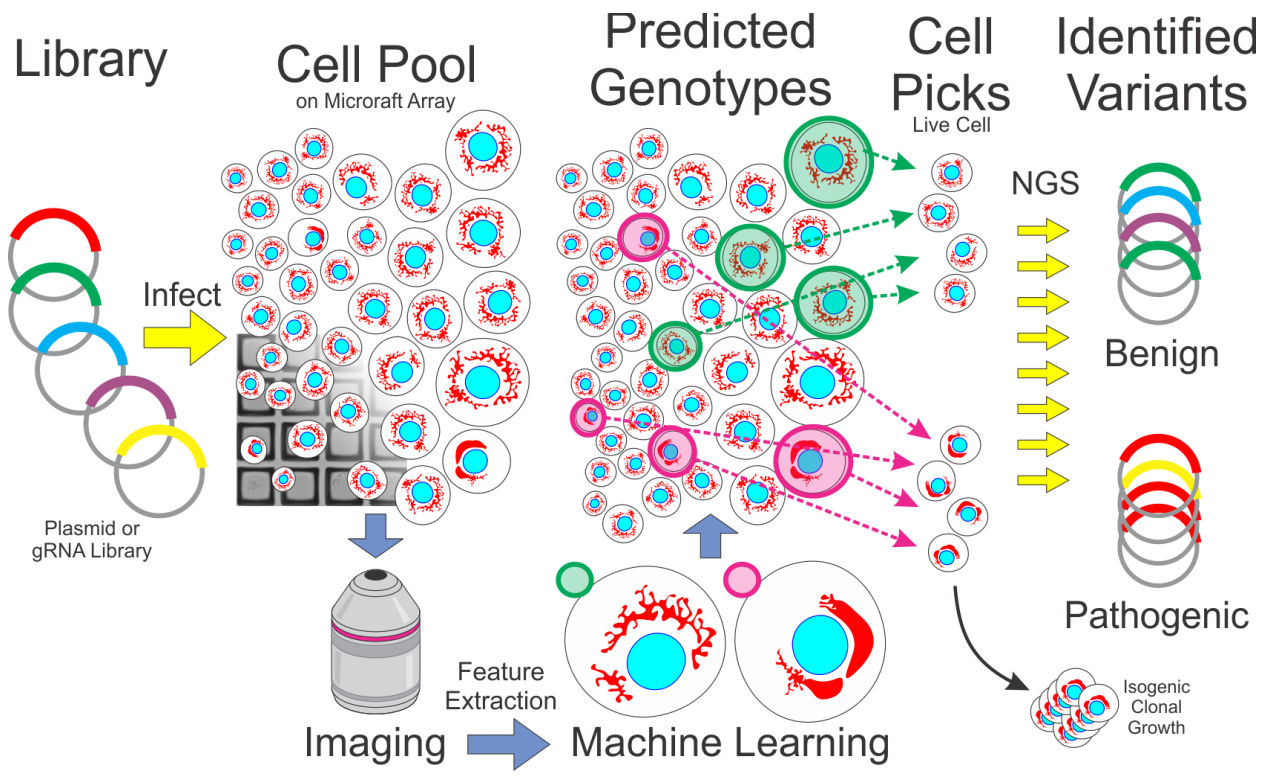
William J Buchser (Corresponding)
Department of Genetics, Washington University School of Medicine
4515 McKinley Ave
St Louis, MO USA
Phone 314-362-0026
wbuchser@wustl.edu

Running Title (Head)
MFN2 Variants Imaged with Raft-Seq

Definitions
MAVE: multiplexed assays of variant effects. DMS: deep mutational scanning. VUS: Variant of Uncertain Significance. Genotype: The specific point mutation present (R280H, WT, D221, etc). Genotype class: WT, Benign, or Pathogenic (aka what is designated by the model). Labeled Cell Population: Used for training the models. Unlabeled Cell Population: admixtures of labeled cells (like benign + pathogenic) or in screening, where the researchers are completely blind to the identity of individual cells. ROC: Receiver Operator Curve. AUC: Area under the curve.

Abstract

Most human genetic variation is classified as VUS - variants of uncertain significance. While advances in genome editing have allowed innovation in pooled screening platforms, many screens deal with relatively simple readouts (viability, fluorescence) and cannot identify the complex cellular phenotypes that underlie most human diseases. In this paper, we present a generalizable *functional genomics* platform that combines high-content imaging, machine learning, and microraft isolation in a new method termed “Raft-Seq”. We highlight the efficacy of our platform by showing its ability to distinguish pathogenic point mutations of the mitochondrial regulator *MFN2*, even when the cellular phenotype is subtle. We also show that our platform achieves its efficacy using multiple cellular features, which can be configured on-the-fly. Raft-Seq enables a new way to perform pooled screening on sets of mutations in biologically relevant cells, with the ability to physically capture any cell with a perturbed phenotype and expand it clonally, directly from the primary screen.



Graphical Abstract. Here, we address the need to evaluate the impact of a multitude of genetic variants. This manuscript depicts the methods of using machine learning on a biologically relevant phenotype to predict specific point mutations, followed by physically capturing those mutated cells.

Background

The explosion of functional genomics in the past decade [1] has enabled a massive shift in the study of the underlying genetics of human pathology. Even so, specific genetic mutations are still difficult to connect with disrupted cellular phenotypes, necessitating a detailed phenotyping-based *functional genomics* platform that can screen large numbers of genetic perturbations/variants—so called multiplexed assays of variant effects MAVE [2]—and work alongside new methods of generating variant libraries, such as deep mutational scanning DMS [3].

Genetic perturbation screens have generally relied on simple phenotypes conducive to enrichment analysis [4], Fluorescence-Assisted Cell Sorting [5–7], or phenotypes measurable by sequencing [8,9]. For more complex phenotyping, high-content imaging/screening (HCI/HCS) [10–14] is performed in an arrayed format for genetics perturbation screens [15–20], but these are difficult to scale and infeasible for combinatorial screens. There are platforms that use an imaging-based approach in pooled genetic perturbation screens, and have been demonstrated on relatively simple phenotypes [21–23] or on precise phenotypes that were known in advance [24,25].

Although some platforms do isolate individual cells [26], most genetic perturbation screens use a population-level measurement as their endpoint. The main challenge for such screens that operate on the single-cell level is mapping the perturbation back to each cell post sequencing. For screens where the phenotype is measurable through sequencing, the perturbation can be found simultaneously with the phenotype [8]. In imaging, a cell's position within a microscopic field of view is the unique piece of identifying information used to map phenotype to perturbation. Some platforms use *in*

situ sequencing [27,28] to generate sequencing results that contain positional data to a specific cell [24,29]. However, a platform that can be integrated into existing next-generation sequencing (NGS) pipelines would be more accessible and robust. Other platforms use a digital micromirror device to photoactivate endogenous fluorophores in specific cells which are then put through FACS-seq [22,25,30].

In this paper, we present Raft-Seq, a pooled screening platform that predicts individual cell perturbations from high-content imaging and machine learning. Raft-Seq improves on other platforms in several important ways: 1, it uses a microraft array plate [31–33] for context-aware isolation of identified cells; 2, it can use vital dyes or stains, so no genetic modification of the cell is required other than the perturbation itself; 3, it is largely phenotype-agnostic, needing only knowledge of the broad physiology beforehand for stain selection and initial feature filtering; 4, it uses machine learning to identify perturbed cells, which allows for the identification of complex cell-autonomous phenotypes; 5, the cells it selects have high viability and clonability, directly from the primary screen.

Here, we use the Raft-Seq platform to examine the neurologically relevant mutations in the *MFN2* gene, which protects against cellular stress from damaged mitochondria by regulating mitochondrial fission and fusion [34]. Clinically relevant *MFN2* mutations (pathogenic variants) primarily result in Charcot-Marie-Tooth Disease (CMT), the most common inherited neuromuscular disorder characterized by peripheral neuropathy with impairment of the central nervous system [35,36]. We find that the phenotype caused by pathogenic *MFN2* variants is distinct, but the difference is not adequately described by a single measurement/feature, necessitating a more complex

feature analysis. Findings from the single perturbation experiments allowed the implementation of scanning mutagenesis with gRNAs across the *MFN2* coding region.

Results

MFN2 Mutants

We generated cell lines that each contain *MFN2* cDNA with single point mutations through lentiviral infection. The mutants studied consisted of six clinically relevant pathogenic *MFN2* variants [37] and five control variants annotated as “benign” from ClinVar [38] (**Figure 1a**). Twelve lentiviral transfer plasmids—one plasmid for each mutated cDNA, and one for wild type cDNA—were used to package lentivirus that was infected into U2OS to create twelve cell lines.

To investigate the phenotypes caused by the *MFN2* variants, we stained each cell to visualize mitochondria and mitochondrial membrane potential using MitoTracker and TMRM, respectively [39]. The cells were imaged using a high-throughput fluorescent confocal microscope and intracellular features measured using Cytiva’s IN Carta analysis software. A visual comparison of the *MFN2* cell lines demonstrates subtle phenotypic differences among them (**Figure 1b**), specifically, perinuclear aggregation of mitochondria in the cells containing the pathogenic mutations. Box plots for a subset of the features are shown in **Figure 1c.** The mean mitochondrial puncta count in the wild-type and benign cell lines is consistent, while the values for the pathogenic cell lines were lower and more variable. A higher average variation in TMRM intensity within a single cell was also observed in the pathogenic mutants, all suggesting mitochondrial aggregation and disrupted mitochondrial membrane potential**.** Additional feature

histograms are shown in **Figure S1**. Although there are significant differences in the population of mutants as a whole, there is no single feature that can separate more than a small fraction of the cells.

We next tested whether multiple features could distinguish the pathogenic-variant cells. UMAP dimensionality reduction was performed on the full set of cellular imaged-based features (**Figure 1d**). Most pathogenic mutant cells are separable from wild type/benign cells in the dimensionality reduction, though some mutants (L76P and R94Q) are more distinct (**Figure S2**), which is consistent with the individual feature data. The UMAP reduction shows that using data from multiple features may more reliably distinguish cells with pathogenic mutants. Thus, to screen individual cells based on subtle phenotypes caused by these pathogenic mutations, we designed an HCS platform with machine learning capabilities that can use information from multiple image features.

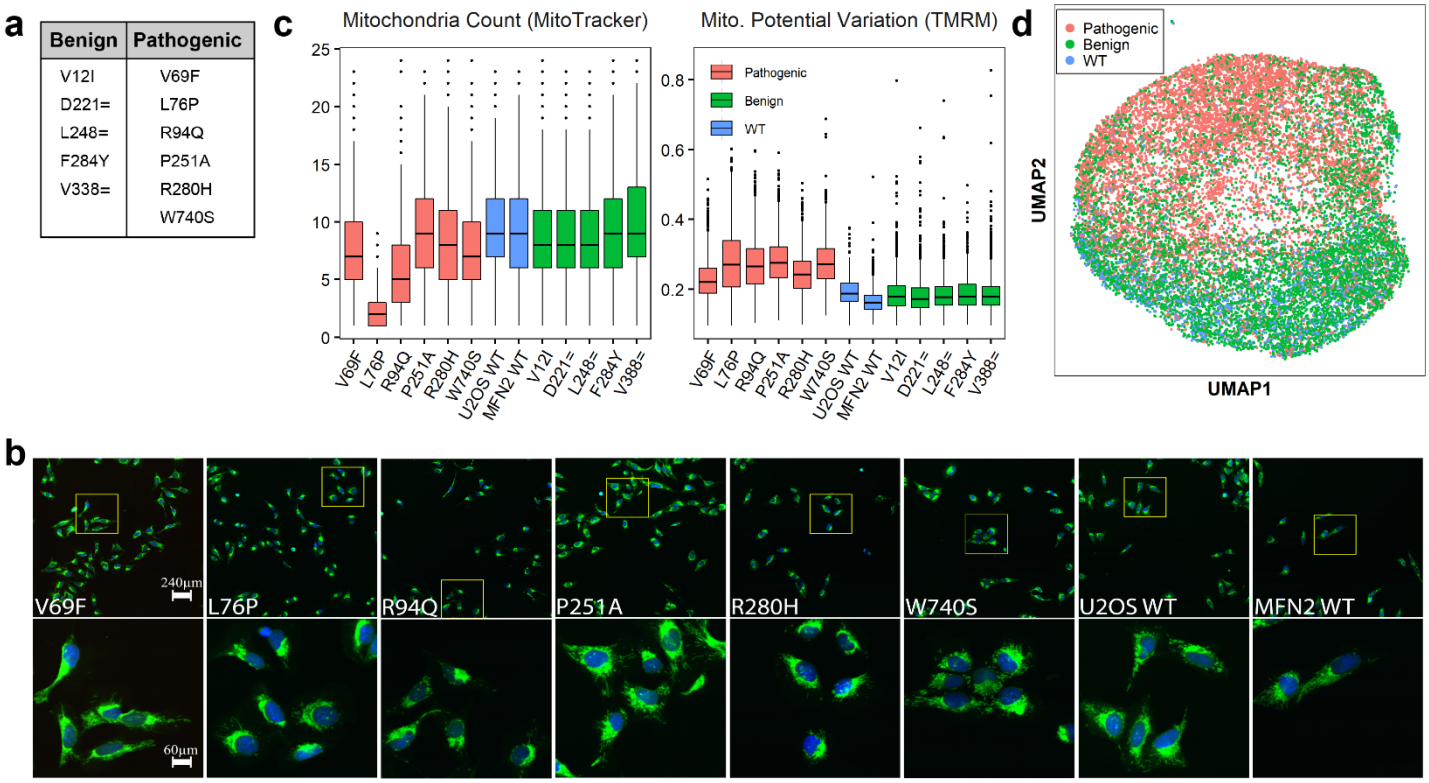


Figure 1. Isogenic *MFN2* Mutant Cell Lines Characterization: a. Table of the mutants used. **b.** Confocal images of mutant cells at two magnifications (blue = nuclear staining by Hoechst, green = mitochondria staining by MitoTracker). Subtle mitochondrial differences can be observed amongst the mutant populations. Known pathogenic mutants display peri-nuclear aggregation of mitochondria and a lack of mitochondrial spreading. In comparison, wild-type cells show ample mitochondrial spreading. **c.** Box plots of two features: Mitochondrial count (based on MitoTracker staining) and coefficient of variation in mitochondrial potential (TMRM intensity). Histograms are colored by genotype class (Benign/Pathogenic/WT). **d.** Scatter plot of UMAP reduction using a set of 26 features measured with TMRM or MitoTracker staining. Each dot represents a single cell, colored by its genotype class (WT/benign/pathogenic).

Experimental Workflow

The overall workflow of Raft-Seq can be summarized in five steps: perturbation, imaging, model building, isolation, and sequencing (**Figure 2a**). In the first step, we introduce a genetic perturbation, either endogenous (e.g., CRISPR gRNA library) or exogenous (e.g., plasmid overexpression library with mutagenesis). The cells are then plated onto a microraft array, stained, and imaged on a high-throughput confocal

microscope. From the cell-feature matrix obtained from feature extraction, we generate many supervised learning models, used to predict whether there is a discernible perturbation in each cell. Several different model types are employed (**Methods: Machine Learning and Model Generation**), which are then evaluated and deployed to determine cells to be isolated. Isolation is performed using the Cell MicroSystems Air Instrument, where rafts containing cells of interest are automatically transferred with a magnet to a well of a 96-well plate. The cells on the isolated rafts are then genotyped to determine the perturbation present. The genotyping pipeline is performed on single cells, not pooled genetic material.

As a preliminary test of raft-based imaging and isolation, we performed Raft-Seq on cells with dual genome-encoded RFP and GFP that had been given a gRNA to knock-out either RFP or GFP. These cells were plated on a microraft plate as a mixture and were imaged. Cells that were given the guide targeting RFP would be expected to appear green, since GFP would continue to be expressed and vice-versa. Using fluorescent marker intensity, we predicted the guide presence, isolated those cells, performed single-cell DNA sequencing to look for the edited genomic targets, and determined the true genotype. The correspondence between the predicted guide and the true guide was near perfect (Accuracy = 98.8%, n = 162) and comparable to a flow-based single-cell sequencing method run in parallel (**Figure 1b** and **Supplemental Methods**).

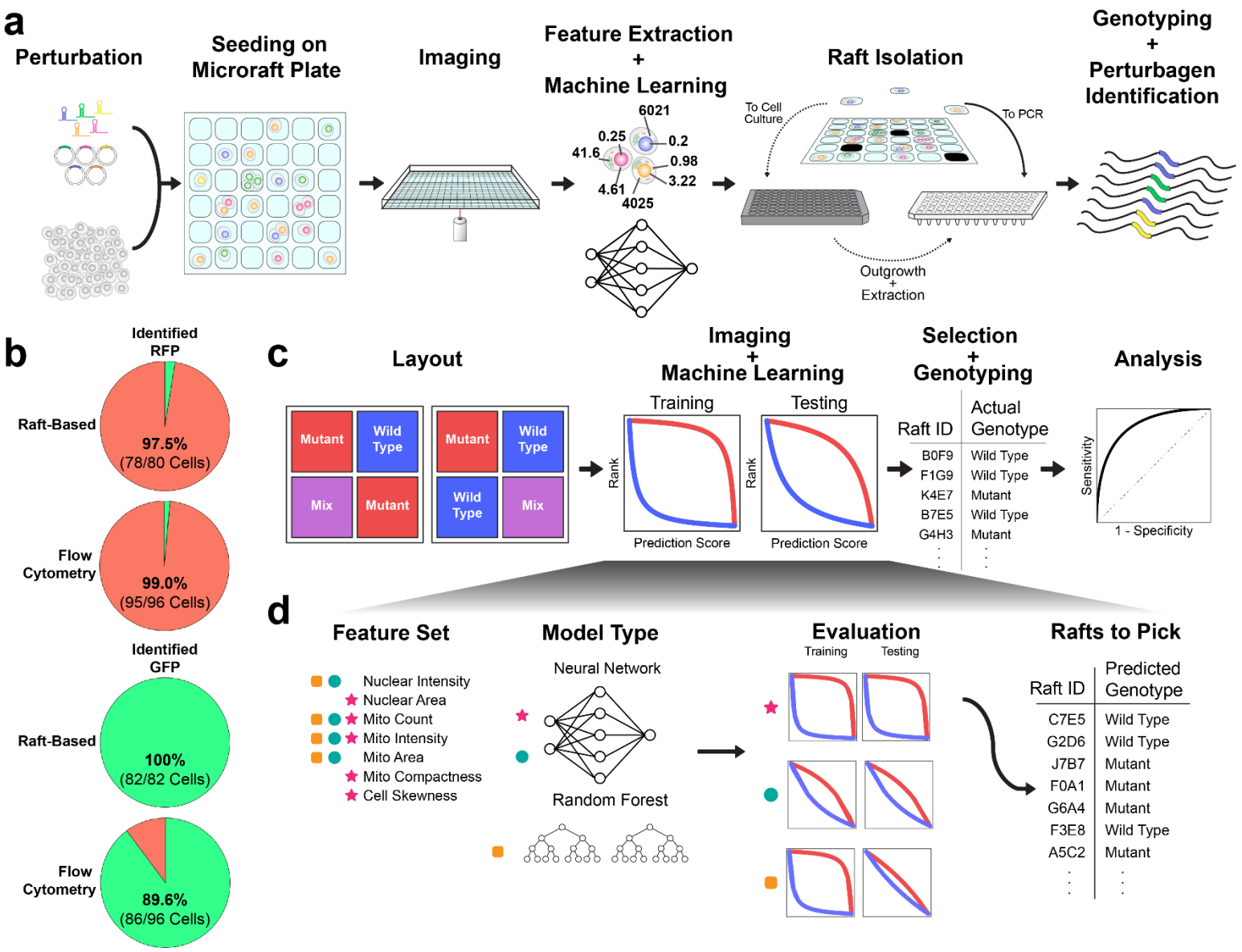


Figure 2. Raft-Seq Experimental Workflow a. A perturbagen library, primarily a lentiviral gRNA or plasmid overexpression library, is introduced to cells, which are then seeded onto a microraft plate. The plate is imaged, and cell feature data are then extracted from the resulting images and used to build machine learning models. The trained model then selects rafts to either be isolated into a PCR plate for immediate analysis or a tissue culture plate for clonal expansion. Following isolation, the cells are genotyped. **b.** Pie charts showing the composition of cells identified as either expressing GFP or RFP, after being separated by either the raft-based approach or flow cytometry. The color represents the expected appearance of the cells based on genotype (green = contains RFP gRNA, red = contains GFP gRNA), and the fractions/percentages are the amount in each group that is correctly identified. **c.** Performance evaluation workflow. The blue and red wells represent the labeled cell populations and the purple well represents the unlabeled population that is a mixture between the two. Following imaging, models are trained and tested on the labeled wells, and the one with the clearest

delineation between the two classes is selected. That model is used to select rafts that are then isolated and single cell genotyped. The comparison between the prediction scores and the true genotype class generates a receiver operator characteristic curve to evaluate model efficacy. **d.** Modeling workflow. Each model comprises a feature set and machine learning algorithm, combinations of which are represented here with different symbols. Each model is evaluated with training and testing cell data, and the model that shows the largest separation between classes and is robust on the testing data is used to generate a list of cells (rafts) to pick, along with their predicted genotype.

Identifying Subtle Mitochondrial Phenotypes in a Mixed-Variant Pool

To validate the efficacy of Raft-Seq in a more complex screen, we attempted to separate a mixture of wild-type and mutant cells by reproducibly predicting a given cell's *genotype* based on its *phenotype*, as determined by a set of features extracted from imaging data (with no fluorescent reporters). For this experiment, we separated cells containing pathogenic *MFN2* mutations from those containing the wild-type *MFN2* cDNA.

The workflow of this validation experiment is shown in **Figure 2c**. The goal was to see if we could identify any of four pathogenic mutants (L76P, R94Q, P251A, R280H) from WT in an admixture of those cells. In the different wells of the microraft plate, we plated either 1) *MFN2* WT cells, 2) a mixture of the 4 *MFN2* mutants (1:1:1:1), 3) a mixture of the wild-type and pathogenic cells at a wild-type:pathogenic ratio of 90:10, and 4) a similar mixture at a ratio of 50:50. We distinguished between the first two groups and the last two by calling the former (1,2) "labeled" and the latter (3,4) "unlabeled" to clarify for which cell populations the genotype class was known prior to imaging and analysis. Following application of a nuclear stain, a plasma membrane stain, and a mitochondrial stain (**Methods: Staining**), the cells were imaged and their

features computed. Next, we designed classification models that best distinguished the labeled populations (*MFN2* WT vs. *MFN2* mutants). General modeling workflow is shown in **Figure 2d**. We used a variety of model algorithms (types) and cell features sets, preferentially choosing features that correlated with genotype class. For model selection, we trained each model on the fly using plate-spanning sets of feature data. For each set of predictions for each model, we graphed the ranked prediction score (**Figure 2d: Evaluation**). If a model appropriately distinguished between the labeled populations, it would assign a score near 1 to the pathogenic mutant cells and a score near 0 to the wild-type cells and the ranked prediction score curves will be widely separated. We then selected the model with the largest separation between the labeled populations in both the training and testing wells (lack of separation in the testing wells indicated overfitting or sensitivity to batch effects) (**Figure 2d**).

Using the selected model, we chose cells for isolation from the unlabeled wells with additional control cells picked from the labeled wells. Over 1000 cells were then individually isolated into wells of several 96-well PCR plates (384 labeled, 758 unlabeled). From there, the cells underwent amplicon library construction, multiplexing, and NGS, from which their genotypes were ascertained.

With the knowledge of each cell's genotype (n_{WT}=159, n_{L76P}=61, n_{R94Q}=74, n_{P251A}=41, n_{R280H}=6), we found the total accuracy of our on-the-fly predictions to be 72.4% (50:50 =75.7%, and 10:90= 64.5%). Therefore, when training with only WT vs. a mixture of *MFN2* mutants, we could predict that individual cells were mutant from an admixture in which we were completely blind to the real genotype using only the subtle mitochondrial phenotype. Confusion bar charts are presented in **Figure 3a.** As

expected, when trying to identify mutants that were the vast minority (in the 10:90 admixture), the model does correctly identify pathogenic mutant cells, but the type II error is large. **Figure 3b** shows further breakdown of all unlabeled wells by specific mutant and by whether the model correctly identified it as a pathogenic mutant (per-mutant metrics are not available in this experiment since we did not train on individual classes of mutants). The mutations most successfully recovered were R94Q and L76P, agreeing with the previous data showing those mutants as having a more severe phenotype. We recovered a similar number of false negatives of the P251A and R280H mutants, despite fewer overall numbers of each mutant, indicating that these mutants likely have a weaker phenotype—leading to a less confident prediction and exclusion from the list of cells to isolate—rather than less penetrance. To check the model quality, we evaluated the Receiver Operating Characteristic (ROC) curves and the resulting area under the curve (AUC) **Figure 3c**. The curves in red show the ROC of our model—for the cells picked from labeled (AUC = 0.94) and unlabeled (AUC = 0.74) populations.

We also took a larger view and looked at the quality of *all* models generated, not just the one used to select cells for isolation. **Figure 3d** shows a scatterplot of AUCs (labeled populations vs. unlabeled populations) and **Figure 3e** shows histograms of all models' unlabeled populations AUCs. Importantly, the majority of models had discriminatory ability. This entire experiment was repeated successfully using a mixture of all six mutants, and histograms and scatterplots of the models generated are shown in **Figure S3.**

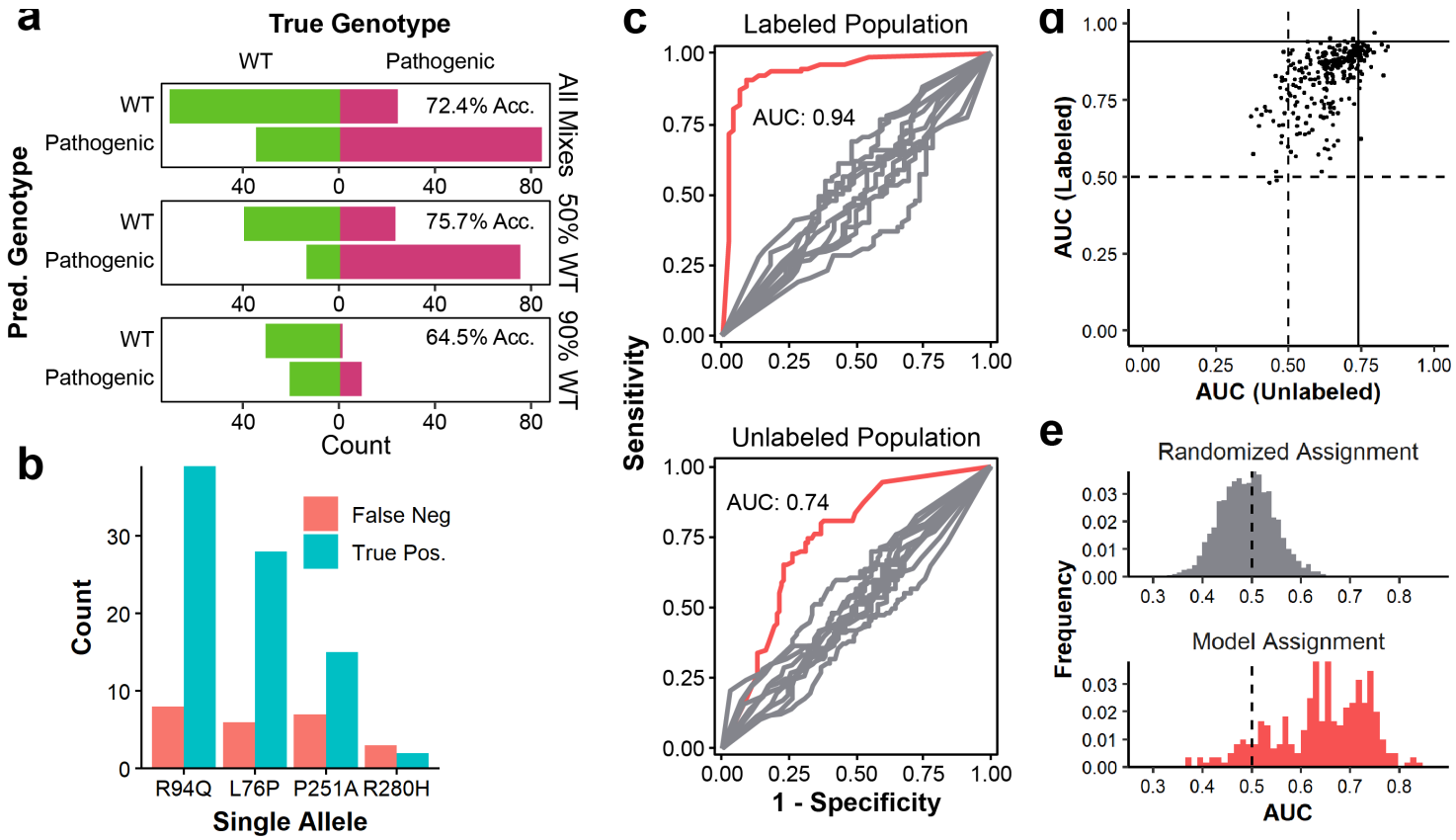


Figure 3. Raft-Seq Proof-of-Principle Results against MFN2 Pathogenic Mutations. a. Confusion bar charts of predicted genotype vs. true genotype for the main model from cells in the WT/Pathogenic mixture. The top bar chart represents all cells isolated from the mixture, while the bottom two are a breakdown of the first according to whether the cells were isolated from the well containing 50% or 90% wild-type cells. **b.** A bar chart of picked mutants separated by allele and counting the final status of individual cells. **c.** ROC Curves generated using the best model identified *a posteriori* separated by data generated from cells picked from control pure wells (upper) and from mix wells (lower). The red curves are the experimental results and the grey curves are a control generated by random shuffling of labels. **d.** A scatterplot showing the performance of all 290 models in detecting mutants in a mixture of wild type cells and four pathogenic mutants. Each point represents a single model and its position is determined by its ability to distinguish cells in the labeled control wells and cells in the mixed wells. The vertical and horizontal lines represent the AUCs of the model that was used to choose cells for isolation. **e.** A histogram of AUCs for models detecting mutants in a mixture of wild-type cells and four pathogenic mutants. On top, a histogram of AUCs generated from randomly assigning models is shown as a comparison.

Identification of Weak Pathogenic *MFN2 Mutants*

Our pipeline also has the ability to recognize relatively weak phenotypes. As shown earlier, of the six pathogenic mutants, the R280H and P251A appear most like benign mutants and wild type (**Figure 2c**). Despite the subtle phenotype, we recovered R280H cells in the experiment above (**Figure 3b**), though comparably fewer than cells with other mutations. Given those results, we explored the ability of Raft-Seq to isolate cells with the R280H mutation compared against *MFN2* wild type. **Figure 4a** shows images of R280H mutation-harboring cells in the microraft plate. As a baseline measurement against more conventional methods of cell separation, we ran separate samples of a *MFN2* WT and R280H mutant cell lines each stained with MitoTracker through flow cytometry (**Figure 4b**). While there is variation between the two in the MitoTracker intensity, there is too much overlap to separate a mixture (**Figure S4** AUC = 0.60).

We then followed the Raft-Seq process described above, replacing the mixture of *MFN2* mutants with just R280H mutant cells, and therefore training the models directly on the weaker phenotype. Additionally, the experiment was done with different culture densities and cross-compared to confirm that culture conditions had no 'residual phenotype' that impacted the models prediction. **Figures 4c** and **4d** show histogram and scatterplots for the resulting models in this experiment. While the AUCs for the best generated models were not as high as those from the earlier experiment containing all mutants, our process still easily discriminated between the R280H mutant and the *MFN2* wild type (AUC of picked model 0.72, best models > 0.8, n_{WT} = 265, n_{R280H} = 205). We also ran a similar experiment comparing the P251A pathogenic mutant cell

line—the other pathogenic mutation with a weaker phenotype—to the D221= benign mutant cell line. Cells containing the synonymous substitution D221= act as another control for point mutations instead of using WT. **Figure S5** shows that the resulting model, when applied to a mixture of the cell lines, was able to discriminate between the two (AUC = 0.8, $n_{D221=}$ = 103, n_{P251A} = 65). The AUCs mentioned above apply to the majority of unlabeled cells that were physically picked. If we limit our results to cells whose class the model is increasingly confident about, then the accuracy of the model gets increasingly better (**Figure 4e, f**). For cells that the model is at least 80% confident about (prediction score ≥ 0.8 or ≤ 0.2), we get high AUCs and accuracies in both the R280H/WT and P251A/D221= experiments (AUC = 0.94,0.98 Acc = 0.88,0.92 n=17,24).

We have shown that Raft-Seq can accurately predict genotypes from strong mitochondrial point mutants as well as weak mutants in *MFN2* and predict them as part of a mixed culture where there was no *a priori* knowledge of individual cell's genotypes. The model's predictions were realistically tested by isolating single cells and genotyping them to reveal the method's accuracy.

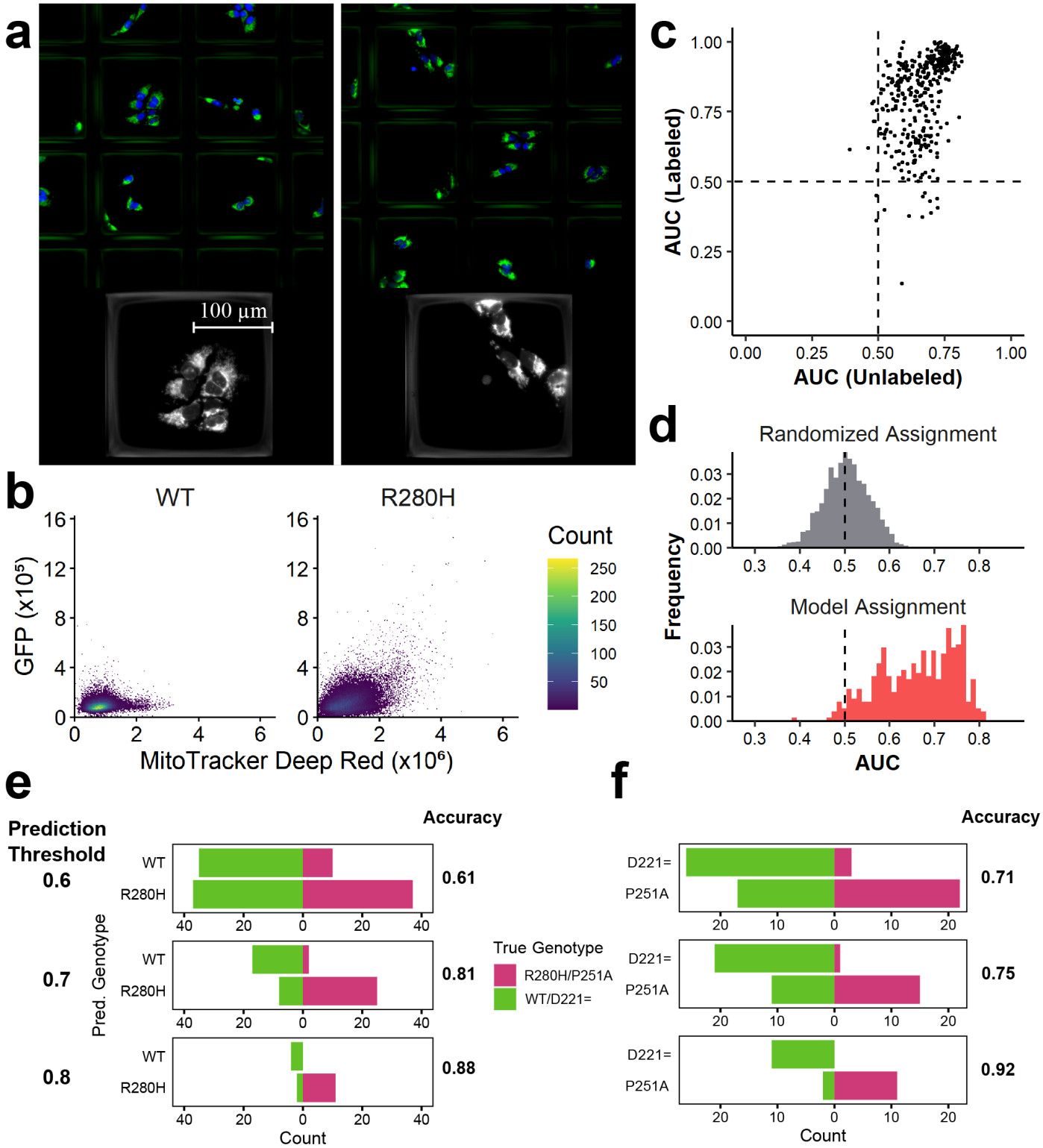


Figure 4. Raft-Seq can Predict *MFN2* WT and *MFN2* R280H mutant cells which have nearly undetectable Phenotypic Differences. a. Images of the R280H mutant and *MFN2* WT cell line in the microraft plate. **b.** Side-by-side flow cytometry scatter plots of the *MFN2* WT and R280H mutant cell (GFP was present equally in both cell lines). **c.** Scatterplot showing the performance

of all 433 models in detecting mutants in a mixture of wild type cells and the R280H mutant. **d.** A histogram of AUCs for models detecting mutants in a mixture of wild-type cells and the R280H mutant. On top, a histogram of AUCs generated from randomly assigned models for comparison. **e,f.** Accuracies and confusion bar charts of the predictions from the un-labeled wells, when only picking cells with prediction scores ≥0.9 (≤0.1), ≥0.8 (≤0.2), and ≥0.7 (≤0.3). **e.** uses the model to predict R280H against WT *MFN2*, while **f.** uses the model to predict P251A against D221=.

Modeling Analysis

Multi-feature models were necessary to yield accurate predictions, warranting exploration into the importance of individual features and the performance impact associated with them. Using data from the experiment containing wild-type *MFN2* cDNA and four pathogenic mutants, we generated additional models by varying the number of features used, sampling from the 8 features in the pick model (**Supplemental Table 1**). Histograms showing the distributions of the AUCs, separated by the number of features are shown in **Figure 5a.** When more features are added, the models can better distinguish between the populations of cells. We also examined the importance of individual features in building models (**Figure 5b-d**). We used the original feature data (**Figure 1**) and generated Kruskal-Wallis χ^2 for each feature between the wild type and pathogenic mutant cells from labeled populations. A ranking of the 16 features associated with the highest χ^2 are shown in **Figure 5b**. We compared the resulting AUCs from models built off each specific feature alone (**Figure 5c**), and to the resulting AUCs from models built from the 16 features minus one (**Figure 5d**). Importantly, we find that single features *are not* the key to discriminating these clinically relevant pathogenic *MFN2* point mutants. Instead, small numbers of relevant features can inform a useful model.

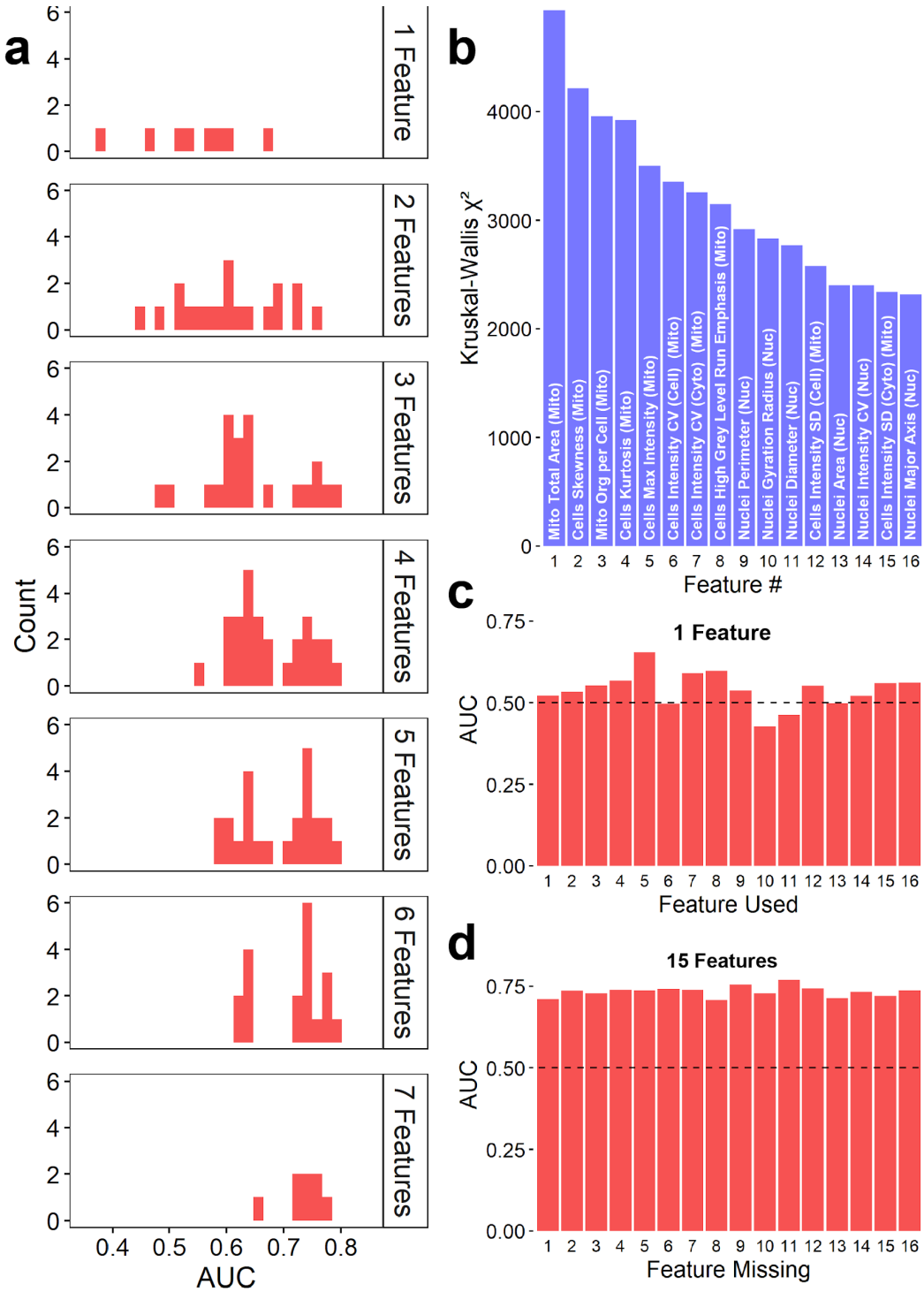


Figure 5. *MFN2* Genotype Prediction does not depend on specific features. a. Histograms of AUCs resulting from models built from combinations of a set of 8 features used when picking, arranged in panels split by the number of features used in each subset. **b.** Bar chart of the Kruskall-Wallis χ^2 values for the 16 features with the highest values. "Mito" and "Nuc" indicate features measured from mitochondrial and nuclear stains, respectively. **c.** Bar chart of the AUCs resulting from models built out of each individual feature. Note that the most significant feature was unable to produce a good model alone. **d.** Bar charts of the AUCs resulting from models

built using all but one feature (leaving 15 features). All AUCs listed indicate the performance of the model trained from labeled data on their ability to predict *unlabeled* cells in admixed conditions.

Scanning Mutagenesis with *MFN2* VUS

While the previous experiments were done by over-expressing a mutant *MFN2* cDNA, the following set of experiments tested genome editing of an endogenous gene. First, we used an existing U2OS line with mutations in the mitochondrial primase *PRIMPOL* [40] and found that it alters mitochondrial morphology. We followed the Raft-Seq process described above, using the mutated line and a wild type U2OS line as the two labeled populations (**Figure 6a**). The final ROC of our best model generated on-the-fly (AUC = 0.90), verifying the platform performs similarly for an endogenous genetic perturbation. We next sought to scan through *MFN2* to elucidate the impact of several VUS. Next, we constructed a CRISPR-Cas9 gRNA library targeting different exonic regions of *MFN2* on sites near known ClinVar variants (**Figure 6b**). The library was delivered by lentivirus to U2OS cells with a dox-inducible Cas9 construct, and the cell population was split in two, with only half receiving doxycycline. In the experiment, 40,000 cells were imaged, and anomaly detection modeling—an unsupervised learning algorithm, unlike those previously mentioned—was used on the feature data to distinguish the non-induced cells from the Cas9-activated cells. The scatterplot of the cells in the anomaly detection space is shown in **Figure 6c**, with a boundary highlighting cells that are outliers. The identification of anomalous cells was also confirmed by cross-referencing against many more anomaly detection models (n_{models} = 8). This was done by keeping track of which cells were consistently part of the top 10, 20, or 30 (out of 40,000) most anomalous. **Figure 6d** shows a bar chart of the number of times a cell

appeared in the top 10, 20, or 30 most anomalous cells, highlighting the 30 cells most frequently identified. Anomalous cells were consistently labeled as such across models, lending credence to the idea that these models were truly identifying cells with unusual phenotypes.

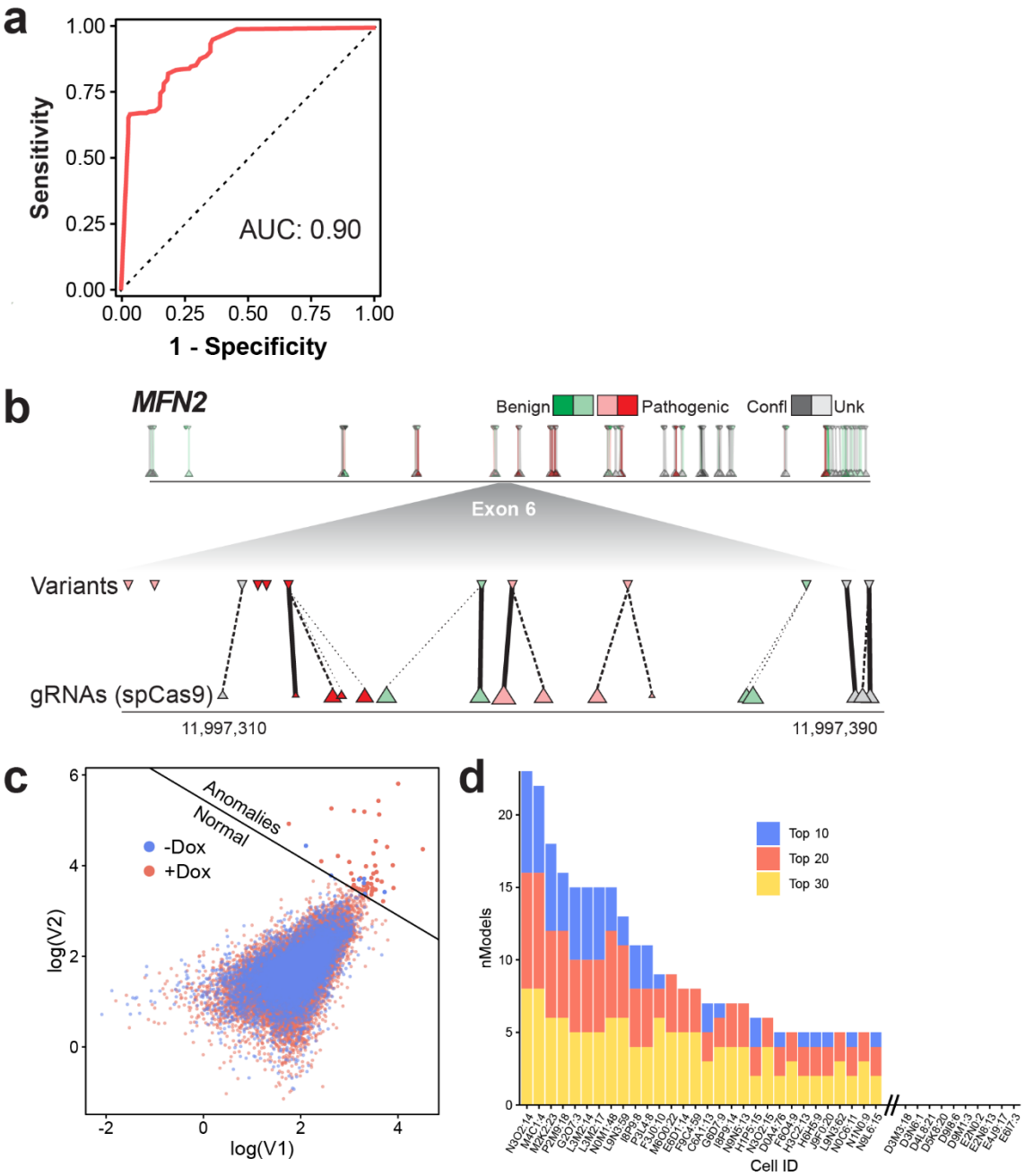


Figure 6. Endogenous Scanning Mutations in MFN2 simulate Variants of Uncertain Significance. a. Raft-Seq demonstrated with an endogenous mitochondrial gene (PRIMPOL) in

the setting of CRISPR-Cas9 indel formation using on-the-fly training. **b.** Diagram of coding sequence with ClinVar variants (downward facing triangles) and their closest gRNA cut sites (upward facing triangles). The upper line shows the entire genomic context of the MFN2 gene, while the lower inset magnifies the variants and gRNA cut sites for exon 6. Lines indicate the 'partner' gRNAs for each ClinVar mutation, where thicker lines indicate closest association. **c**. U2OS cells with inducible Cas9 were first selected for the presence of a library gRNA against MFN2 and then induced with doxycycline (+Dox) or left un-induced as a control (-Dox). Two anomaly detection models were generated from the control cells and used to identify cells from the +Dox conditions that had anomalous mitochondrial phenotypes. **d.** Bar chart of the cells (rafts) with the most perturbed mitochondria, where multiple machine-learning models often identified the cells.

Discussion

We have developed a method, Raft-Seq, to efficiently screen many genetic variants based on their impact on a cell's phenotype. We have shown that it can effectively discriminate between wild-type cells and cells containing different pathogenic point mutations of the *MFN2* and *PRIMPOL* genes. Though we utilize a particular individual model for the selection of cells for isolation, we found that most of the models that we generate can identify pathogenic mutant cells. Since our feature selection process is only mildly dictated by the actual phenotype that we are looking for, our ability to predict a cell's genotype comes from quickly generating complex computational models. This will enable the pipeline to exploit many more phenotypes that the scientific literature is not yet familiar with, therefore making almost any gene in the genome amenable to this functional screening and thus enable the re-cataloging of VUS as benign or pathogenic.

We developed Raft-Seq primarily as a screening platform to work in concert with Deep Mutational Scanning [41], a method for creating a library of every single possible mutation in a gene. Our results show that such a screen would correctly call pathogenic

variants, since we were able to recover pathogenic *MFN2* mutants from a mixture with wild-type cells. However, our results also show some limitations. For one, based on the accuracy difference between the two unlabeled wells which were at different proportions, our ability to recover pathogenic variants appears to decrease the smaller proportion of pathogenic variants present. This would indicate that such a screen would perform significantly better on genes for which mutations are more likely to result in a pathogenic variant, but this could be countered with a higher n. Second, we were able to identify several mutants (L76P, R94Q)—which had more extreme phenotypes—more frequently than other mutants, meaning that a scaled-up screen would most likely result similarly, and more extreme variants would be overrepresented. However, because we were able to identify the weaker mutants (R280H, P251A) with relative ease when they were not mixed with any other mutants, we can assume that overrepresentation of specific variants is not due to the absolute strength of the resulting phenotype. Instead, it is due to the relative strength when compared to other variants. Penetrance may also play a role since weaker phenotypes may have a strong phenotype in some cells but have low penetrance in the population.

Since Raft-Seq isolates cells for sequencing individually rather than in pools, we have a few advantages over platforms. For one, we find the specific genotype of every isolated cell, rather than perform batch measurements, meaning that we can find effects of combinations of perturbations. We can also theoretically have as many categories in our machine learning models as cells, though using more than a handful of features causes the modeling and analysis to deteriorate, likely due to overfitting (but again is countered by higher n). Lastly, Raft-Seq is amenable to generate isogenic cell lines

directly from the primary screen, since the entire process is performed alive, and cells are picked with no flow pressure while they stay adhered to their raft. These clonal/isogenic lines can then be used for a number of downstream assays like biochemistry, metabolism, bulk sequencing, and more.

An advantage of this approach is the flexibility that comes from using machine learning to identify phenotypes. Since we can easily combine features for more accuracy, we can theoretically screen for any visible phenotype, given a strong enough signal provided by staining or other fluorescence. Unsupervised learning methods, like clustering, are also possible and allow for a simpler setup since no labeled wells would be needed.

Since the raft identifies the location of the cell stably over time, we can do on-the-fly training from the entire experiment, then go back and pick those cells. We have also found that Raft-Seq can be used to apply past training (labeled) data to future experiments. We are now working on a normalization framework to accrete our previous training data and utilize it for future models (allowing a gain of accuracy in identifying specific phenotypes over time).

Raft-Seq is slower than flow cell sorting, possibly limiting scalability. All the image-based techniques require time for imaging, and this system is no exception. As evidenced by the recent genome-wide screen [25], the throughput of this system is comparable, and the number of cells captured by that system is slightly lower over an experiment of comparable scale. Regardless, even higher throughput is still advantageous. There are fewer limits in terms of the number of cells able to be imaged, so by increasing the pool of isolation candidates, cells are selected with a higher level of

confidence. There are two steps in Raft-Seq that are done manually but we are currently in the process of automating. The first is image quality control which can be automated by a convolutional neural network similar to [33]. The second is model selection, which can be automated by choosing among several model performance metrics on the labeled data.

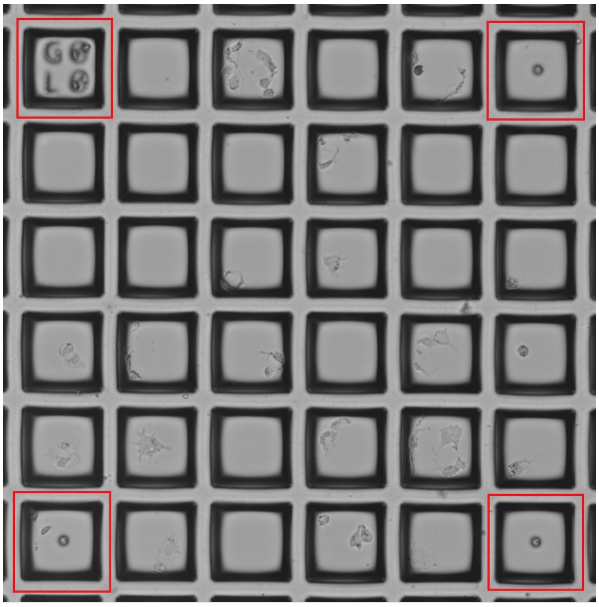
We are optimistic that with increased scale, Raft-Seq can be used to sensitively find variants across a wide breadth of perturbed cellular phenotypes. An increase in scale would provide more training data to make models more accurate and identify more cells to which the models assign a high prediction score, which we have shown to raise overall accuracy. Altogether, we have shown that using various cell lines (HEK293, U2OS, A549), with multiple types of genetic perturbation (endogenous disruption with Cas9, overexpression of mutant genes), that we can deploy flexible machine learning (logistic regression, decision trees, SVM, neural networks) from data within an experiment (on-the-fly) or from previous experiments (pre-trained) to select clinically relevant point mutations in a screening setting. Since this system captures the cells individually, we can carefully evaluate many 'selection criteria' by generating a variety of different machine-learning models and finding how well they perform. This technique will be an important tool in the advancement of precision medicine to identify variants relevant to disease.

Methods

All custom software can be found at https://gitlab.com/buchserlab/FIVTools

Microraft Plate

CytoSort array plates are received from Cell Microsystems (Durham, North Carolina). For all experiments in this paper, either 100x100 or 200x200 micron quad reservoir arrays were used. Each plate is made up of rafts each having a four-character alphanumeric coordinate (Raft ID). Fiduciary markers located at fixed locations on the plate (shown below) in conjunction with custom software were used to locate individual rafts for map generation (see **Image Post Processing** and **Image and Cell Quality Control**).



Cropped example field of CytoSort array plate: Alphanumeric fiduciary markers are etched on every 10 rafts (e.g. G0L0), and dot fiduciary markers are etched on every 5 rafts. Rafts with fiduciary markers are highlighted with red outlines.

All microraft isolation was performed using the Cell MicroSystems CellRaft Air System. Given a list of raft coordinates, the Air System uses a needle to eject each individual raft and transfer the raft to a 96-well PCR plate via a magnetic wand. Raft isolation is confirmed twice through post ejection imaging of the raft location and through identification of a raft in the PCR collection plate. Rafts can also be transferred into 96-well tissue culture plates for clonal cell growth.

Plates were prepared by rinsing with 1mL 1x PBS prior to plating. $1.2x10^4$ cells per well were plated and incubated overnight (14-16 hours).

Cell Culture and Transfection

Human osteosarcoma (U2OS, ATCC HTB-96) cell lines were maintained in McCoy's 5A Modified Medium (16600082, Gibco, Gaithersburg, MD, USA) supplemented with 10% fetal bovine serum (FBS) (16000044, Gibco). Human embryonic kidney (HEK) 293T cells (CRL-11268, ATCC) were cultured in Dulbecco's Modified Eagle's Medium (11965-092, Gibco, Gaithersburg, MD, USA) supplemented with 10% (16000044, Gibco, Gaithersburg, MD, USA), 1% Penicillin-Streptomycin (15140122, Gibco) and 1% non-essential amino acids (11140050, Gibco).

All cell lines were maintained in an incubator at 37°C, 5% CO_2 and observed daily for growth and overall health. Cells were passaged using 0.25% Trypsin-EDTA 1x (25200056 Gibco, Gaithersburg, MD, USA) at a subcultivation ratio of 1:10. All live cell counting necessary was performed with the BioRad TC20 automated cell counter. All centrifugation of cell cultures was performed at 1200 rpm for 3 minutes. STR profiling, to confirm cell type, was performed using NGS-based analysis by Genome Engineering and iPSC Center (GEiC) at Washington University in St. Louis (Chen et al).

Virus Production

MFN2 lentiviral expression plasmids were cloned into the CCIV lentiviral expression plasmid with a GFP expression marker [37]. In preparation for lentiviral packaging, 8.0×10^5 HEK293T cells were plated into each well of a six well tissue culture plate and incubated overnight. The

cells were then transfected with TransIT Lenti-transfection reagent (MIR 6600, Mirus Bio, Madison, WI, USA) using an envelope plasmid (pVSVg: Addgene plasmid # 8454), a packaging plasmid (psPAX2: Addgene plasmid # 12260), and each individual MFN2 expression plasmid in a mass ratio of 0.5/1/0.5 for a total of 2µg. After 48 hours, media was collected, centrifuged, and sterile filtered before being concentrated with Lenti-X Concentrator (631232,Takara Bio, Kusatsu, Shiga, Japan). Concentrated virus was resuspended in 200µL 1xPBS per well created, and stored at -80°C.

Viral Stable Cell Line Production

U2OS cells were infected with 70µL of concentrated lentivirus and given polybrene (NC9840454 Santa Cruz Biotechnology, Texas) at a final concentration of 10µg/mL, then incubated for 24 hours, media was changed after infection. Cells were taken to the Washington University Siteman Flow Core for fluorescent sorting on the Sony Synergy HAPS1, 100-micron sorter. Cells were sorted, based on viability and GFP expression. GFP expression levels were compared both within and across generated cell lines to ensure population purity and comparable fluorescent expression levels. We received the PRIMPOL KO U2OS cell line from the Vindigni lab which was produced at the Washington University Genome Engineering and iPSC Center (GEiC).

CRISPR/Cas9 gRNA Library Infection and Induction

We generated a dox-inducible Cas9 (iCas9) U2OS cell line via CRISPR-mediated homology directed repair. The Cas9 protein, gRNA, and donor construct were introduced via nucleofection. Presence of the construct was validated via junction PCR [42]. Isogenic iCas9

clones were isolated using the Air system and then propagated for further experiments. The MFN2 scanning gRNA library was generated and constructed by Washington University GEiC, then virus was generated (as above). iCas9 U2OS cells were infected with the lentiviral library at an MOI of <0.2, followed by 8μg/mL puromycin selection for 7 days. At 60-70% confluency, Doxycycline (Cat#: D9891-1G, Millipore Sigma) was added at a final concentration of 5 mg/mL, and the cells were incubated at 37°C for 48-60 hours to before proceeding with staining and imaging.

Staining

In this manuscript the following cellular structures were stained using vital dyes (Table 1): DNA labeling (Hoechst, Thermo Fisher H3570), mitochondria (MitoTracker Deep Red, Thermo Fisher M2246), and mitochondrial membrane potential (TMRM, Thermo Fisher I34361). MitoTracker and TMRM were incubated for 40 minutes at concentrations of 500nM and 1μM respectively. Hoechst was incubated for 15 minutes at a concentration of 10μg/mL. Each plate was rinsed twice with culture media prior to imaging.

Imaging

All imaging was conducted using a Cytiva INCell 6500HS high content confocal microscope. Custom plate form factor files were developed to capture the entirety of the plate surface that contained rafts (available upon request). Images were captured using a 20x .45 objective. Software autofocus was performed in the 405 nm (nuclear) channel for every field. Exposure times for Hoechst (405 nm) and TMRM (561 nm) averaged around 0.15 seconds while MitoTracker Deep Red (642 nm) averaged 0.05 seconds. Confocality was used in the 405 and 642 wavelengths to decrease the background fluorescence of the raft array plate. Each field of

view overlapped by 12% of their area, allowing for better maintenance of focus throughout imaging. Imaging settings were held constant throughout the course of an experiment. Following imaging, an additional 500µL medium was added to each well to ensure enough volume during isolation.

Image Post Processing

Following plate imaging, raft coordinates were mapped to images to allow for the selection of individual microraft using the Air System. Custom software was developed that uses three manually identified rafts to calculate the rotation angle and shear of the raft array plate in the image set. This information allows the program to map microscope coordinate data to the row and column of every microraft using a series of linear transformations. Images are also curated semi-manually to ensure that out-of-focus images are discarded.

Image Tracing and Feature Extraction

Following imaging, tracing and feature extraction was performed using Cytiva's INCarta software. Cells were identified using nuclear segmentation, then the cell body was defined as a collar expanded out 20 µm. Mitochondrial puncta were identified and counted for each cell as well as a range of texture features. Feature analysis settings remained unchanged for all imaging for a single experiment to ensure consistency.

Image and Cell Quality Control

The cell feature dataset was joined with the image quality data and raft position mapping data described above by custom software (included in GitLab repository). Post tracing quality control

was performed with each dataset in Tibco Spotfire Analyst. First, aberrant tracing artifacts were excluded based on nuclear area, form factor, and proximity to the raft grid. Next, non-nuclear debris and dead nuclei were excluded using nuclear area, intensity, and cell intensity. One set of cells traced in the overlap region were excluded and plates were checked for sagging and other errors. Distribution of cells and tracing error / dead cells were scrutinized in the QC workflow to look for and exclude systemic errors. Rafts with too many cells (>6) were excluded from the analysis. Finally, rafts were excluded for failing initial manual image quality control and if they contained a fiduciary marker. This filtered set of cells was used as the input for machine learning downstream.

Machine Learning and Model Generation

After exporting the quality-controlled cell-based feature table, we next needed to build a machine learning model that could distinguish pure populations of WT from pathogenic mutant cells that were plated as controls with the unlabeled wells. The models were generated quickly since the cells were alive and needed to be physically picked within the next few hours. Thus a variety of machine learning platforms and algorithms were applied in order to predict an unknown cell's genotype and give a higher chance of success. Four different platforms were utilized in this study, including Microsoft AzureML Studio, Tibco Spotfire, Tibco Statistica, and H2O.ai. Generally, logistic regression was performed in Spotfire, then Random Forests, Boosted Trees, Support Vector Machines, Gradient Boosted Machines, and Artificial Neural networks were trained in the other platforms. Models were evaluated on labeled populations that were withheld during training. Based upon model performance on the testing dataset, a model was selected and deployed to the unlabeled cell populations. Starting with the best prediction scores, a list of cells with raft locations was generated.

Cell Capture and DNA Extraction

Cells were isolated using the Air System (see **Microraft Plate**). The order of raft picking was randomized. Rafts with cells-of-interest were isolated directly into 5µL extraction buffer (molecular grade water with 10mM Tris-HCl (pH 8.0), 2mM EDTA, 200 µg/mL Proteinase K, and 0.2% TritonX-100) in 96-well PCR plates. Genomic DNA was extracted in a thermocycler immediately following raft isolation by incubating at 65°C for 15 minutes then 95°C for 5 minutes. All plates were manually inspected to ensure rafts were deposited successfully.

Single-Cell DNA Amplification

Amplification of single-cell DNA prior to library preparation consists of two separate amplifications. An initial preamplification is conducted using extracted DNA with KOD Hot Start DNA Polymerase (71842-4, Millipore Sigma, Burlington, MA, USA) according to manufacturer's instructions using all 5µL of extracted DNA in a total reaction volume of 20µL. Pre-amped product was processed through an AMPure XP (Catalog: A63882, Beckman Coulter, Brea, CA, USA) bead clean up according to the manufacturer's instructions using 10mM Tris-HCl pH 8.5 as elution buffer. The second amplification uses the cleaned template and BioLine MyTaq HS Red Mix 2x (C755G97,Meridian Life Sciences, Memphis, TN, USA), according to manufacturer's instructions, including 5% by volume DMSO. Primers in the second amplification contained universal 5' tags to be compatible with Illumina library preparation (Forward tag: 5′-CACTCTTTCCCTACACGACGCTCTTCCGATCT-3′, Reverse tag: 5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3′).

For amplification of *MFN2* cDNA, primers amplifying the entire cDNA were used in the first amplification step, followed by multiplexed amplification of two specific regions containing the

relevant mutations. Genotyping of the RFP-GFP cells used multiplexed primers that amplified specific regions in both the RFP and GFP regions. All primers are listed below in table 2.

Name	Sequence (Excluding tags, where necessary)	PCR Stage
pMFN2.All.F	GCTCTTCTCTCGATGCAACTCT	1
pMFN2.All.R	GCAGGTACTGGTGTGTGAAC	1
pMFN2.1.F	CACATGGCTGAGGTGAATGC	2
pMFN2.1.R	GCAGGAAGCAATTGGTGGTG	2
pMFN2.2.F	CTCAGAGTCCACCCTGATGC	2
pMFN2.2.R	CACTTGAAAGCCTTCTGCGAG	2
RFP.F	GTTCATGCGCTTCAAGGTGC	1, 2
RFP.R	CAAGTAGTCGGGGATGTCGG	1, 2
GFP.F	TGAAGTTCATCTGCACCACCG	1, 2
GFP.R	TCGCCCTCGAACTTCACCTC	1, 2
PRIMPOL.F	GCAACCCAGTTTTGAAACCA	1, 2
PRIMPOL.R	TCGATGTCCAGCTTTCCTCT	1, 2

Table 2. Genotyping Primers.

Illumina Library Preparation

These methods are expanded from Connelly et al. and Bell et al. After amplification with universal illumina tags, each plate is amplified with specific forward and reverse primers that indicate the PCR plate position and a unique library ID. PCR amplification was performed with EconoTaq PLUS GREEN 2X Master Mix according to the manufacturer protocol. We generated 2 × 250 reads with the Illumina MiSeq platform at the Center for Genome Sciences and Systems Biology (Washington University).

Sequence Analysis

Illumina paired reads were demultiplexed by the core facility and FastQ files were returned. The rest of the analysis was performed with laboratory software available on Gitlab (called “Library Aligner”). Reads were joined and trimmed then aligned with small sequence fragments at the genetic sites of interest containing the sequence to mutant or WT alleles. The end result was a ‘counts’ table that gave the number of reads containing each 20-mer for each well. 20-mer search fragments listed below in **Table 3**. After accounting for isolation and genomic amplification errors, around 80% of the isolated cells genotypes’ were captured.

Name	20-mer (relevant mutations/deletions bolded)
MFN2_V69	TGGACCCCGTTACCACAGAA
MFN2_V69F	TGGACCCC**T**TTACCACAGAA
MFN2_L76	ACAGGTTCTGGACGTCAAAG
MFN2_L76P	ACAGGTTC**C**GGACGTCAAAG
MFN2_R94	TGCTGGCTCGGAGGCACATG
MFN2_R94Q	TGCTGGCTC**A**GAGGCACATG
MFN2_D221=	CTGGATGCTGATGTGTTTGT
MFN2_D221=	CTGGATGCTGA**C**GTGTTTGT
MFN2_P251	CTCTCCCGGCCAAACATCTT
MFN2_P251A	CTCTCCCGG**G**CAAACATCTT
MFN2_R280	CATGGAGCGTTGTACCAGCT
MFN2_R280H	CATGGAGC**A**TTGTACCAGCT
MFN2_W740	AAAGCCGGTTGGTTGGACAG
MFN2_W740S	AAAGCCGGTT**C**GTTGGACAG
RFP_guide	GGCCACGAGTTCGAGATCGA
RFP_control	AAGGTGCGGATGGAGGGCAG

GFP_guide	TGCCCGAAGGCTACGTCCAG
GFP_control	CTACCCCGACCACATGAAGC
PRIMPOL_WT	GATAGCGCTCCAGAGA**C**AAC
PRIMPOL_del	GATAGCGCTCCAGAGAAACA

Table 3. 20-mer fragments.

For *MFN2* cDNA genotyping, each mutation locus was given a %mutant score calculated as the number of mutant reads divided by total number of reads at that locus. Cells were designated as wildtype if no locus had >50% mutant score, otherwise they were designated as a specific mutant based on which locus had the highest mutant score. Cells with a highest %mutant score around 50% were excluded from downstream analysis, due to ambiguity. Lastly, a flat file was exported containing each picked raft and its assigned genotype. All sequencing analysis after generation of the flat file was performed in Tibco Spotfire.

Accuracy and ROC Measurements

Using our custom software (same package as above, the “AUC” feature), we joined the modeling and genotyping flat files to find overall accuracy and generate ROC curves for each model. We also generated random noise ROC curves by shuffling the assigned genotypes. The final exported data was then analyzed in Tibco Spotfire. For the AUC analysis, the program filtered out cells with prediction scores that were between 0.4 and 0.6 (the prediction scores were poor meaning the particular model was unable to classify these cells). For the data presented in **Figure 4e,d**, this threshold was further adjusted.

HEK293 Cell Nucleofection

SF cell line solution stock was prepared from the Lonza SF Cell line 4D-Nucleofector LV Kit XL (V4LC-2520) kit by combining 82µL of SF Cell Line Nucleofector™ Solution with 18µL Supplement 1. 20µL of stock solution was then combined with 2µL Cas9 protein (QB3 MacroLab) and 2µL gRNA. The complexes were incubated on the benchtop for 10 minutes. Cell suspensions of $2x10^5$ cells were placed in 1.5 mL microcentrifuge tubes and spun down. Supernatant was carefully removed avoiding the cell pellet. The cells were rinsed in 1x PBS and centrifuged a second time. Supernatant is carefully removed, and the cells were resuspended in the final combined SF cell line solution and transferred to a nucleofector cuvette, found in the nucleofector kit. The cuvette was placed in the Lonza 4D-Nucleofector Unit. (Lonza AAF-1002X, AAF-1002B) and nucleofected with Pulse code CM130. Nucleofected material was added to a prewarmed 6 well plate with 5mL of DMEM media in each well. Nucleofected cells were incubated for 48 hours for recovery. The cell line solution is not healthy for the cells, so speed is a priority upon resuspension in cell line solution.

Supplementary Methods

RFP-GFP Cell Perturbation and Isolation

A HEK293 cell line expressing both RFP and GFP (Gentarget #SC009) was used in this experiment. Cells were then nucleofected with a gRNA targeting GFP (TGCCCGAAGGCTACGTCCAG) or RFP (GGCCACGAGTTCGAGATCGA) and Cas9. All gRNAs were ordered from Synthego. Nucleofection was conducted using a Lonza 4D-Nucleofector Unit. (Lonza AAF-1002X, AAF-1002B). Following the Raft-Seq workflow, cells were imaged and the guide presence was predicted by a combination of RFP and GFP intensity features. Cells were selected and isolated into 96-well plates. Alternatively, cells were sorted using a Sony SH800S cell sorter individually by their RFP and GFP fluorescence. Cells were then genotyped and designated as being given the RFP or GFP guide if the locus that the respective guide targets had been altered.

Modeling Considerations

We implement several specific design criteria during the experimental setup and model selection. First, the ‘pure’ samples are split into testing wells and training wells so that we can decrease overfitting during model selection. For most of the experiments presented herein, our training datasets comprised two of the three labeled population wells of each type—the two wells must come from different plates—and use the remaining two labeled wells as “testing sets”. We split the data by well rather than randomly assigning cells from all pure population wells to mitigate the influence of batch effects during modeling. It is important to note that the actual data used for ultimate validation of the models will be the identities of the perturbagen in

the unlabeled cell population. These are independent of the model scores of the pure populations, so we do not have to be concerned about hyperparameter overfitting from examining the testing data. Second, the testing wells are spread across different plates to hopefully account for batch effects. Since training data and testing data technically come from separate samples, we observed several odd phenomena such as prediction quality improving for testing data. These phenomena would most likely not be observed if the model-generating data were truly randomized, but we felt it was important to retain any variance caused by batch effects, since batch effects present a large and consistent obstacle to cell imaging analysis (Caicedo et al.).

Relatedly, we make sure not to create any artificial batch effects generated by inconsistencies in the screening process. All stains used are prepared as a single batch to be used across plates. All imaging and feature extraction settings are kept constant across each plate in a screening experiment.

Another aspect of model design that we consider is the exclusion of ‘leaky’ variables, or variables that happen to correlate with the data labels. Since our labeled data exists on specific positions of the plate, including a variable that is effectively a proxy for cell position would render the model useless on unlabeled data. We have found that several variables that are in the standard output of IN Carta feature extraction are leaky, such as features measured on a global basis, and we exclude them before starting the modeling process.

We also face a time constraint in our modeling pipeline. This time constraint is caused by the necessity of the locations of cells on the plate staying constant between imaging and isolation. Though we can fix the cells, a modeling pipeline based on that assumption would limit our ability to perform Raft-Seq for the isolation of live cells. Optimally, we complete feature selection, modeling, and raft selection within six hours of imaging, and continue on to isolation. This of course, requires a lot of computational power.

Supplementary Data

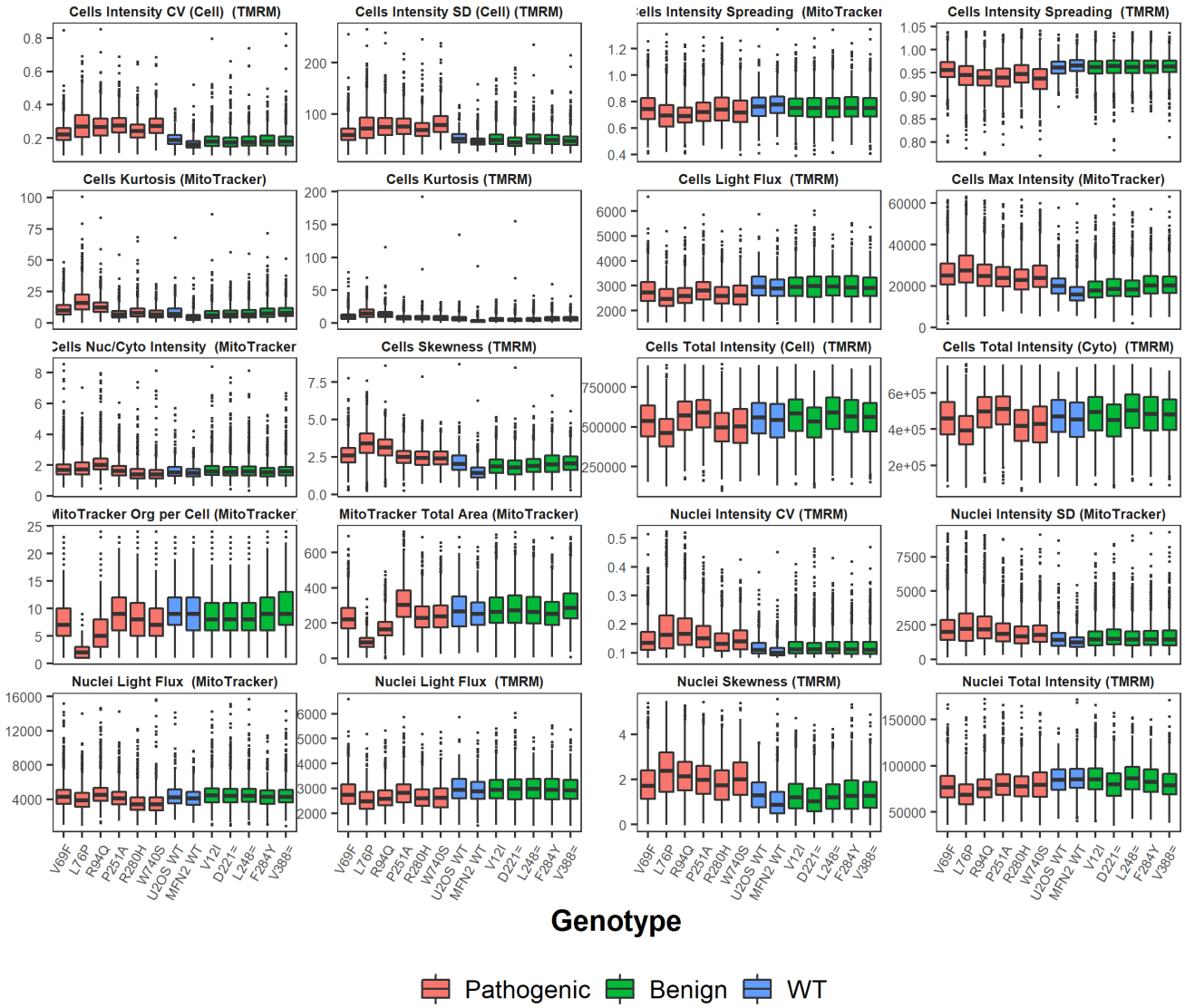
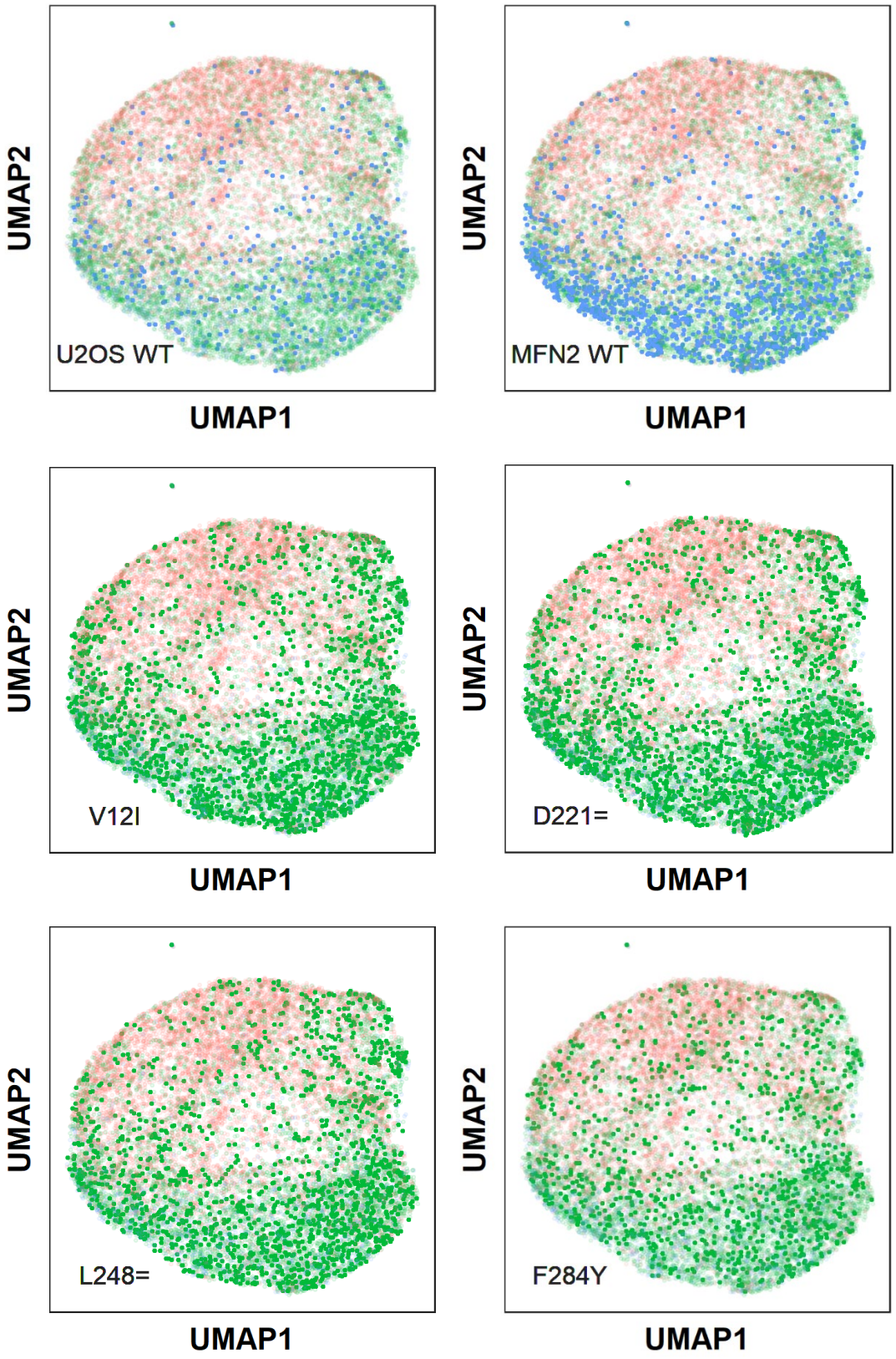
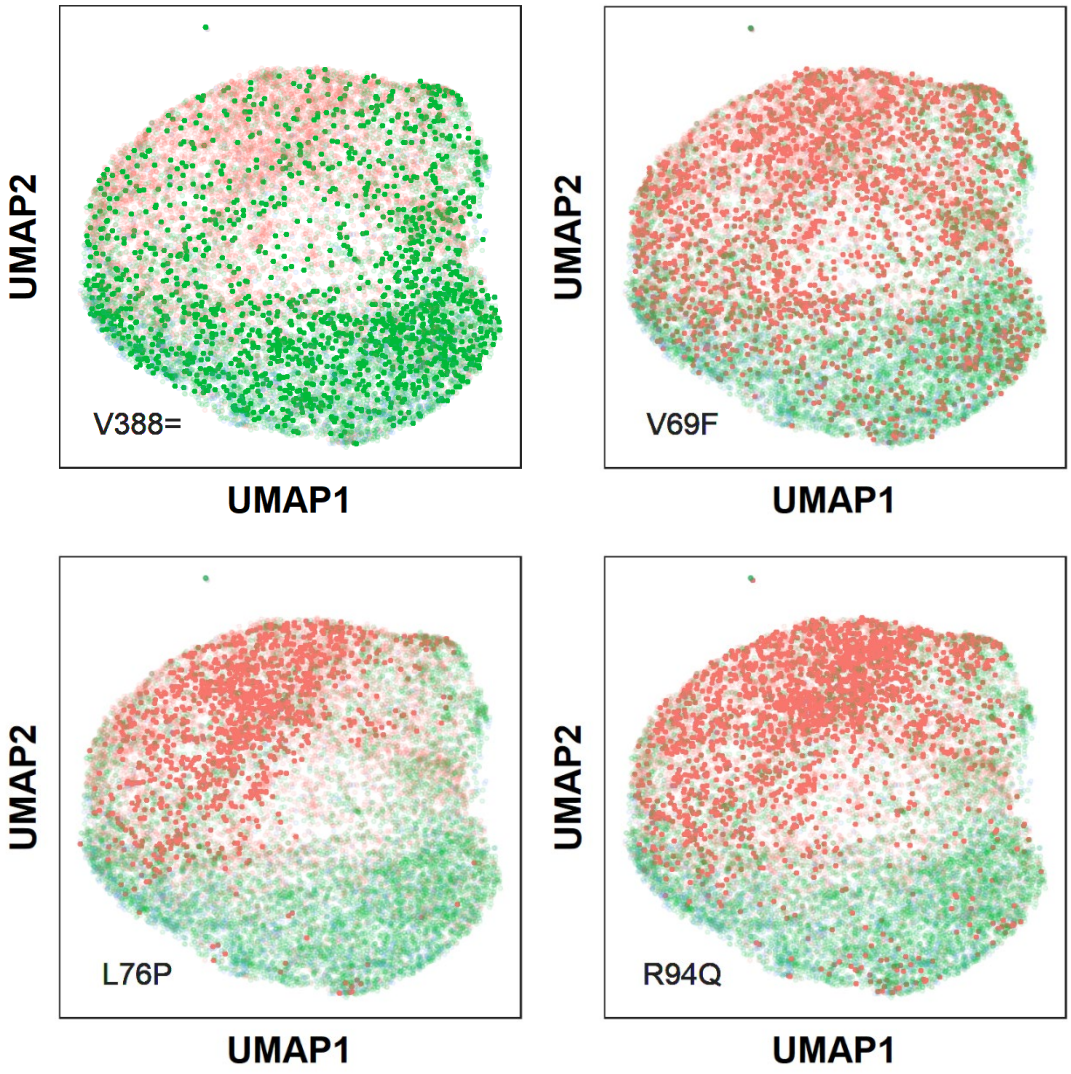


Figure S1: Box plots of important features extracted from TMRM and MitoTracker staining. Histograms are colored by whether the cell was wild type (unmodified wild type or modified with a wild type MFN2 cDNA), contained MFN2 cDNA containing a benign mutant, or contained MFN2 cDNA containing a pathogenic mutant. Features that showed no apparent difference between mutants were omitted.


UMAP2
U2OS WT
UMAP1
UMAP2
MFN2 WT
UMAP1
UMAP2
V12I
UMAP1
UMAP2
D221=
UMAP1
UMAP2
L248=
UMAP1
UMAP2
F284Y
UMAP1


UMAP2
V388=
UMAP1
UMAP2
V69F
UMAP1
UMAP2
L76P
UMAP1
UMAP2
R94Q
UMAP1

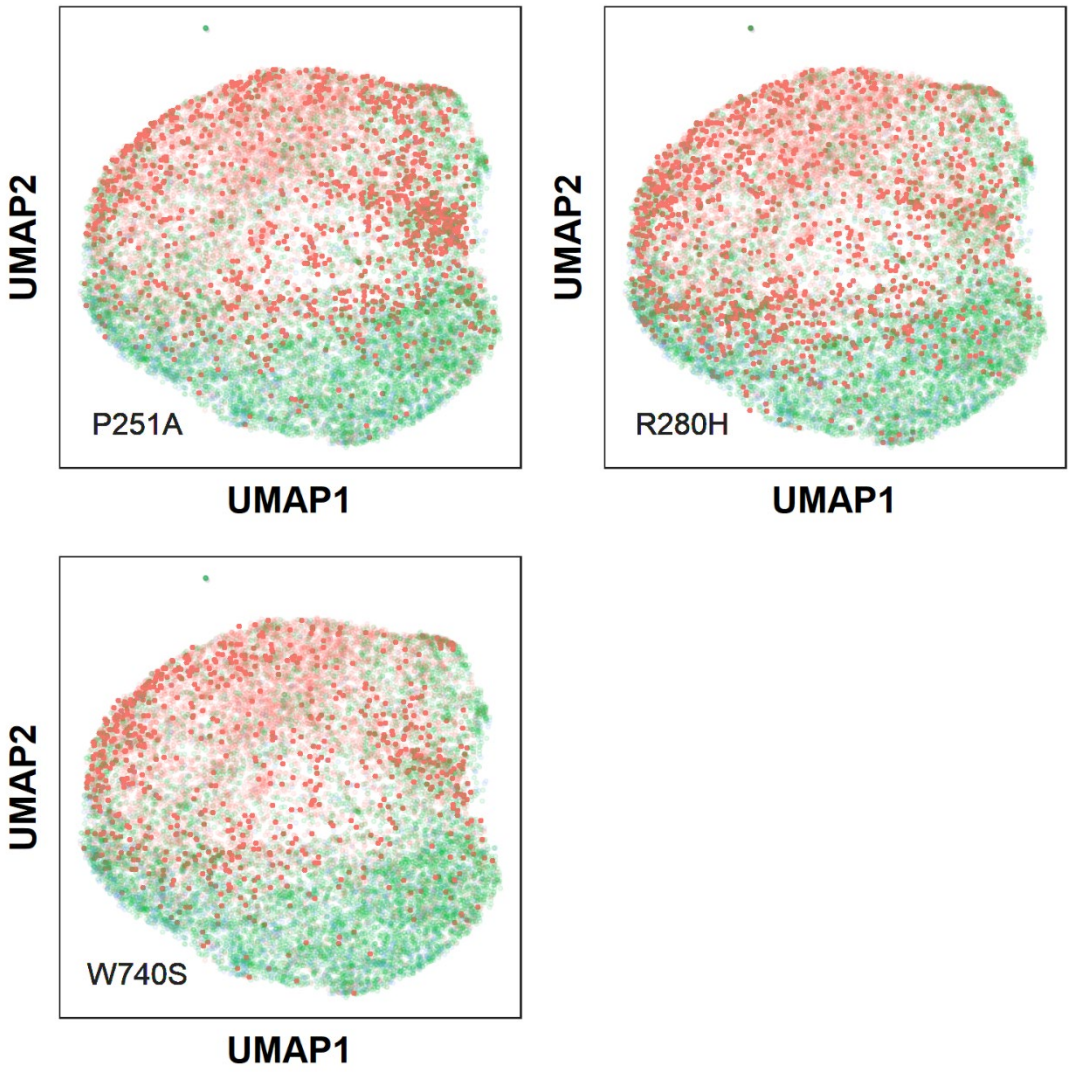


Figure S2: Scatter plot for cell feature data UMAPs, highlighted by cell genotype. U2OS WT refers to cells uninfected with lentivirus, while *MFN2* WT refers to cells infected with lentivirus containing the wild type *MFN2* gene. Otherwise, the highlighted genotype refers to the mutation present in the introduced *MFN2* gene.

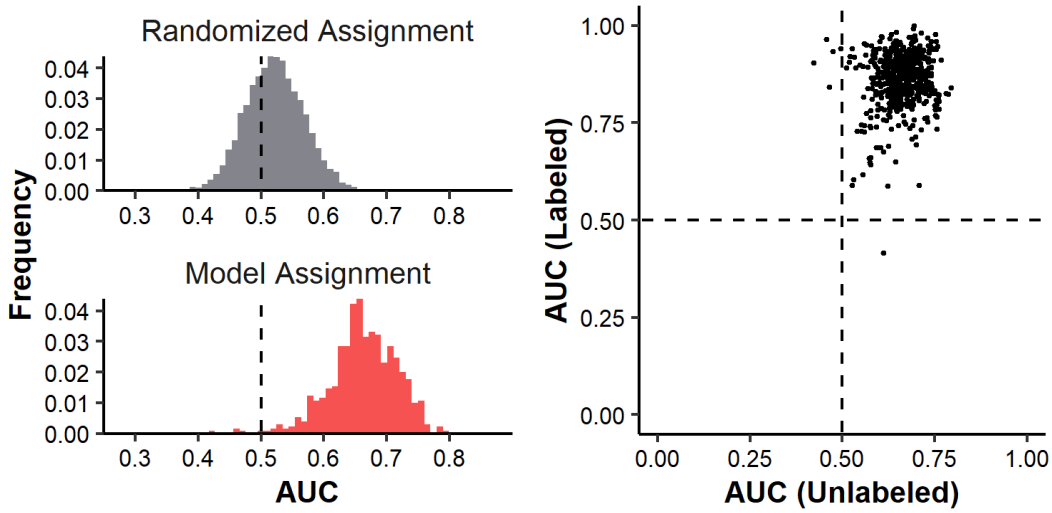


Figure S3 Model AUCs for identifying a mixture of wild type and six pathogenic mutant cell lines: a: A scatterplot showing the performance of 610 models in detecting mutants in a mixture of wild type cells and four pathogenic mutants. Each point represents a model and its position is determined by its ability to distinguish cells in the pure control wells and cells in the mixed wells. The vertical and horizontal lines represent the AUCs of the model that was used to choose cells for isolation. **b**: A histogram of AUCs for models detecting mutants in the unlabeled population of wild-type cells and mutants. On top, a histogram of AUCs generated from randomly assigning models is shown as a comparison.

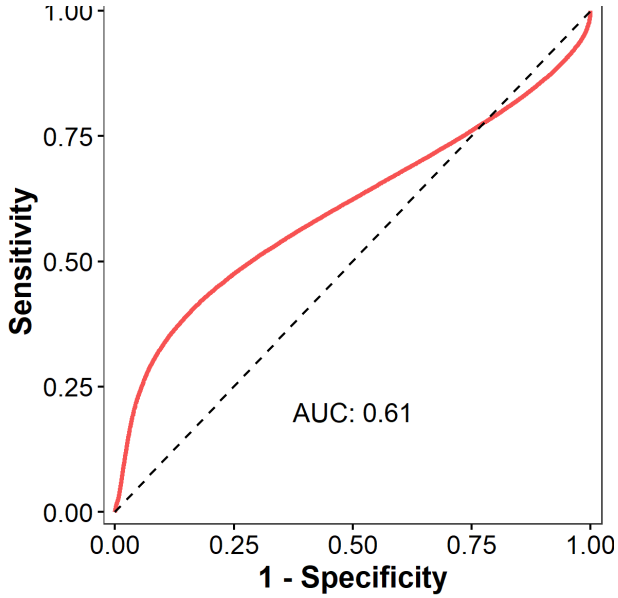


Figure S4 Modeled ROC from flow cytometry: ROC Curve based on flow data for *MFN2* WT vs. *MFN2* R280H mutants on the PerCP-Cy5.5-A-Compensated channel

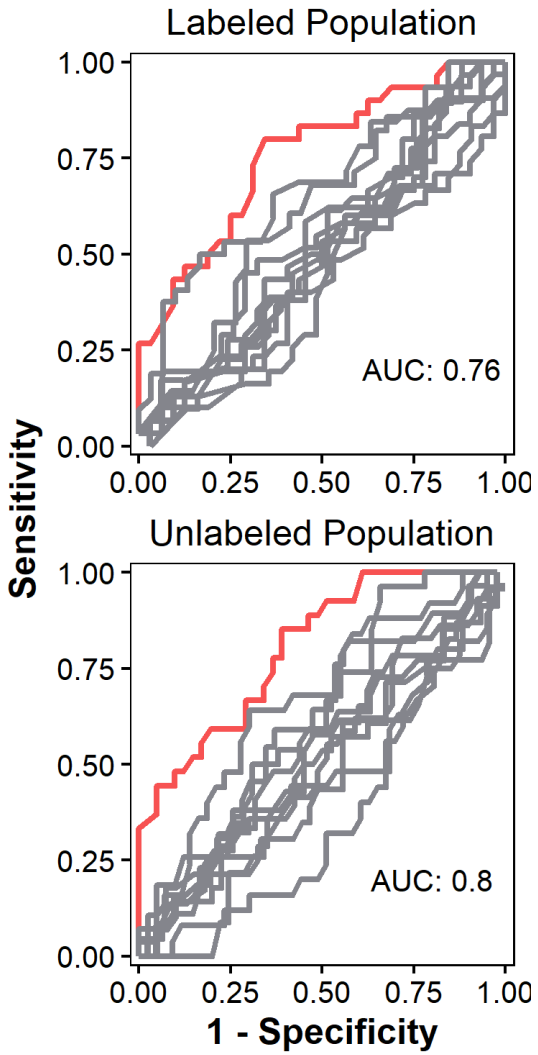


Figure S5 Results from Distinguishing P251A Pathogenic vs. D221 Benign: ROC Curves generated using the best model separated by data generated from cells picked from wells in the labeled cell population (upper) and from the unlabeled cell population (lower). The red curves are the curves for the models in question and the grey curves are those for control models generated by random shuffling of labels

Feature	Stain
Mito Total Area	Mitotracker Red
Nuclei Low Grey Level Run Emphasis	Mitotracker Red

Mito Intensity	Mitotracker Red
Cells Entropy	Mitotracker Red
Nuclei Intensity CV	Mitotracker Red
Cells Skewness	Mitotracker Red
Cells Total Intensity (Cyto)	Mitotracker Red
Cells Intensity Spreading	Mitotracker Red

Table S1: Table of the 8 features used for cell isolation and random feature assembly with the corresponding stain.

Acknowledgements

We would like to thank Kate Matsunaga, Will Lee, and Silas Kuang for their help in the laboratory. We would also like to thank the Milbrandt lab for their continued support and equipment use. Genome Engineering and iPSC Center at Washington University School of Medicine also provides invaluable support to all members of our laboratory. We also thank the Vindigni lab for allowing us to utilize the engineered lines. Finally, we would like to thank the genetics administration and our cleaning staff.

Author Contributions

ALY, JCB, and WJB wrote the manuscript. WJB and JCB designed the experiments. Initial planning and ideas brought by WJB, JDM, RDM. JEW, CLK, MJL, XEX, MAV performed the experiments. JEW, CLK, MJL, XEX, JCB, ALY, and WJB analyzed results. VC wrote some of the key software. All reviewed the paper and gave suggestions.

References

1. Guigo, R. & de Hoon, M. Recent advances in functional genome analysis [version 1; referees: 2 approved]. *F1000Research* **7**, (2018).
2. Weile, J. & Roth, F. P. Multiplexed assays of variant effects contribute to a growing genotype–phenotype atlas. *Human Genetics* (2018). doi:10.1007/s00439-018-1916-x
3. Esposito, D. *et al.* MaveDB: An open-source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome Biol.* (2019). doi:10.1186/s13059-019-1845-6
4. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647–661 (2014).
5. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
6. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020).
7. Reilly, S. K. *et al.* HCR-FlowFISH: A flexible CRISPR screening method to identify cis-regulatory elements and their target genes. *bioRxiv* 2020.05.11.078675 (2020). doi:10.1101/2020.05.11.078675
8. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).
9. Klann, T. S. *et al.* CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat. Biotechnol.* **35**, 561–568 (2017).
10. Buchser, W. J., Slepak, T. I., Gutierrez-Arenas, O., Bixby, J. L. & Lemmon, V. P. Kinase/phosphatase overexpression reveals pathways regulating hippocampal neuron morphology. *Mol. Syst. Biol.* **6**, 391 (2010).
11. Gerdts, J., Sasaki, Y., Vohra, B., Marasa, J. & Milbrandt, J. Image-based screening identifies novel roles for IkappaB kinase and glycogen synthase kinase 3 in axonal degeneration. *J. Biol. Chem.* **286**, 28011–8 (2011).
12. Caicedo, J. C., Singh, S. & Carpenter, A. E. Applications in image-based profiling of perturbations. *Curr. Opin. Biotechnol.* **39**, 134–142 (2016).
13. Pegoraro, G. & Misteli, T. High-Throughput Imaging for the Discovery of Cellular Mechanisms of Disease. *Trends Genet.* **33**, 604–615 (2017).
14. Lalli, M. A., Avey, D., Dougherty, J. D., Milbrandt, J. & Mitra, R. D. High-throughput single-cell functional elucidation of neurodevelopmental disease-associated genes reveals convergent mechanisms altering neuronal differentiation. *Genome Res.* **30**, 1317–1331 (2020).
15. Jaeger, P. A. *et al.* Systematic Gene-to-Phenotype Arrays: A High-Throughput Technique for Molecular Phenotyping. *Mol. Cell* **69**, 321-333.e3 (2018).
16. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787-799.e17 (2019).
17. Groot, R., Lüthi, J., Lindsay, H., Holtackers, R. & Pelkmans, L. Large-scale image-based profiling of single-cell phenotypes in arrayed CRISPR-Cas9 gene perturbation screens. *Mol. Syst. Biol.* **14**, (2018).
18. Simm, J. *et al.* Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery. *Cell Chem. Biol.* **25**, 611-618.e3 (2018).
19. Bray, M. A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).
20. Kang, Z. B. *et al.* Fluopack screening platform for unbiased cellular phenotype profiling. *Sci. Rep.* **10**, 1–6 (2020).

21. Gudla, P. R., Nakayama, K., Pegoraro, G. & Misteli, T. SpotLearn: Convolutional Neural Network for Detection of Fluorescence In Situ Hybridization (FISH) Signals in High-Throughput Imaging Approaches. *Cold Spring Harb. Symp. Quant. Biol.* **82**, 57–70 (2017).
22. Hasle, N. *et al.* High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Mol. Syst. Biol.* **16**, 856476 (2020).
23. Lee, J. *et al.* Versatile phenotype-activated cell sorting. *Sci. Adv.* **6**, eabb7438 (2020).
24. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787-799.e17 (2019).
25. Kanfer, G. *et al.* Image-based pooled whole-genome CRISPRi screening for subcellular phenotypes. *J. Cell Biol.* **220**, (2021).
26. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst.* **3**, 385-394.e3 (2016).
27. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
28. Payne, A. C. *et al.* In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* (2020). doi:10.1126/science.aay3446
29. Rappez, L. *et al.* Spatial single-cell profiling of intracellular metabolomes in situ 1. doi:10.1101/510222
30. Yan, X. *et al.* High-content imaging-based pooled CRISPR screens in mammalian cells. *J. Cell Biol.* **220**, (2021).
31. Attayek, P. J. *et al.* Automated microraft platform to identify and collect non-adherent cells successfully gene-edited with CRISPR-Cas9. *Biosens. Bioelectron.* **91**, 175–182 (2017).
32. Welch, J. D. *et al.* Selective single cell isolation for genomics using microraft arrays. *Nucleic Acids Res.* **44**, 8292–301 (2016).
33. Wheeler, E. C. *et al.* Pooled CRISPR screens with imaging on microraft arrays reveals stress granule-regulatory factors. *Nat. Methods* **17**, 636–642 (2020).
34. Youle, R. J. & Van Der Bliek, A. M. Mitochondrial fission, fusion, and stress. *Science* **337**, 1062–1065 (2012).
35. Pareyson, D., Saveri, P. & Pisciotta, C. New developments in Charcot-Marie-Tooth neuropathy and related diseases. *Curr. Opin. Neurol.* **30**, 471–480 (2017).
36. Stuppia, G. *et al.* MFN2-related neuropathies: Clinical features, molecular pathogenesis and therapeutic perspectives. *J. Neurol. Sci.* **356**, 7–18 (2015).
37. Baloh, R. H., Schmidt, R. E., Pestronk, A. & Milbrandt, J. Altered axonal mitochondrial transport in the pathogenesis of Charcot-Marie-Tooth disease from mitofusin 2 mutations. *J. Neurosci.* **27**, 422–30 (2007).
38. Landrum, M. J. *et al.* ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
39. Perry, S. W., Norman, J. P., Barbieri, J., Brown, E. B. & Gelbard, H. A. Mitochondrial membrane potential probes and the proton gradient: A practical usage guide. *BioTechniques* **50**, 98–115 (2011).
40. Tirman, S., Cybulla, E., Quinet, A., Meroni, A. & Vindigni, A. PRIMPOL ready, set, reprime! *Critical Reviews in Biochemistry and Molecular Biology* **56**, 17–30 (2021).
41. Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741–746 (2010).
42. Ihry, R. J. *et al.* P53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* (2018). doi:10.1038/s41591-018-0050-6