Label-free Identification of Protein Aggregates Using Deep Learning

Khalid A. Ibrahim[1,2], Kristin S. Grußmayer[3,*], Nathan Riguet[2], Lely Feletti[1], Hilal A. Lashuel[2,*] and Aleksandra Radenovic[1,*]

[1]Laboratory of Nanoscale Biology, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[2]Laboratory of Molecular and Chemical Biology of Neurodegeneration, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[3]Department of Bionanoscience and Kavli Institute of Nanoscience Delft, Delft University of Technology, Delft, Netherlands

**Correspondence: k.s.grussmayer@tudelft.nl, hilal.lashuel@epfl.ch, aleksandra.radenovic@epfl.ch*

Abstract

Protein misfolding and aggregation play central roles in the pathogenesis of various neurodegenerative diseases (NDDs), including Huntington's disease, which is caused by a genetic mutation that leads to a polyglutamine repeat length > 35 in exon 1 of the Huntingtin protein (Httex1). Current research on protein aggregation often involves the use of fluorescent labels to visualize and monitor the dynamics of protein expression, which can alter the biophysical properties of proteins and the final ultrastructure, composition, and toxic properties of the formed aggregates. To overcome this limitation, we present a method for label-free identification of NDD-associated aggregates (LINA). Our approach utilizes deep learning to detect unlabeled and unaltered Httex1 aggregates in living cells from quantitative phase images, without the need for fluorescent labeling. We developed pixel-classification and pixel-regression models and validated them on aggregates formed by different constructs of Httex1. Our results reveal that Httex1 proteins with shorter polyglutamine repeat lengths form aggregates with a higher average dry mass and area, highlighting the differences in their ultrastructure and aggregation mechanisms. LINA enables the dynamic identification of label-free aggregates and measurement of their dry mass and area changes during their growth process. Our models offer high speed, specificity, and simplicity to analyze label-free protein aggregation dynamics and obtain high-fidelity information.

Introduction

Neurodegenerative diseases (NDDs), such as Alzheimer's disease (AD), Parkinson's disease (PD), and Huntington's disease (HD), are major global health concerns afflicting tens of millions of patients worldwide and are currently incurable[1,2]. They are characterized by progressive damage to the structure and function of neurons and the accumulation of misfolded proteins in specific brain regions and neuronal populations[3–6]. Although these misfolded protein aggregates represent the hallmark of several NDDs, including AD, PD, and HD, our understanding of their mechanisms of formation and role in the disease pathogenesis remains incomplete. While some studies suggest that their formation is a major driver of neurodegeneration, other studies propose that they represent neuroprotective mechanisms that lead to inactivation of toxic aggregates[7–12]. Furthermore, the nature of toxic species remains unknown. One primary reason for this knowledge gap is the lack of tools and methods that allow for direct monitoring of the different phases of protein misfolding and aggregation, leading to the formation of the pathological hallmarks found in the brains of affected individuals.

To enable monitoring protein aggregation in living cells or animals, fluorescent proteins (FPs), such as GFP, YFP, mCherry, and others, are usually fused to the C- or N-terminus of the protein of interest. Protein aggregation is usually seen as the transition from diffuse signal to the formation of puncta or foci structures. However, as useful and valuable as FPs can be, they also have limitations. Several studies have shown that fluorescently-tagged proteins exhibit altered biochemical, biophysical or cellular properties[13–22]. In the context of protein aggregation linked to NDDs, the addition of fluorescent proteins to proteins of different sizes has been shown to alter not only the kinetics of aggregation and the final size of the aggregates, but also the interactome of the aggregate and their ultrastructural organization in the final inclusions that accumulate in cells or brain tissues. This discrepancy has been observed in several NDD-associated proteins, including alpha-synuclein[18,19], tau[20], amyloid beta[21,22] (Aβ), and exon1 of the Huntingtin (Htt) protein[14–17] (Httex1), one of the primary components of intracellular protein aggregates found in Huntington's disease post-mortem brains[23–25]. Httex1 overexpression in multiple cellular and animal models of HD recapitulates many of the

key features of HD human pathology, including Htt aggregation and inclusion formation, neurodegeneration, and brain atrophy[8,26–30].

In a recent study[14], we compared the biochemical and ultrastructural properties of cellular inclusions formed by mutant Httex1, with expanded polyglutamine (polyQ) repeats of 16-72Q, in the absence or presence of GFP on the C-terminus of the protein. Our results showed that the addition of GFP to mutant Httex1 not only altered the biochemical and ultrastructural properties of the inclusions but also their toxic properties and interactions with cellular organelles. For example, the inclusions formed by unlabeled mutant Httex1 displayed distinctive core and shell structures that contain lipids and membranous organelle and within which the Htt fibrils colocalize with and interact with the endoplasmic reticulum (ER) structure and membranous structures. In contrast, mutant Httex1-GFP formed a mesh of cytoplasmic fibrils that exhibited minimal interactions with cellular organelles (Fig. 1a). These observations are consistent with previous studies showing that fusion of large proteins to mutant Httex1, such as GFP, alter their dimensions as well as their mechanical and surface properties. GFP-labeled fibrils were approximately 3 nm thicker[15], and exhibited increased stiffness[16] compared to the unlabeled Httex1 fibrils. Using coarse-grained simulations, we recently showed that fluorescent tags like GFP cause a size-dependent surface occlusion and thus alter the fibril's interactome[17]. These observations could explain the distinct interactome observed for untagged and GFP-tagged mutant Htt inclusions in mammalian cells and primary neurons[14]. While it is tempting to attribute these changes in the biophysical and biochemical properties to the large size of fluorescent proteins, it has been shown that even the fusion of small peptide tags like HA to different N-terminal Htt fragments could markedly change their aggregation and toxic properties[31]. Altogether, these observations underscore the critical importance of developing label-free methods that enable investigation of the mechanisms of protein aggregation and inclusion formation of the native proteins.

Although recent studies have suggested using label-free methods, such as Raman microscopy[32], to monitor protein aggregation in cells, these methods lack the specificity and contrast needed to analyze them sufficiently, often lack the needed temporal resolution, or require deuterium labeling. Quantitative phase imaging (QPI)[33–35] is an

attractive label-free microscopy modality that produces a quantitative image of an unlabeled biological specimen based on the phase shift at each pixel in the field of view (FOV), enabling the extraction of quantitative parameters, such as the dry mass or morphology. Since low-intensity illumination can be used, quantitative information can be obtained from live specimens with minimal photodamage and photobleaching (compared to fluorescence imaging) at rapid time scales. However, the lack of a specific label complicates distinguishing between different structures inside cells. In recent years, the concept of virtual staining has emerged[36–39], in which a deep learning model is taught to map between a label-free imaging modality (e.g., brightfield, QPI, or auto-fluorescence) and fluorescence imaging, such that the model can predict the fluorescence signal strictly from the label-free input, yielding the necessary contrast and specificity. A recent technique called VISTA uses sample expansion to achieve super-resolved Raman imaging and virtual labeling of Aβ plaques and mutant Htt aggregates[40]. However, this technique was not validated on label-free Htt aggregates, was not quantitatively validated on any Htt aggregates, had a large variation in Aβ-plaque model performance (roughly between 30-90%), requires substantial sample manipulation and perturbation during expansion, and is fundamentally incompatible with live-cell imaging.

In this work, we employ the concept of virtual labeling[36–39] on a well-characterized cellular model of mutant Httex1 aggregation and inclusion formation by training a neural network on data collected using widely available imaging modalities (brightfield/QPI and widefield fluorescence) to generate label-free identification of NDD-associated aggregates (LINA) models (Fig. 1b). We quantitatively validate the LINA models and demonstrate that they can be used to accurately identify unlabeled mutant Httex1 aggregates with areas as small as 3 μm^2 and at exposure times as low as 3 ms. By applying our neural network models to identify aggregates formed by different label-free mutant Httex1 constructs, we are able to measure and compare their dry masses. We can also identify aggregates from live-cell imaging data and are able to measure the dry mass and area of aggregates as they form. We experimentally validate the models' robustness to different acquisition settings and cell lines, showing the high applicability of the method. The aforementioned issues related to labeling illustrate the need for label-free techniques for identifying and analyzing native neurodegenerative protein

aggregates to avoid tag-induced alterations. The label-free method we have developed to visualize and analyze protein aggregates demonstrates, for the first time to our knowledge, the applicability of virtual labeling on unlabeled and unaltered protein aggregates in living cells, with rapid speed, high accuracy, and relative simplicity. LINA paves the way for more accurate analysis of NDD-associated protein aggregates, better recapitulating their true neurobiological nature and offering higher-fidelity information.

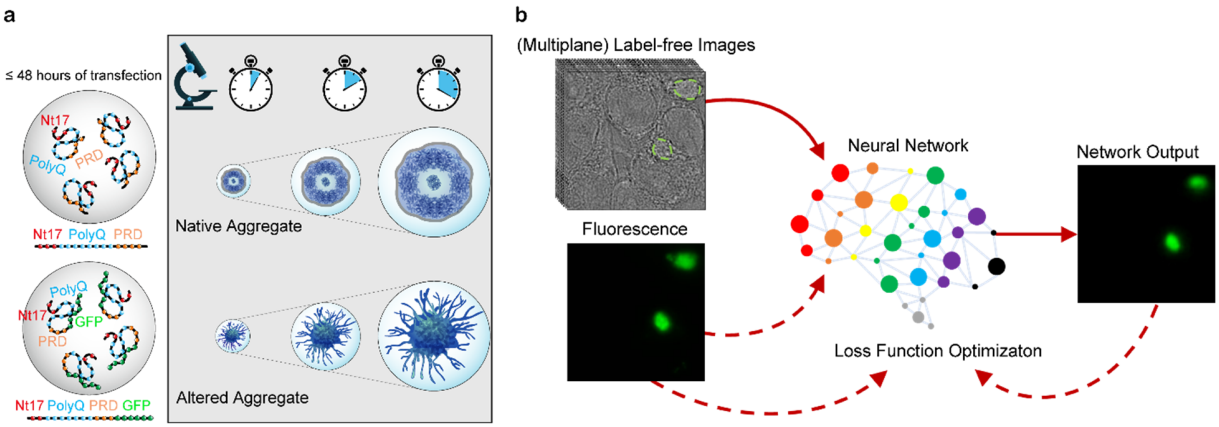

Figure 1: Label-free identification of NDD-associated aggregates (LINA). (a) The NDD-associated Httex1 protein forms aggregates in cells within 48 hours. When the protein is unlabeled, these aggregates have a core and shell ultrastructure. Labeled Httex1 (e.g., with GFP) forms altered aggregates that lack this structure, instead resembling a mesh of fibrils, and have altered biochemical and biophysical properties (e.g., different proteome composition, stiffness, and fibril length). (b) To enable label-free imaging of unaltered Httex1 aggregates, we trained a neural network to map between label-free (brightfield or quantitative phase) image inputs (single or multiple planes) and fluorescence images, such that the network is then able to identify aggregates using only the label-free input. Dashed arrows represent training-only steps.

140 # Results

141 **Label-free identification of aggregates formed by different Httex1 constructs with**
142 **high accuracy**

143 The choice of a well-controlled biological model system is crucial for the development of
144 robust and reproducible image analysis tools. To apply virtual labeling on NDD-
145 associated aggregates and develop LINA models, we chose to use a cellular model of
146 HD that is well-characterized. As reported previously, when a mutant Httex1 construct
147 with a polyQ repeat length ≥ 39 is overexpressed in HEK 293 (HEK) cells, aggregates are
148 formed within a period of 48 hours. With our custom-built, multi-modal, multi-plane
149 microscope[41], it is possible to acquire ultrafast 4D brightfield and fluorescence images
150 using an image-splitting prism that introduces path length differences in the detection
151 path. This prism allows the acquisition of an image stack consisting of 8 z-planes
152 simultaneously. An algorithm is then used to retrieve phase information from the
153 brightfield image stack using Fourier filtering, transforming it into a QPI stack. We used
154 this microscope to collect a dataset of over 1,000 8-plane brightfield and fluorescence
155 pairs of images of fixed HEK cells overexpressing a mutant Httex1 construct with a polyQ
156 repeat length of 72, fused to GFP (Httex1-72Q-GFP).

157 We processed the dataset to obtain pixel-registered 8-plane QPI (which we also
158 refer to as 'phase' images) and fluorescence images. We then used a subset of the
159 dataset (the training and validation sets) to train a convolutional neural network (CNN)
160 (with a U-Net architecture[42], more information is provided in the Methods) and produce
161 LINA models for both pixel regression and pixel classification (Fig. 2a). Here, each ground
162 truth label is generated as the maximum z-projection of each 8-plane image stack so that
163 it incorporates information from the 8 planes, reducing the complexity of the mapping.
164 The pixel classification model is trained on segmentation masks generated from these
165 maximum z-projection images, and it classifies the pixels into either 1's or 0's, i.e., part of
166 an aggregate or not. The pixel regression model is more general in that it is trained directly
167 on the maximum z-projection fluorescence images, and can predict intensity values rather
168 than just 1's or 0's. In Fig. 2a, the phase image is a color-coded maximum z-projection,
169 where each z-plane is represented using a different color, illustrating the spread of

170 information available in the 8 planes. The network outputs were observed to match the
171 ground truths quite well, as can be verified further in Fig. 2b.

172 To precisely quantify how well the LINA pixel-regression model performs, we
173 measured two standard metrics: the structural similarity index (SSIM) and the Pearson
174 correlation coefficient (r), between the network output and the ground truth (Fig. 2c). Both
175 metrics were consistently high (~96% mean values) over the entire test set, a set of
176 images which the model had never seen before, illustrating the high reliability of the
177 model. The pixel-classification model also performs well, achieving an F1-score of 0.9
178 and a mean Jaccard index of 0.78. Further visual test-set examples for both model types
179 are shown in Supplementary Fig. 1 and Supplementary Fig. 2.

180 To determine if the models work on the unlabeled protein aggregates and if they
181 generalize to other constructs than the one it was trained on, we sought to validate the
182 network on different label-free constructs of mutant Httex1. The constructs we tested are
183 unlabeled mutant Httex1 with a polyQ repeat length of 39 (Httex1-39Q), unlabelled mutant
184 Httex1 with a polyQ repeat length of 72 (Httex1-72Q), and unlabelled mutant Httex1 with
185 a polyQ repeat length of 72 with a truncated Nt17 domain (Httex1-ΔNt17-72Q). These
186 constructs, which either have a shorter polyQ repeat length or have a truncated Nt17
187 domain, have been shown to influence the kinetics of aggregation, as well as the
188 ultrastructural properties of Httex1 inclusions when comparing 39Q and 72Q[14]. Therefore,
189 it was important to verify if LINA can identify aggregates formed using these constructs
190 despite the differences in aggregate properties. To obtain a ground truth, we performed
191 immunocytochemistry (ICC) on the three constructs and labeled the aggregates with an
192 anti-Htt antibody raised against the Proline-rich domain (PRD) (MAB5492, Millipore). We
193 acquired QPI images of the same cells before and after ICC, using grid-marked
194 coverslips, to ensure that LINA works well in both cases. Fig. 2d shows the input, ground
195 truth and network output for each construct. For all three constructs, the network correctly
196 identifies the aggregates, whose peripheries are shown in the ground truth images; this
197 is because the antibody we used is known to show a strong immunoreactivity to the
198 periphery of the aggregate rather than the core[14], while the network was trained solely on
199 GFP-labeled protein aggregates. Although the CNN was only trained on images of
200 Httex1-72Q-GFP, the model is nonetheless generalizable to different kinds of label-free

Httex1 constructs. For Httex1-ΔNt17-72Q, the model is able to discern between rather small closely-neighboring aggregates (areas of around 3 μm^2), instead of seeing them as one larger aggregate.

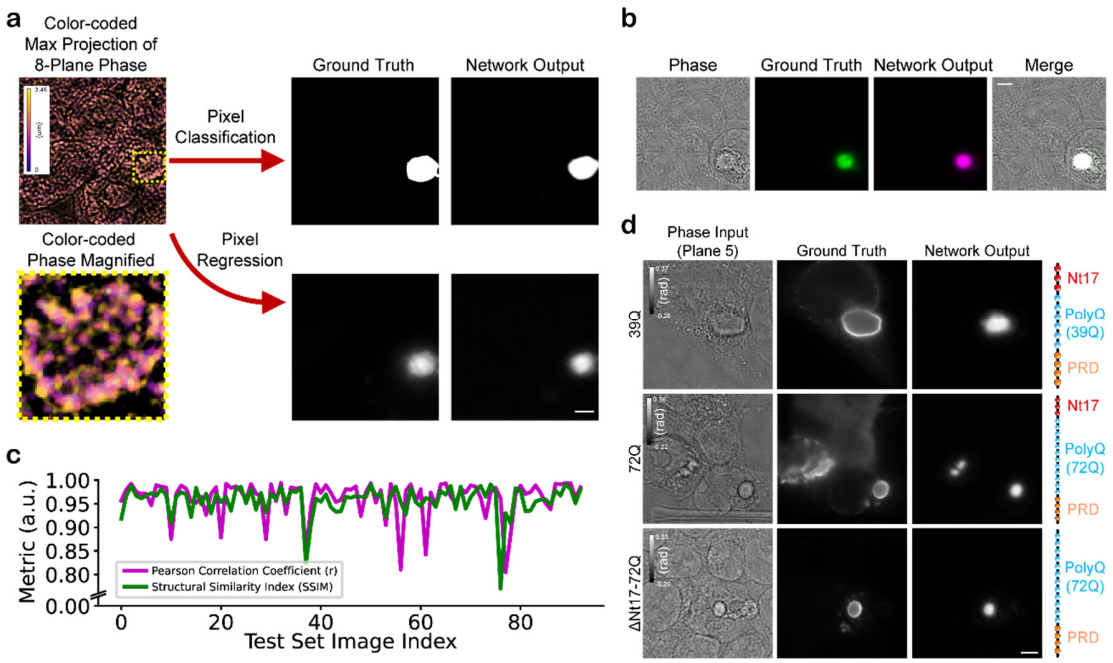

Figure 2: Validation of deep learning models for label-free identification of Httex1 protein aggregates. (a) Convolutional neural networks for both pixel classification and regression have been trained using pixel-registered pairs of 8-plane quantitative phase images and maximum-projected 8-plane fluorescence images or corresponding segmented masks. The 8-plane input images are represented by a color-coded maximum *z*-projection image; the aggregate in the image is highlighted by the dashed square and is shown magnified. The phase signal was thresholded (T = 0 rad) prior to color-coding to enhance the contrast. The labels used are generated as the maximum z-projection of the fluorescence images in the 8 planes for the pixel regression network and the corresponding segmented masks (Otsu thresholding) for pixel classification. (b) Test set prediction example, with images of the phase, fluorescence, network output, and a merge of all three, showing where the three images colocalize in white. (c) The pixel regression network was validated quantitatively on the test set using two different standard metrics, and both corresponded well with the ground truth (mean ~0.96). (d) The regression network was further validated on label-free Httex1 aggregates. Despite being trained only on a 72Q-GFP construct of Httex1, the model works well on label-free constructs of different polyQ repeat lengths (39Q, 72Q) and types (72Q with a truncated Nt17 domain). Scale bars: 5 μm.

Robustness and generalizability of LINA

Testing the robustness of LINA in various image acquisition, input type, label quality, and cell line configurations is vital to ensure the models' robustness and generalizability to different kinds of data, and to facilitate the usability of the models by new users.

Supplementary Note 1 provides a more extensive discussion of the robustness and generalizability of LINA. We observed, in test-set examples that are being seen by the network for the first time, that the pixel-classification model can identify aggregates that are missing in the label due to segmentation errors (Supplementary Fig. 3a). The model outputs outperform direct segmentation from the ground truth, showing that it can learn very well what constitutes an aggregate despite some erroneous labels.

The signal-to-noise ratio (SNR) of microscopy images is a key determinant of how well image analysis pipelines and models perform. Increasing amounts of noise can reduce the accuracy and precision of segmentation and classification models, leading to incorrect or inconsistent results. All our models were trained on images taken at a 50 ms exposure time and a fixed illumination, so we wanted to test whether it is possible to identify aggregates from images taken at lower SNRs. We verified LINA's robustness to noise both using simulated noise and experimentally (Supplementary Fig. 3b-f). Through a simple pre-filtering step, we show that it is possible to identify aggregates even at excessive amounts of noise (variance of 0.48 rad), where nothing can be visually identified in the image anymore. We then experimentally verified the model's robustness to different SNRs, as we measured a consistently-high SSIM at various exposure times, as low as 3 ms, without the need for pre-filtering, despite the model being solely trained on images acquired at a 50 ms exposure time.

Another aspect we wanted to assess was the models' generalizability to different cell lines than what we used for training (HEK). Therefore, we quantified LINA's ability to identify aggregates expressed in HeLa cells, and the model consistently produced highly-accurate predictions, as quantified by measuring the SSIM between the network outputs and the ground truth images (Supplementary Fig. 3g). We also evaluated LINA's dependence on the number of planes of the input by training new pixel-regression models using QPI inputs with one, two and four planes, respectively. Supplementary Fig. 4a quantifies and compares the performance of the 8-plane model to these three new models. We found that all three new configurations perform very well. We expected to possibly need transfer learning using a small set of new data to be able to identify aggregates correctly at much lower SNRs or for different cell lines, however, we found

254 LINA to be highly robust to these changes, promising high generalizability to different
255 kinds of imaging conditions.

256 Our method to produce quantitative phase images[41] is widely adoptable; as
257 previously mentioned, the phase information can be extracted using an algorithm from a
258 z-stack of images. However, the simplest and most basic label-free technique is
259 brightfield microscopy. How well would LINA perform on the simplest and most widely-
260 available label-free technique? To answer this question, we trained pixel-regression
261 models using only brightfield inputs, first with 8 planes then with a single plane
262 (Supplementary Fig. 4b). We quantitatively compared both these models' performance
263 with the model trained on 8-plane QPI images, and we found that both brightfield models
264 perform very well on the test set. The 8-plane brightfield model yielded very similar results
265 to the QPI model, with a slightly larger variance and the drawback of not having the
266 possibility to extract the dry mass from the images. To confirm LINA's generalizability and
267 ease of use, we tested it on a completely different imaging setup with different illumination
268 and a lower-numerical-aperture microscope objective (Supplementary Fig. 4c). We were
269 able to identify aggregates from images acquired using this different setup and the 1-
270 plane brightfield LINA model, without the need to rely on transfer learning. We, therefore,
271 expect LINA to work well for new users acquiring images at different conditions, perhaps
272 needing some transfer learning, or as shown here, possibly not even requiring it.

273 **Automatic scanning and image acquisition**

274 We enhanced our image acquisition throughput by implementing an automatic image-
275 scanning method, called xy-scan mode, to scan an arbitrary number of fields of view
276 (FOV) across the sample (Fig. 3a). The xy-scan mode permits us to set the step size and
277 the number of steps in the x and y directions. The illumination source remains stationary
278 while a piezo nano-positioning stage moves under the control of the software. This allows
279 us to acquire images of thousands of FOVs, for cells expressing different Httex1
280 constructs, which our LINA models can then process to determine the presence or
281 absence of aggregates (Fig. 3b).

282 Using the automatic xy-scan mode, we efficiently collected images of cells
283 overexpressing mutant Httex1-39Q, Httex1-72Q, Httex1-ΔNt17-72Q and Httex1-72Q-

GFP, then processed them to determine which FOVs contain aggregates. To minimize the occurrence of false positives, we combined the predictions from multiple models. While this approach may result in a higher chance of false negatives, our priority is to ensure for the analysis results to consider only true aggregates. The output of our model, which indicates the label-free images that contain aggregates, can then be used to extract their dry masses.

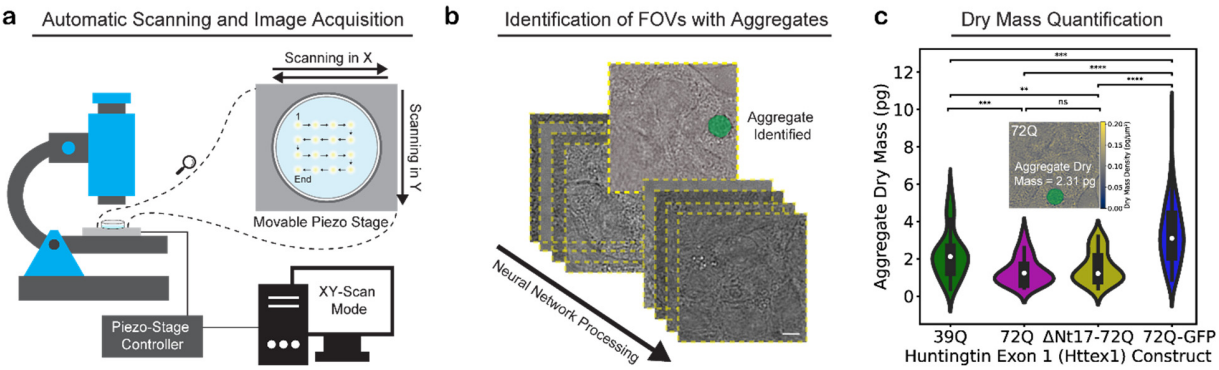

Figure 3: Automatic image acquisition, identification, and dry mass quantification of different kinds of label-free Httex1 aggregates. (a) A motorized piezo nano-positioning stage is controlled using software to scan the sample in the x and y directions and collect images at various fields of view (FOVs) for different label-free constructs of Httex1 (39Q, 72Q, ΔNt17-72Q) and a GFP-labeled construct (72Q-GFP). (b) The images are post-processed and our trained model is used to identify FOVs that contain aggregates. (c) The dry mass of aggregates produced by different constructs of Httex1 is extracted from the quantitative phase images, as shown in the inset image. ΔNt17-72Q and 72Q had similar dry mass distributions, however, both label-free constructs had smaller dry masses on average than 39Q, and all three had smaller dry masses than the labeled 72Q-GFP. Scale bars: 5 μm.

Dry mass quantification and comparison between aggregates of different Httex1 constructs

The dry mass of a biological specimen can be extracted from a QPI image through a proportional relation (Methods)[35]. For this relation, we used a constant value of 0.19 for the refractive increment α, as recommended in the literature, due to α having an extremely narrow distribution[35,43]. The quality of the QPI images, e.g., in terms of resolution or SNR, determines the accuracy and precision of the dry mass measurement[35]. After the QPI image is translated into a dry mass density map, the dry mass of an aggregate can be measured by integrating over its area in the image (more details in Supplementary Fig.

309 5). To compare the dry mass of the different Httex1 constructs, we used the model outputs
310 from the automatic xy-scan images to curate images with aggregates, producing
311 segmentation masks that were then multiplied by the corresponding dry mass images.
312 This allows for efficient and accurate measurement of the dry mass of the different kinds
313 of aggregates (Fig. 3c).

314 Our findings indicate that there is no statistically-significant difference between the
315 mean dry masses of mutant Httex1-72Q and Httex1-ΔNt17-72Q. Both of these constructs
316 have been shown to form inclusions with the same core-and-shell ultrastructure[14]. Thus,
317 they likely have a similar mechanism of aggregation, which explains their similar dry
318 masses. We observed that fewer aggregates were detected for Httex1-39Q (8% of
319 automatically-scanned FOVs) compared to the other two constructs, which had similar
320 rates (~12% of automatically-scanned FOVs). This is in agreement with the literature, as
321 increasing polyQ repeat length is known to accelerate aggregation and toxicity[44–46].
322 However, interestingly, we found that Httex1-39Q, which lacks the core-and-shell
323 ultrastructural arrangement[14], produced aggregates with a larger average area (16.9 ±
324 7.6 μm^2), compared to Httex1-72Q (11.5 ± 4.7 μm^2) and Httex1-ΔNt17-72Q (12.7 ± 4.9
325 μm^2), and consequently a larger average dry mass. The difference in ultrastructure, and
326 therefore the aggregation mechanism, leads to fewer aggregates being produced overall,
327 yet at the same time, a fraction of the aggregates that are produced grow to larger final
328 sizes.

329 We also compared the mean dry masses of the three unlabeled constructs with
330 that of the labeled one (Httex1-72Q-GFP) and found around a two-fold increase in the
331 mean dry mass of the labeled construct compared to the two constructs with the same
332 polyQ repeat length, as well as a 1.5-fold increase compared to Httex1-39Q. Similar to
333 Httex-39Q, Httex1-72Q-GFP aggregates have been shown to lack the core-and-shell
334 ultrastructural arrangement[14], which for both constructs leads to increasing the size of the
335 formed aggregates. The considerable increase in size of Httex1-72Q-GFP aggregates is
336 likely due to a mixture of the large size of the GFP tag and its effect on the size of the
337 fibrils, the difference in ultrastructure, and the difference in properties (e.g., proteome
338 composition) that it is known to cause, supporting previous studies' results and
339 demonstrating why such label-free methods are needed.

340 **Live, label-free identification and analysis of an aggregate as it forms**

341 Capturing the growth dynamics of NDD-associated protein aggregation is crucial to
342 deciphering the role of different stages and intermediates on the pathology formation
343 pathways in the pathogenesis of NDDs and potentially identifying new targets for
344 therapeutic intervention. For example, if the aggregate growth rate is controlled by a
345 particular cellular mechanism, drugs that target this mechanism may slow or stop the
346 disease development and progression. Furthermore, a better understanding of the
347 characteristics of different aggregates on the pathway to pathological inclusions should
348 facilitate the development of new molecular probes and diagnostic tools for monitoring
349 early oligomerization events that are thought to contribute to and correlate with early
350 stages of disease development. LINA enables the live identification, imaging, and
351 analysis of native aggregates as an important step toward reaching these goals.

352 Figure 4a shows our LINA pixel-regression model's performance on dynamic, live-
353 cell, time-lapse images, taken every 2 minutes, of Httex1-72Q-GFP in a HEK cell, starting
354 from a diffuse protein state and aggregating over time. LINA can distinguish between the
355 diffuse protein and the aggregated state, as it only starts to predict an aggregate at the
356 time-point where the aggregate begins to form, as verified visually from the ground-truth
357 fluorescence images. It is able to correctly detect the aggregate as it grows in size from
358 then on, and as it moves along different subcellular localizations. Fig. 4b illustrates the
359 normalized mean intensity of the output images along time. As the aggregate grows and
360 more pixels are predicted as parts of an aggregate, the mean intensity of the output image
361 increases. The mean intensity follows the expected sigmoidal behavior and can be
362 divided into three regimes. First, for diffuse protein, the prediction intensities are very low,
363 which shows successful predictions that there are no aggregates. Then, for aggregate
364 formation, the intensity starts increasing gradually while the aggregate grows, until the
365 third regime is reached, where the intensity stabilizes into a steady state and the
366 aggregate stops growing.

367 The dry mass of aggregates can be extracted from the phase images and the
368 segmentation masks of the network output images in the same fashion as described
369 previously for the fixed cells. Furthermore, the area can also be extracted from the mask

of the identified aggregate. Both the dry mass and area are plotted in Fig. 4c, and they are highly correlated to each other and to the normalized mean intensity. Once again, they can be split into three regimes: diffuse protein, aggregate formation, and stabilizing to a steady state. The ability to measure the dry mass and area of living aggregates as they grow opens up many possibilities for dynamics studies.

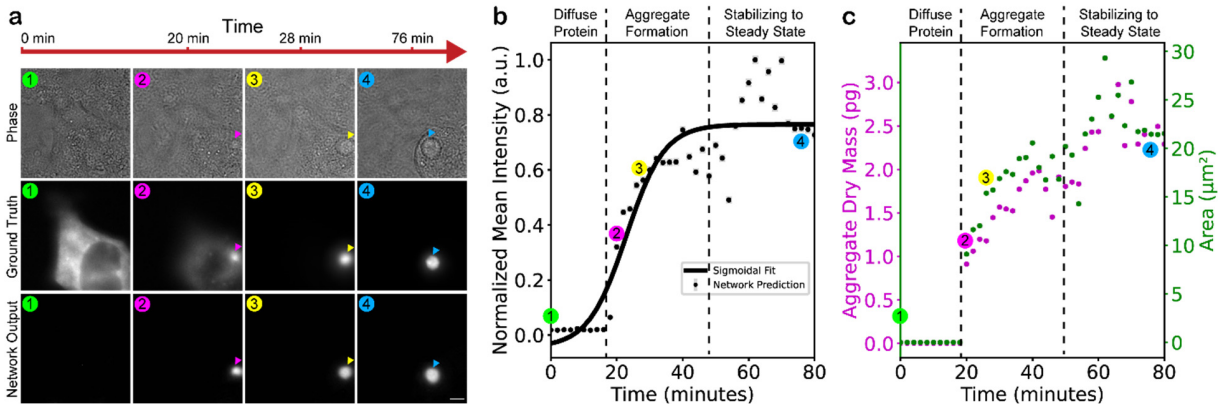

Figure 4: Live, label-free identification and analysis of Httex1 protein aggregation. (a) Time-lapse images, acquired every 2 minutes, of Httex1-72Q-GFP in a HEK cell, starting from a diffuse protein state and aggregating over time. LINA is able to distinguish between the diffuse protein and the aggregated state, correctly predicting the aggregate as it grows in size and moves along different subcellular localizations. (b) Normalized mean intensity of the network output images as the aggregate grows, following a three-regime sigmoidal behaviour. (c) Dry mass and area changes extracted from the network output images, split into the same three regimes. (c) Scale bar: 5 μm.

383 ## Discussion

384 LINA is a label-free method for virtually labeling and identifying protein aggregates using
385 a trained deep-learning CNN, enabling more accurate analysis of the native aggregates,
386 and avoiding label-induced alterations. We have validated LINA quantitatively, showed
387 its applicability on different constructs of Httex1, and verified the models' consistently-
388 accurate performance over various imaging conditions, including varying SNR, cell line,
389 and input type, illustrating the universality of our method. Furthermore, LINA is fully
390 compatible with long-term live-cell imaging, circumventing the issues of photobleaching
391 and phototoxicity, and can be used to identify aggregates and analyze their quantitative
392 parameters (e.g., area, dry mass, morphology, and interaction dynamics). In support of
393 prior studies[14–17], our characterization and comparison of the dry masses of unlabeled
394 and labeled aggregates further demonstrate the need for using label-free methods to
395 study protein aggregation and inclusion formation in NDDs. We found that mutant Httex1
396 with a truncated Nt17 domain and a polyQ repeat length of 72Q forms aggregates of the
397 same dry mass on average as the construct with the same polyQ repeat length but
398 maintaining the Nt17 domain. We found that the aggregates formed by Httex1 with a
399 polyQ repeat length of 39Q, which lack the core-and-shell ultrastructural arrangement of
400 the two other constructs, are larger on average, despite a reduction in the number of
401 formed aggregates. This is likely due to the different ultrastructure and aggregation
402 mechanism of Httex1-39Q. Further analysis and comparisons of the characteristics of
403 different label-free Httex1 constructs will be possible using LINA, which could lead to new
404 discoveries into the pathological mechanisms of Huntington's disease and uncover
405 possible therapeutic targets.

406 What is the minimal amount of data needed to successfully train a LINA model?
407 To answer this question, we first computationally estimated this by taking increasingly-
408 sized samples from our complete dataset, which consists of over 1,000 8-plane phase
409 and fluorescence image stacks. We calculated the mean aggregate area and circularity
410 for each sample size, repeated the sampling 100,000 times, and took the average. We
411 plotted the discrepancy from the ground truth standard deviation to indicate the sample
412 size at which we obtain a similar distribution as that of the ground truth (Supplementary

Fig. 6); here, we consider the ground truth standard deviation to be that of the entire dataset, since it leads to well-performing models and should be large enough to capture the diversity of the different kinds of aggregates. Supplementary Fig. 7 shows the morphological distribution of the entire dataset. The simulation showed a saturation at around 30-50% of the size of the complete dataset, indicating that there could be diminishing returns past this point. We then aimed to answer this question experimentally by attempting to train models with varying amounts of data (10%, 20%, 30%, and 50% of the entire dataset). It is possible to train LINA models even with only 10% of our dataset (Supplementary Fig. 8). However, performance does degrade slightly for the models trained with lower amounts of data, except for the model trained with 50% of the data, which yields very similar results to our standard model trained with 90% of the data (the remaining 10% is used as the test set), in agreement with our simulation results. While it is possible to train a LINA model with just 10% of the data, we visually observed this model to not do as well as the 90% model on some test set images. It is crucial that the training and validation datasets capture the diversity of the subject of the data to ensure high generalizability and accuracy of the trained model, and, therefore, more data will very often lead to better results. However, care should be taken to only provide high quality, useful data to the CNN, as low-quality data can actually confuse the network and lead to hallucinations[47]. The quality of the dataset is just as important as its quantity, making data curation a crucial step in successfully training a deep learning model.

Nowadays, artificial intelligence (AI), particularly machine learning, is extensively being used in microscopy for various purposes, including denoising, image-to-image translation, segmentation, classification, and even medical diagnosis[36–39,48–53]. There have also been developments aiming to ease the adoption of these models and democratize their use[54–56]. As with any new technology, there can be mistrust and a reluctance to rely on it, especially when it is a “black box” that is difficult to interpret – fortunately, there has been considerable progress in deciphering how deep learning models make their predictions[47]. In the case of AI-enhanced microscopy, concern can be warranted because of the field’s fundamental nature and it being a basis for biomedical discoveries, which makes it truly important for any results to be unequivocally true, as this affects future studies and research directions which, depending on their success, can

either facilitate or delay the treatment of devastating illnesses. This is especially the case when it comes to unsupervised learning using generative models, such as those using a generative adversarial network (GAN). GANs create highly-plausible outputs that can be hard to distinguish from real data, and it can be difficult to determine if biological phenomena identified by such a model are true or a hallucination. To mitigate such concerns, it is important to validate the AI model as much as possible and to discuss the kind of problem that it excels at solving, as well as the model's limitations. That is why we validated our models as much as possible, quantitatively, on different protein constructs and at different conditions.

We believe that our method has significant potential for characterizing protein aggregates and inclusions associated with other NDDs. This could be achieved by following a similar approach of training a CNN from scratch on a new dataset, or by simply performing transfer learning on our trained model with the new data. Our method has the advantage of not requiring manual annotations, as the training labels are directly obtained from the fluorescence images. Therefore, researchers who are interested in extending our approach using new data only need access to a microscope capable of imaging in a label-free modality and any kind of fluorescence microscopy. Additionally, LINA paves the way for studying the molecular interaction dynamics of protein aggregates and the evaluation of the cellular state in cells containing aggregates, without the need to rely on labels. It can be extended to virtually label other cellular components, such as organelles or cytoskeletal proteins, which are known to interact with NDD-associated protein inclusions, and several studies have already shown their capability to be virtually labeled[36,37,57]. Moreover, LINA can be further developed to classify the cellular state[36] of label-free cells containing aggregates to elucidate the mechanisms that disrupt cellular viability or introduce toxicity. We also envisage several possibilities for LINA to be correlated with other imaging techniques, opening up new avenues for correlative microscopy of neurodegenerative diseases.

471 # Methods

472 ## Sample preparation

473 HEK 293 cells were cultured at 37 °C and 5% CO_2 using DMEM high glucose without
474 phenol red (Gibco, Thermo Fisher Scientific), supplemented with 10% fetal bovine serum,
475 1% penicillin-streptomycin and 4 mM L-glutamine (all three from Gibco, Thermo Fisher
476 Scientific). Cells were plated at a density of 120,000 per dish either on high precision #1.5
477 25 mm coverslips (Marienfeld) or on FluoroDish Sterile Culture Dishes 35 mm, 23 mm
478 well (World Precision Instruments), coated with fibronectin. Cells were transfected one
479 day after plating using polyethylenimine (PEI) transfection. 2 µg of DNA are mixed in 100
480 µl of OptiMEM Reduced-Serum Medium (Life Technologies), 6 µl of PEI are mixed in 100
481 µl of OptiMEM, and then both mixtures are mixed and incubated for 5 minutes at room
482 temperature (RT), then added drop-wise and carefully distributed over the cells. Cells
483 were then returned to the incubator and left either for 48 hours before fixation or for shorter
484 durations before live-cell imaging followed by fixation at 48 hours post-transfection. For
485 fixation, cells were washed twice with PBS pH 7.4 (1X) (Life Technologies, Switzerland)
486 and fixed in 3.7% formaldehyde (Sigma-Aldrich, Switzerland) in PBS (PFA) for 15 min at
487 room temperature (RT). Cells were then washed in PBS then mounted in PBS.

488 For immunocytochemistry, after a blocking step with 3% BSA (Sigma-Aldrich,
489 Switzerland) diluted in 0.1% Triton X-100 (Applichem, Germany) in PBS (PBST) for
490 30 min at RT, cells were incubated with the primary antibody (anti-Htt raised against the
491 Proline-rich domain (PRD) (MAB5492, Millipore) at a dilution of 1/500 in PBST for 2 h at
492 RT. Cells were then rinsed five times in PBST and incubated for 1 h at RT with the
493 secondary donkey anti-mouse Alexa488 (Life Technologies, Switzerland) used at a
494 dilution of 1/800 in PBST. Cells were then washed five times in PBST, and finally washed
495 once in double-distilled H2O, before being mounted in PBS.

496 ## Data acquisition and processing

497 All imaging was performed with a custom-built microscope equipped with a temperature
498 and CO_2 controlled incubator for live cell imaging, as described in previous work from our
499 group[41]. Live-cell imaging was performed in DMEM without phenol red at 37 °C and 5%

500 CO_2. The microscope is controlled using custom LabVIEW software. For fluorescence
501 imaging, a 120 mW, 488 nm laser (iBeam smart, Toptica), is focused into the back focal
502 plane of an Olympus UPLSAPO 60XW 1.2 NA objective for wide-field epi-fluorescence
503 illumination. The fluorescence light was filtered using a combination of a dichroic mirror
504 (zt405/488/532/640/730rpc, Chroma) and an emission filter. For phase imaging, we used
505 the white-light Koehler illumination module of a Zeiss Axiovert 100 M microscope
506 equipped with a halogen lamp to collect brightfield images which are later processed into
507 QPI images. The detection path is arranged as a sequence of four 2-f configurations to
508 provide image–object space telecentricity. The image splitter placed behind the last lens
509 directs the light into eight images, which are registered by two synchronized sCMOS
510 cameras (ORCA Flash 4.0, Hamamatsu; back projected pixel size of 111 nm). For
511 translating the sample, the microscope is equipped with piezoLEGS stage (3-PT-60-
512 F2,5/5) and Motion-Commander-Piezo controller (Nanos Instruments GmbH). While
513 collecting our dataset, we acquired any FOV that had an aggregate, not filtering for
514 particular sizes or shapes, to have a dataset which is as general as possible.

515 We used custom MATLAB (Mathworks) scripts (available here) to retrieve the
516 phase information from the brightfield images and produce quantitative phase images.
517 These scripts are also used for pixel-registration in the 8 z-planes for both phase and
518 fluorescence images. The images are cropped to a size of 352 pixels x 352 pixels.

519 We used Fiji[58] scripts to produce maximum z-projection fluorescence images, to
520 segment these images using Otsu thresholding and produce the labels for pixel
521 classification, to prepare the color-coded maximum z-projection phase image shown in
522 Fig. 2a, and for the image processing in Fig. 2b. Fiji was also used to segment network
523 predictions and produce masks which are used to measure the area, circularity or dry
524 mass of aggregates. Python was used for data analysis and plotting.

525 **Neural network training**

526 Our models are trained on a deep CNN with a U-Net[42] architecture. Compared to the
527 original architecture, we reduced the number of feature maps by a factor of 4 which led
528 to a reduction in the trainable parameters by a factor of 15. This notably reduces GPU
529 memory usage and training time, while still enabling excellent performance. We used

TensorFlow and Keras to build our network, and training was done on a workstation equipped with an NVIDIA GeForce RTX 3090 GPU. We used the adaptive moment estimation (Adam) optimizer with a learning rate of 1e-4 and a mean squared error loss function. Before training our models, we normalize both the phase and fluorescence images by rescaling each image to be between 0 and 1. We used 10% of our dataset as the test set and split the rest of the dataset into training and validation sets with 20% being used for validation. We used the 'EarlyStopping' (on the validation loss) and 'ModelCheckpoint' callbacks to avoid overfitting and to save the best, most general models.

Dry mass quantification

Supplementary Fig. 5 summarizes the dry mass extraction process. To map between phase information and dry mass, we use the following equation[35]:

$$\sigma(x, y) = \frac{\lambda}{2\pi\alpha}\phi(x, y)$$

where σ and ϕ are the dry mass and the phase, respectively, at the location defined by the x and y coordinates, λ is the wavelength of the illumination (here chosen as the average wavelength of our white-light source), and α is the refractive increment, treated as a constant value of 0.19 $\mu m^3 \cdot pg^{-1}$, as recommended in the literature, due to α having an extremely narrow distribution[35,43].

Acknowledgments

We thank Samuel M. Leitao for testing the method on a different imaging setup and Anne-Laure Mahul Mellier for the helpful discussions. This work was financially supported by the SV-iPhD program. A.R. and L.F. were supported by funding through the Swiss National Science Foundation (SNSF) through the National Centre of Competence in Research Bio-Inspired Materials and SNF grant CRSII5_193740.

Author contributions

K.G. conceived the idea and initiated the project, which was driven by the observations of N.R.; K.A.I., H.A.L., and A.R. designed the experiments; K.A.I., L.F., and N.R. prepared the samples; K.A.I. performed the experiments, acquired, analyzed, and visualized the data, developed and optimized the neural network training pipeline, validated the trained models, and developed the pipeline to extract the dry mass and area measurements; K.A.I. and H.A.L. wrote the paper, with input from all authors. H.A.L. and A.R. supervised the project; All authors discussed the results and commented on the manuscript.

562 # References

563 1. Forman, M. S., Trojanowski, J. Q. & Lee, V. M.-Y. Neurodegenerative diseases: a decade of
564 discoveries paves the way for therapeutic breakthroughs. *Nat Med* **10**, 1055–1063 (2004).
565 2. Hansson, O. Biomarkers for neurodegenerative diseases. *Nat Med* **27**, 954–963 (2021).
566 3. Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature Reviews*
567 *Neuroscience* **4**, 49-60, (2003).
568 4. Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nature Medicine* **10**,
569 10-17, (2004).
570 5. Goedert, M. Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled Aβ,
571 tau, and α-synuclein. *Science* **349**, (2015).
572 6. Soto, C. & Pritzkow, S. Protein misfolding, aggregation, and conformational strains in
573 neurodegenerative diseases. *Nature Neuroscience* **21**, 1332-1340, (2018).
574 7. Ross, C. A. *et al.* Huntington disease: natural history, biomarkers and prospects for therapeutics.
575 *Nature Reviews Neurology* **10**, 204-216, (2014).
576 8. Davies, S. W. *et al.* Formation of neuronal intranuclear inclusions underlies the neurological
577 dysfunction in mice transgenic for the HD mutation. *Cell* **90**, 537-548, (1997).
578 9. Sisodia, S. S. Nuclear inclusions in glutamine repeat disorders: are they pernicious, coincidental, or
579 beneficial? *Cell* **95**, 1-4, (1998).
580 10. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation
581 reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431**, 805-810, (2004).
582 11. Saudou, F., Finkbeiner, S., Devys, D. & Greenberg, M. E. Huntingtin acts in the nucleus to induce
583 apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* **95**, 55-66,
584 (1998).
585 12. Yu, Z.-X., Li, S.-H., Nguyen, H.-P. & Li, X.-J. Huntingtin inclusions do not deplete polyglutamine-
586 containing transcription factors in HD mice. *Human Molecular Genetics* **11**, 905-914, (2002).
587 13. Swulius, M. T. & Jensen, G. J. The helical mreb cytoskeleton in Escherichia coli MC1000/pLE7 is an
588 artifact of the N-terminal yellow fluorescent protein tag. *Journal of bacteriology* **194**, 6382-6386,
589 (2012).
590 14. Riguet, N. *et al.* Nuclear and cytoplasmic huntingtin inclusions exhibit distinct biochemical
591 composition, interactome and ultrastructural properties. *Nature communications* **12**, (2021).
592 15. Dahlgren, P. R. *et al. Atomic force microscopy analysis of the Huntington protein nanofibril*
593 *formation*. (NBM, 2005).
594 16. Bäuerlein, F. J. *et al.* In situ architecture and cellular interactions of polyQ inclusions. *Cell* **171**, 179-
595 187, (2017).
596 17. Shillcock, J. C., Hastings, J., Riguet, N. & Lashuel, H. A. Non-monotonic fibril surface occlusion by GFP
597 tags from coarse-grained molecular simulations. *Computational and Structural Biotechnology*
598 *Journal* **20**, 309-321, (2022).
599 18. Caputo, A. *et al.* Snca-GFP Knock-In Mice Reflect Patterns of Endogenous Expression and
600 Pathological Seeding. *eNeuro* **7**, (2020).

601 19. Afitska, K., Fucikova, A., Shvadchak, V. V. & Yushchenko, D. A. Modification of C Terminus Provides
602 New Insights into the Mechanism of α-Synuclein Aggregation. *Biophysical Journal* **113**, 2182–2191
603 (2017).
604 20. Kaniyappan, S. *et al.* FRET-based Tau seeding assay does not represent prion-like templated
605 assembly of Tau filaments. *Molecular Neurodegeneration* **15**, 39 (2020).
606 21. Hussein, R., Hashem, R. & Rashed, L. Evaluation of the amyloid beta-GFP fusion protein as a model
607 of amyloid beta peptides-mediated aggregation: A study of DNAJB6 chaperone. *Frontiers in*
608 *Molecular Neuroscience* **8**, (2015).
609 22. Azizyan, R. A. *et al.* Establishment of Constraints on Amyloid Formation Imposed by Steric Exclusion
610 of Globular Domains. *Journal of Molecular Biology* **430**, 3835–3846 (2018).
611 23. Neueder, A. *et al.* The pathogenic exon 1 Htt protein is produced by incomplete splicing in
612 Huntington's disease patients. *Scientific Reports* **7**, (2017).
613 24. Wellington, C. L. *et al.* Caspase cleavage of mutant Huntingtin precedes neurodegeneration in
614 Huntington's disease. *Journal of Neuroscience* **22**, 7862-7872, (2002).
615 25. Schilling, G. *et al.* Characterization of huntingtin pathologic fragments in human Huntington disease,
616 transgenic mice, and cell models. *Journal of Neuropathology & Experimental Neurology* **66**, 313-320,
617 (2007).
618 26. Martindale, D. *et al.* Length of huntingtin and its polyglutamine tract influences localization and
619 frequency of intracellular aggregates. *Nature Genetics* **18**, 150-154, (1998).
620 27. Mangiarini, L. *et al.* Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a
621 progressive neurological phenotype in transgenic mice. *Cell* **87**, 493-506, (1996).
622 28. Scherzinger, E. *et al.* Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-
623 like fibrils: Implications for Huntington's disease pathology. *PNAS* **96**, 4604-4609, (1999).
624 29. Wellington, C. L. *et al.* Inhibiting caspase cleavage of huntingtin reduces toxicity and aggregate
625 formation in neuronal and nonneuronal cells. *Journal of Biological Chemistry* **275**, 19831-19838,
626 (2000).
627 30. Lunkes, A. *et al.* Proteases acting on mutant huntingtin generate cleaved products that differentially
628 build up cytoplasmic and nuclear inclusions. *Molecular Cell* **10**, 259-269, (2002).
629 31. Chongtham, A. *et al.* Effects of flanking sequences and cellular context on subcellular behavior and
630 pathology of mutant Htt. *Human Molecular Genetics* **29**, 674–688 (2020).
631 32. Miao, K. & Wei, L. Live-cell imaging and quantification of polyQ aggregates by stimulated Raman
632 scattering of selective deuterium labeling. *ACS Central Science* **6**, 478-486, (2020).
633 33. Cuche, E., Marquet, P. & Depeursinge, C. Simultaneous amplitude-contrast and quantitative phase-
634 contrast microscopy by numerical reconstruction of Fresnel off-axis holograms. *Applied Optics* **38**,
635 6994-7001, (1999).
636 34. Popescu, G. *Quantitative phase imaging of cells and tissues*. (McGraw-Hill Education, 2011).
637 35. Park, Y. K., Depeursinge, C. & Popescu, G. Quantitative phase imaging in biomedicine. *Nature*
638 *Photonics* **12**, 578-589, (2018).
639 36. Christiansen, E. M. *et al.* In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**,
640 792-803, (2018).

641 37. Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F. & Johnson, G. R. Label-free prediction of
642 three-dimensional fluorescence images from transmitted-light microscopy. *Nat Methods* **15**, 917–
643 920 (2018).
644 38. Rivenson, Y. *et al.* PhaseStain: the digital staining of label-free quantitative phase microscopy
645 images using deep learning. *Light, Science & Applications* **8**, (2019).
646 39. Rivenson, Y. *et al.* Virtual histological staining of unlabelled tissue-autofluorescence images via deep
647 learning. *Nat Biomed Eng* **3**, 466–477 (2019).
648 40. Lin, L.-E., Miao, K., Qian, C. & Wei, L. High spatial-resolution imaging of label-free in vivo protein
649 aggregates by VISTA. *Analyst* **146**, 4135–4145 (2021).
650 41. Descloux, A. *et al.* Combined multi-plane phase retrieval and super-resolution optical fluctuation
651 imaging for 4D cell microscopy. *Nature Photonics* **12**, 165-172, (2018).
652 42. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image
653 segmentation. in *Medical Image Computing and Computer-assisted Intervention—MICCAI 2015*
654 (2015).
655 43. Zhao, H., Brown, P. H. & Schuck, P. On the distribution of protein refractive index increments.
656 *Biophysical Journal* **100**, 2309-2317, (2011).
657 44. Rosenblatt, A. *et al.* The association of CAG repeat length with clinical progression in Huntington
658 disease. *Neurology* **66**, 1016–1020 (2006).
659 45. Rosenblatt, A. *et al.* Age, CAG repeat length, and clinical progression in Huntington's disease.
660 *Movement Disorders* **27**, 272–276 (2012).
661 46. Tabrizi, S. J. *et al.* Predictors of phenotypic progression and disease onset in premanifest and early-
662 stage Huntington's disease in the TRACK-HD study: analysis of 36-month observational data. *The*
663 *Lancet Neurology* **12**, 637–649 (2013).
664 47. Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence
665 image reconstruction. *Nat Methods* **16**, 1215–1225 (2019).
666 48. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy.
667 *Nat Methods* **15**, 1090–1097 (2018).
668 49. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-
669 resolution localization microscopy. *Nat Biotechnol* **36**, 460–468 (2018).
670 50. Nehme, E., Weiss, L. E., Michaeli, T. & Shechtman, Y. Deep-STORM: super-resolution single-molecule
671 microscopy by deep learning. *Optica* **5**, 458 (2018).
672 51. Buggenthin, F. *et al.* Prospective identification of hematopoietic lineage choice by deep learning.
673 *Nat Methods* **14**, 403–406 (2017).
674 52. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature*
675 **542**, 115–118 (2017).
676 53. Kermany, D. S. *et al.* Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep
677 Learning. *Cell* **172**, 1122-1131.e9 (2018).
678 54. Haberl, M. G. *et al.* CDeep3M—Plug-and-Play cloud-based deep learning for image segmentation.
679 *Nat Methods* **15**, 677–680 (2018).

680 55. Chamier, L. *et al.* Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nature*
681 *Communications* **12**, 2276-2276, (2021).
682 56. Gómez-de-Mariscal, E. *et al.* DeepImageJ: A user-friendly environment to run deep learning models
683 in ImageJ. *Nature Methods* **18**, 1192-1195, (2021).
684 57. Sekh, A. A. *et al.* Physics-based machine learning for subcellular segmentation in living cells. *Nature*
685 *Machine Intelligence* **3**, 1071-1080, (2021).
686 58. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–
687 682 (2012).