

1 **CenSegNet: a generalist high-throughput deep learning framework for**  
2 **centrosome phenotyping at spatial and single-cell resolution in**  
3 **heterogeneous tissues**

4 Jiaoqi Cheng<sup>1,2,8</sup>, Keqiang Fan<sup>2,3,8</sup>, Miles Bailey<sup>1,2</sup>, Xin Du<sup>4</sup>, Rajesh Jena<sup>5</sup>, Costantinos Savva<sup>6,7</sup>,  
5 Mengyang Gou<sup>1,2</sup>, Ramsey Cutress<sup>2,6</sup>, Stephen Beers<sup>2,7\*</sup>, Xiaohao Cai<sup>2,3\*</sup> and Salah Elias<sup>1,2,9\*</sup>

6

7 <sup>1</sup>School of Biological Sciences, University of Southampton, Southampton SO17 1BJ, UK

8 <sup>2</sup>Institute for Life Sciences, University of Southampton, Southampton SO17 1BJ, UK

9 <sup>3</sup>School of Electronics and Computer Science, University of Southampton, Southampton SO17  
10 1BJ, UK

11 <sup>4</sup>The Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge, CB3  
12 0FZ, UK

13 <sup>5</sup>Department of Oncology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK,  
14 University of Cambridge, CB2 0QQ, UK

15 <sup>6</sup>University Hospital Southampton NHS Foundation Trust, Southampton, SO16 6YD, UK

16 <sup>7</sup>Centre for Cancer Immunology, School of Cancer Sciences, Faculty of Medicine, University of  
17 Southampton, Southampton, SO16 6YD, UK

18 <sup>8</sup>These authors contributed equally

19 <sup>9</sup>Lead contact

20 \*Correspondence: [s.a.beers@soton.ac.uk](mailto:s.a.beers@soton.ac.uk), [x.cai@soton.ac.uk](mailto:x.cai@soton.ac.uk) and [s.k.elias@soton.ac.uk](mailto:s.k.elias@soton.ac.uk)

21

22 **Abstract**

23 Centrosome amplification (CA) is a hallmark of epithelial cancers, yet its spatial complexity and  
24 phenotypic heterogeneity remain poorly resolved due to limitations in conventional image analysis.  
25 We present CenSegNet (Centrosome Segmentation Network), a modular deep learning framework  
26 for high-resolution, context-aware segmentation of centrosomes and epithelial architecture across  
27 diverse tissue types. Integrating a dual-branch architecture with uncertainty-guided refinement,  
28 CenSegNet achieves state-of-the-art performance and generalisability across both  
29 immunofluorescence and immunohistochemistry modalities, outperforming existing models in  
30 accuracy and morphological fidelity. Applied to tissue microarrays (TMAs) containing 911 breast  
31 cancer sample cores from 127 patients, CenSegNet enables the first large-scale, spatially resolved  
32 quantification of numerical and structural CA at single-cell resolution. These CA subtypes are  
33 mechanistically uncoupled, exhibiting distinct spatial distributions, age-dependent dynamics, and  
34 associations with histological tumour grade, hormone receptor status, genomic alterations, and  
35 nodal involvement. Discordant CA profiles at tumour margins are linked to local aggressiveness and  
36 stromal remodelling, underscoring their clinical relevance. To support broad adoption and  
37 reproducibility, CenSegNet is released as an open-source Python library. Together, our findings  
38 establish CenSegNet as a scalable, generalisable platform for spatially resolved centrosome  
39 phenotyping in intact tissues, enabling systematic dissection of the biology of this organelle and its  
40 dysregulation in cancer and other epithelial diseases.

41

## 42 Introduction

43 Centrosomes, composed of a pair of orthogonally arranged centrioles surrounded by pericentriolar  
44 material (PCM), function as the principal microtubule-organising centres (MTOCs) in animal cells.  
45 They play essential roles in diverse cellular processes, including vesicular trafficking, cell polarity,  
46 motility, ciliogenesis, and the assembly of a bipolar mitotic spindle during cell division<sup>1,2</sup>. Centrosome  
47 number is tightly regulated, with duplication occurring once per cell cycle during S phase, ensuring  
48 the formation of the mitotic spindle and equal inheritance of chromosomes by daughter cells<sup>3-5</sup>.

49 Centrosome amplification (CA) can lead to multipolar spindle formation, chromosomal  
50 missegregation, and aneuploidy<sup>3, 6, 7</sup>—a hallmark of cancer<sup>3, 4, 7</sup>. The hypothesis that CA-induced  
51 aneuploidy contributes to tumorigenesis was first proposed by Theodor Boveri over a century ago<sup>8</sup>.  
52 In recent years, CA has been documented in several solid tumours including breast, prostate, colon,  
53 ovarian, and pancreatic cancers<sup>3, 7, 9-13</sup>, as well as haematological malignancies such as multiple  
54 myeloma, lymphomas, and leukaemias<sup>14, 15</sup>. While its role in tumour initiation remains debated<sup>6, 9, 16-</sup>  
55 <sup>18</sup>, CA is consistently associated with aggressive disease features, including high-grade histology,  
56 poor prognosis, recurrence, and metastasis<sup>3, 7, 9, 19</sup>. Despite its clinical relevance, CA remains poorly  
57 characterised at scale due to the lack of robust, high-throughput tools capable of resolving  
58 centrosome phenotypes in complex tissue architecture.

59 Mechanistically, CA arises from both numerical and structural centrosome defects. Numerical  
60 amplification results from centriole overduplication, *de novo* centriole assembly, cytokinesis failure,  
61 mitotic slippage and cell–cell fusion<sup>20-29</sup>. Disruption of cell-cycle progression, such as prolonged G2  
62 arrest, can trigger premature centriole disengagement and reduplication *via* PLK1 activation<sup>30, 31</sup>.  
63 Fragmentation of the PCM, driven by dysregulation of proteins such as pericentrin, γ-tubulin, PLK4,  
64 PLK1, and Aurora-A, also contributes to numerical CA<sup>7, 32-34</sup>. Structural CA, on the other hand,  
65 involves aberrant accumulation of PCM<sup>4, 7, 9, 35, 36</sup> or defects in centriole architecture<sup>37-39</sup>. Among these  
66 centriole architectural defects, over-elongation and fragmentation can lead to unstable centriole

67 structures and further overduplication<sup>37</sup>, suggesting a mechanistic link between numerical and  
68 structural centrosome aberrations. Yet, their differential contributions to cancer biology remain to be  
69 determined.

70 Manual centrosome annotation remains the standard but is time-consuming, low-throughput, and  
71 prone to observer bias. Semi-automated pipelines have emerged to address these limitations. For  
72 example, CenFind—a deep learning pipeline based on SpotNet—accurately detects and counts  
73 centrioles in cultured cells using immunofluorescence images but does not support structural  
74 phenotyping<sup>40</sup>. Other machine learning-based approaches have quantified centriole number and  
75 linked supernumerary centrioles to PCM expansion in breast cancer cells<sup>41</sup>. Semi-automated  
76 frameworks have also been developed for centrosome quantification in human breast histological  
77 sections, including a HistoQuest-aided method detecting structural CA<sup>42</sup> and an IMARIS-based  
78 pipeline integrating both structural and numerical CA<sup>7, 43</sup>. Similarly, a 3D imaging-based pipeline  
79 quantified structural centriole abnormalities across cancer types<sup>37</sup>. However, these approaches  
80 require manual curation and offer only moderate throughput. A recent high-throughput platform using  
81 a Harmony™ software-based framework revealed heterogeneous CA phenotypes in ovarian cancer  
82 tissues<sup>44</sup>, yet lacked single-cell resolution and subtype discrimination. Moreover, most existing tools  
83 are tailored to immunofluorescence imaging and are not compatible with standard chromogenic  
84 immunohistochemistry workflows used in clinical pathology, limiting their diagnosis and translational  
85 utility.

86 To address these limitations, we developed CenSegNet (Centrosome Segmentation Network), a  
87 versatile deep learning framework for fully automated, pixel-level detection and segmentation of  
88 centrosomes in both immunohistochemistry and immunofluorescence images at single-cell  
89 resolution. CenSegNet integrates three state-of-the-art models: Ultralytics YOLOv11, a recent  
90 evolution of the You Only Look Once family optimised for fast and accurate performance<sup>45</sup>, U-Net,  
91 an encoder-decoder convolutional network designed for precise pixel-wise segmentation<sup>46</sup>, and  
92 StarDist for shape-aware instance segmentation pipeline that models objects as star-convex

93 polygons to improve instance separation in dense cellular contexts<sup>47</sup>, enabling robust delineation of  
94 epithelial cell boundaries within histopathological specimens. This systematically engineered  
95 architecture supports multiscale analysis of centrosomal features in morphologically complex tissue  
96 environments. We present a publicly accessible, expert-annotated dataset comprising human and  
97 murine breast tissues and human mammary epithelial cell cultures (MECs). Implemented in Python  
98 3.10 with a PyQt5-based graphical interface, CenSegNet enables streamlined data input, real-time  
99 PyTorch-based inference, and modular extensibility. Benchmarking against expert annotations and  
100 alternative models, CenSegNet achieves pathologist-level accuracy across imaging modalities.  
101 Using CenSegNet, we perform the first high throughput, spatially resolved quantification of numerical  
102 and structural CA in clinical breast carcinomas. Our analyses reveal that these CA subtypes are  
103 mechanistically uncoupled and evolve along orthogonal spatial gradients: numerical CA  
104 predominates in proliferative tumour cores, whereas structural CA accumulates at invasive margins,  
105 reflecting distinct evolutionary pressures and microenvironmental cues. These spatial trajectories  
106 correlate with histological grade, hormone receptor status, HER2 expression, nodal involvement,  
107 and germline alterations, underscoring the role of centrosome dysregulation in driving intratumoral  
108 heterogeneity and progression. Importantly, we validated CenSegNet in other human epithelial  
109 tissues including kidney, colon, and appendix, demonstrating its generalisability and potential for  
110 broad application in spatial pathology and organelle-level phenotyping across diverse healthy and  
111 disease tissue contexts. To support widespread adoption, CenSegNet is released as an open-source  
112 Python library, available at <https://github.com/SKELab/CenSegNet/> and  
113 <https://zenodo.org/records/17131573>.

114

115 **Results**

116

117 **Development of CenSegNet for robust centrosome segmentation across imaging modalities**

118 We generated tissue microarrays (TMAs) comprising 911 breast tissue cores from normal breast  
119 tissue, breast tumours, and adjacent areas, sampled from 127 patients enrolled in the ethically  
120 approved and clinically well-characterised BeGIN cohort (Investigating Outcomes from Breast  
121 Cancer: Correlating Genetic, Immunological and Nutritional Predictors), from University Hospital  
122 Southampton (UHS) (**Fig. 1a**; see Methods). Immunohistochemistry was performed using pericentrin  
123 antibody to label centrosomes, with haematoxylin counterstaining for nuclei (**Fig. 1a**). For training  
124 dataset construction, we manually annotated 14,679 centrosomes within 2,486 epithelial and 108  
125 stromal compartments across 108 selected images (**Fig. 1a**, see Methods). To complement this,  
126 immunofluorescence was performed on human MECs and mouse mammary epithelium, labelling  
127 pericentrin and GT335 (centriole), with DAPI as nuclear counterstain. From this, an  
128 immunofluorescence training dataset was assembled comprising 1,285 annotated centrosomes  
129 from 143 cells from mouse tissue and 841 human MECs, revealing strong segmentation  
130 concordance between pericentrin and GT335 labelling [mouse tissue:  $R^2 = 0.9954$ ; human MECs:  
131  $R^2 = 0.9666$  (MCF10A), 0.9085 (MCF10A-PLK4)] (**Fig. 1b**, **Supplementary Fig. 1a–f**).

132 Using these datasets, we initially benchmarked established segmentation models. U-Net, an  
133 encoder-decoder convolutional neural network optimised for pixel-wise segmentation<sup>46</sup>, was  
134 selected for its extensive use in biomedical imaging. In our datasets, U-Net detected 82.98% of  
135 centrosomes in immunohistochemistry images and 97.6% in immunofluorescence, but despite  
136 achieving an overall F1 score of 0.85 in immunofluorescence, often the model either under-predicted  
137 or over-predicted centrosomes (**Supplementary Fig. 2a, b**). We next evaluated SegNet (**Fig. 1c**),  
138 another encoder-decoder model leveraging max-pooling indices for efficient upsampling<sup>48</sup>. It  
139 achieved 72.73% and 85% detection in immunohistochemistry and immunofluorescence,  
140 respectively, with a precision of 0.90 but reduced recall (0.75) in immunofluorescence and a

141 suboptimal overall F1 score of 0.68 in immunohistochemistry (0.82 in immunofluorescence)

142 (**Supplementary Fig. 2c, d**). We also evaluated DeepLabv3+, a semantic segmentation model

143 utilising atrous convolutions and atrous spatial pyramid pooling for multi-scale contextual

144 information<sup>49</sup>. Despite its proven success in complex image domains, DeepLabv3+ achieved only

145 63.25% and 75.1% detection rates in immunohistochemistry and immunofluorescence, respectively

146 (**Supplementary Fig. 2e, f**). These findings highlight the limitations of conventional segmentation

147 pipelines when applied to structurally heterogeneous tissues.

148 To overcome these constraints, we developed CenSegNet, a novel modular framework that

149 integrates object robust detection and segmentation (**Fig. 1c**). Rather than relying on whole-slide

150 segmentation, CenSegNet focuses on single-cell regions while modelling the entire ROI, explicitly

151 integrating object detection with region-based segmentation in a three-step architecture (**Fig. 1c**).

152 This design enables a more comprehensive capture of centrosome spatial distributions and

153 morphological features, addressing key challenges such as occlusion, small object detection, and

154 the disambiguation of overlapping structures. The framework, first employs YOLOv11<sup>45</sup>, a state-of-

155 the-art object detection model, to identify potential centrosome candidates. YOLOv11 comprises a

156 convolutional backbone for feature extraction, a neck for multi-scale aggregation, and a head for

157 classification and localisation. We fine-tuned YOLOv11-seg<sup>45, 50</sup>, a segmentation-optimised variant

158 to further enhance detection accuracy, and applied a range of data augmentations—including hue,

159 saturation, and value (HSV) adjustments, as well as geometric transformations such as translation,

160 scaling, shearing, and horizontal flipping—to enhance generalisability (**Fig. 1c**; see Methods). To

161 refine segmentation, we integrated U-Net in the second step. For all detected centrosomes, we

162 extracted centrosome-centred patches (256 × 256 pixels, with 40-pixel padding) as input to a U-Net

163 skip-connected encoder-decoder architecture for precise delineation. These patches underwent

164 similar augmentation strategies as in the detection step, ensuring consistency. To accurately quantify

165 centrosome numbers per cell, we incorporated StarDist<sup>47</sup> in the third step, a deep learning based

166 instance segmentation tool widely adopted for nuclear and cell boundary segmentation in biomedical

167 imaging (**Fig. 1c**). In our immunohistochemistry datasets, StarDist segmented 764,354 cells  
168 (**Supplementary Fig. 1g, h**). To validate StarDist-aided cell-level centrosome assignment in  
169 complex tissues, we analysed keratin 8 (KRT8)-labelled mouse immunofluorescence (620 cells) and  
170 human immunohistochemistry (620 cells) tissues. StarDist-based segmentation demonstrated  
171 strong concordance with KRT8 (immunofluorescence:  $R^2 = 0.99$ ; immunohistochemistry:  $R^2 =$   
172 0.9005) (see **Supplementary Fig. 1e-i**).

173 Collectively, these results demonstrate that CenSegNet's multistep architecture effectively  
174 overcomes the limitations of conventional segmentation approaches, delivering robust,  
175 generalisable centrosome detection and quantification across diverse imaging modalities and  
176 complex tissue contexts.

177

#### 178 **Validation of CenSegNet for scalable and high-precision centrosome segmentation**

179 To evaluate CenSegNet's performance, we compiled independent test datasets of 25  
180 immunohistochemistry TMA cores and 17 immunofluorescence images. Predicted centrosome  
181 counts were compared to manually annotated ground truth, revealing strong correlations in both  
182 immunohistochemistry ( $R^2 = 0.9999$ ) and immunofluorescence ( $R^2 = 0.9873$ ) (**Fig. 2a–c**).  
183 CenSegNet consistently outperformed U-Net, SegNet, and DeepLabv3+ in segmentation accuracy  
184 and boundary resolution across both modalities, with the YOLOv11 detection module significantly  
185 enhancing overall performance (**Fig. 2a–c**). To further assess precision, we used an independently  
186 annotated subset of 550 centrosomes from tumour regions in the immunohistochemistry test set.  
187 Again, CenSegNet demonstrated superior pixel-level segmentation compared to all benchmarks  
188 (**Fig. 2d**). F1 score analysis across modalities yielded a mean of 0.82 for CenSegNet, outperforming  
189 U-Net (0.72), SegNet (0.68), and DeepLabv3+ (0.65) (**Fig. 2a–e**). We next benchmarked  
190 performance on 6,475 expert-annotated centrosomes—921 from normal tissue, 2,694 from edge  
191 regions, and 2,860 from tumour cores (**Fig. 2f**). These annotations, stratified by size (0.5–1.0  $\mu\text{m}^2$  to  
192 >10.5  $\mu\text{m}^2$ ), were compared with the full cohort of 333,148 automatically segmented centrosomes

193 (0–0.5  $\mu\text{m}^2$  to >12.0  $\mu\text{m}^2$ )—10,834 from normal tissue, 91,803 from edge tissue, and 230,511 from  
194 tumour tissue (**Fig. 2g**). Both datasets revealed similar size distributions, with the majority falling  
195 within 2.0–3.0  $\mu\text{m}^2$  and a peak at 2.0–2.5  $\mu\text{m}^2$ . Normal tissue showed a higher proportion of 1.0–2.0  
196  $\mu\text{m}^2$  centrosomes, while edge regions had more large centrosomes than either normal or tumour  
197 cores. Using a previously established threshold of 6.5  $\mu\text{m}^2$ -size to define structural CA<sup>42</sup> (see  
198 Methods), both datasets confirmed that centrosomes exceeding this size were absent in normal  
199 tissue. Together, these results establish CenSegNet as a scalable, high-precision tool for  
200 centrosome segmentation, with superior performance across image modalities, tissue  
201 compartments, and size distributions—supporting its utility for high-throughput analysis in complex  
202 tissue environments.

203

204 **CenSegNet enables spatial profiling of CA subtypes and clinical correlates in breast cancer**  
205 To dissect the contribution of numerical (Num CA) and structural (Stru CA) CA in breast cancer, we  
206 applied CenSegNet to multiplexed TMA data. Num CA and Stru CA were not correlated within either  
207 tumour edge or core regions (**Supplementary Fig. 3a**), suggesting mechanistic independence. At  
208 the single-cell level, increasing centrosome number per cell was associated with a reduction in the  
209 size of individual centrosomes (**Supplementary Fig. 3b**), which become smaller, more uniform in  
210 cells containing >4 centrosomes—implying a compensatory constraint on total centrosome volume.  
211 These observations suggest that Stru CA and Num CA represent orthogonal axes of centrosome  
212 dysregulation. In contrast, Num CA levels correlated positively between edge and tumour  
213 compartments ( $R^2 = 0.4857$ ; **Supplementary Fig. 3c**), indicating progressive numerical amplification  
214 from tumour margins inward. This was reflected in an increasing proportion of cells with  $\geq 4$   
215 centrosomes from normal tissue to the edge and tumour core (**Fig. 3a**). Stratifying patients into four  
216 CA phenotypes (Stru<sup>−</sup>Num<sup>−</sup>, Stru<sup>+</sup>Num<sup>+</sup>, Stru<sup>+</sup>Num<sup>−</sup>, Stru<sup>−</sup>Num<sup>+</sup>) revealed that CA was widespread,  
217 detected in 89.8% of edge and 95% of tumour regions (**Supplementary Fig. 3d**). Notably, 73% of  
218 tumour regions exhibited both Stru CA and Num CA, representing a ~19.3% increase compared to

219 edge regions (**Supplementary Fig. 3d**). While the Stru<sup>+</sup>Num<sup>-</sup> group accounted for 18.4% of edge  
220 regions but only 2% of tumour regions, the opposite trend was observed for the Stru<sup>-</sup>Num<sup>+</sup> group  
221 (10.2% *versus* 20%, respectively). These findings allow us to conclude that structural and numerical  
222 centrosome aberrations are mechanistically uncoupled and highlight a spatial shift from centrosome  
223 enlargement at the tumour edge to centrosome accumulation in the tumour core.

224 To investigate how CA correlates with ages, we stratified patients into four groups: <50, 50–60, 60–  
225 70, >70 years. Num CA was elevated in patients aged 50–70 years across edge and tumour regions  
226 (**Fig. 3a, b**). The frequency of patients with detectable Num CA and Stru CA at the tumour edge  
227 markedly decreased in the 50–70 age group, resurging in patients over 70 years (**Fig. 3c**). Although  
228 fewer 50 to 60-year-old patients exhibited detectable CA at the edge, those that did had high levels  
229 of both CA subtypes (**Fig. 3d, e**). CA levels in tumour cores were relatively stable across age.  
230 Notably, normal tissues exhibited age-dependent increases in centrosome size, especially in  
231 patients over 70 (**Fig. 3f, g**), consistent with previous reports linking age to centrosome expansion  
232 via cumulative DNA damage<sup>51</sup>. In edge and tumour regions, centrosome size decreased between  
233 ages 50–70 before rising again in patients over 70 (**Fig. 3h, i**), while centrosome numbers followed  
234 the opposite trend, increasing in the 50–70 group before decreasing in the oldest cohort (**Fig. 3j**).  
235 This dynamic was mirrored by a higher proportion of cells with three centrosomes or more in the 50–  
236 60 age group (**Fig. 3k**). Finally, we did not observe significant differences in age distribution across  
237 the four CA-defined patient groups (Stru<sup>-</sup>Num<sup>-</sup>, Stru<sup>+</sup>Num<sup>+</sup>, Stru<sup>+</sup>Num<sup>-</sup>, and Stru<sup>-</sup>Num<sup>+</sup>)  
238 (**Supplementary Fig. 4a**). Thus, while age alone may not initiate CA in breast cancer, it modulates  
239 its spatial distribution and severity—particularly by promoting numerical amplification in the 50–70  
240 age group and structural amplification in patients over 70.

241 We next evaluated associations between CA subtypes and clinicopathological features. Tumours  
242 lacking CA (Stru<sup>-</sup>Num<sup>-</sup>) displayed smaller total and invasive areas in the tumour core relative to the  
243 edge (**Supplementary Fig. 4b, c**), a pattern absent in other CA groups. Structural and numerical

244 CA prevalence increased with histological grade and nodal involvement (**Fig. 4a, b**). Stru CA was  
245 inversely associated with tumour size across tumour (T) stages both edge and tumour regions, while  
246 Num CA displayed region-specific variation (**Fig. 4c**). Stru CA and Num CA were more frequent at  
247 the edge in invasive ductal carcinoma, but more abundant in the core of mixed-subtype tumours  
248 (**Fig. 4d**). To further assess CA heterogeneity, patients were classified by composite CA defects into  
249 Stru<sup>high</sup>Num<sup>high</sup>, Stru<sup>high</sup>Num<sup>low</sup>, Stru<sup>low</sup>Num<sup>low</sup>, Stru<sup>low</sup>Num<sup>high</sup> CA groups (**Fig. 4e, f**). CA burden was  
250 not age-associated (**Fig. 4g**), and distinct CA subtypes were linked to tumour behaviour. At the edge,  
251 Stru<sup>high</sup>Num<sup>low</sup> and Stru<sup>low</sup>Num<sup>high</sup> CA tumours were associated with greater nodal involvement than  
252 either the Stru<sup>high</sup>Num<sup>low</sup> or Stru<sup>low</sup>Num<sup>high</sup> CA groups (**Fig. 4h**), whereas no such differences were  
253 observed across CA groups in the tumour core. Stru<sup>low</sup>Num<sup>high</sup> CA tumours exhibited the smallest  
254 overall size, while the largest tumours were enriched in Stru<sup>high</sup>Num<sup>low</sup> (edge) and Stru<sup>high</sup>Num<sup>high</sup>  
255 (tumour) CA groups (**Fig. 4i, j**). Stru<sup>high</sup>Num<sup>low</sup> CA tumours grew more aggressively at the edge, while  
256 Stru<sup>high</sup>Num<sup>high</sup> CA tumours expanded predominantly in the core (**Fig. 4k**). While Stru<sup>low</sup>Num<sup>low</sup> CA  
257 was the most frequent subtype among large tumours across both regions, neither Num CA (in tumour  
258 cores) nor Stru CA (at the edge) alone stratified tumour size (**Fig. 4l**). About 40% of tumours within  
259 the Stru<sup>high</sup>Num<sup>high</sup> CA group exhibited nodal involvement greater than N1 in the core (**Fig. 4m**),  
260 consistent with increased metastatic potential. Many tumours with Stru<sup>low</sup>Num<sup>low</sup> CA status at the  
261 edge required nodal clearance, suggesting local aggressiveness independently of global CA burden  
262 (**Fig. 4n**). Identified germline variants were exclusively observed in the Stru<sup>high</sup>Num<sup>high</sup> CA group (**Fig.**  
263 **4o**), consistent with associations between CA and BRCA1-driven genomic instability<sup>52</sup>. We also  
264 evaluated the relationship between CA subtypes and body composition. Tumours with Stru<sup>low</sup>Num<sup>high</sup>  
265 CA in the core were associated with increased fat-free mass index (FFMI), (**Supplementary Fig.**  
266 **5a**), while fat mass index (FMI), waist circumference, and weight showed no differences across  
267 groups or regions (**Supplementary Fig. 5b-d**). Taller patients more frequently exhibited  
268 Stru<sup>low</sup>Num<sup>low</sup> CA status in the edge regions, though this association was not observed in cores  
269 (**Supplementary Fig. 5e**). Collectively, these findings indicate that spatial patterns of CA are linked

270 to tumour clinical parameters. Stru<sup>high</sup>Num<sup>high</sup> CA tumours define a high-risk subgroup marked by  
271 large size, nodal spread, and genomic alterations, while Stru<sup>low</sup>Num<sup>low</sup> CA tumours—despite low CA  
272 burden—can display local aggressiveness.

273 Single-cell analyses revealed further divergent CA subtype dynamics. At the edge regions, Num CA  
274 decreased with increasing tumour size (**Supplementary Fig. 6a**), while Stru CA was enriched in  
275 poorly differentiated but smaller tumours (**Supplementary Fig. 6b**). In the core, the proportion of  
276 cells harbouring either CA subtype increased with both total and invasive tumour size  
277 (**Supplementary Fig. 6c, d**). Num CA showed no significant differences across histological tumour  
278 grades, although it was more prevalent in well-differentiated tumours (**Supplementary Fig. 6c**). Stru  
279 CA levels were enriched in poorly differentiated tumours (**Supplementary Fig. 6d**). No significant  
280 subtype-specific CA differences were observed across histological subtypes (**Supplementary Fig.**  
281 **6a-d**), but cells from mixed tumours harboured more centrosomes in both compartments  
282 (**Supplementary Fig. 7**), highlighting elevated Num CA as linked to increased tumour heterogeneity.  
283 Together, these findings indicate a dynamic evolution of CA during tumour progression, with early-  
284 stage tumours characterised by numerical defects and advanced tumours accumulating structural  
285 abnormalities.

286 Finally, we explored CA patterns in the context of hormone receptor status, a key clinical determinant  
287 in breast cancer<sup>53, 54</sup>. Stru CA levels varied by receptor status in both edge and tumour core  
288 compartments (**Supplementary Fig. 8a**). HER2<sup>-</sup> tumours had the highest Stru CA at the edge, while  
289 PR<sup>-</sup> and ER<sup>-</sup> tumours had the lowest. In contrast, ER<sup>-</sup> tumours displayed elevated Stru CA in the  
290 core, with HER2<sup>+</sup> tumours showing the lowest levels. Num CA also showed compartment-specific  
291 trends: HER2<sup>+</sup> tumours exhibited high Num CA at the edge but low levels in the core; ER<sup>-</sup> tumours  
292 showed the inverse pattern (**Supplementary Fig. 8b**). Despite broadly similar spatial trends  
293 between Stru CA and Num CA across receptor-defined subtypes, HER2<sup>+</sup> tumours emerged as an  
294 exception (**Supplementary Fig. 8c, d**). HER2<sup>moderate</sup> (2+) tumours had significantly lower Num CA

295 levels than HER2<sup>high</sup> (3+) tumours in both compartments (**Supplementary Fig. 8e**), suggesting  
296 HER2 dosage impacts centrosome number. HER2<sup>+</sup> tumours were overrepresented in discordant CA  
297 phenotypes—Stru<sup>low</sup>Num<sup>high</sup> (28.6%) and Stru<sup>high</sup>Num<sup>low</sup> (35.7%)—at the edge (**Supplementary Fig.**  
298 **8f**), indicating HER2 may differentially regulate centrosome structure and number depending on  
299 spatial context. This distribution was not observed in the tumour core. These observations point to  
300 spatially resolved, hormone receptor-specific influence of CA patterns, particularly via HER2  
301 signalling, in a microenvironment-dependent manner.

302

### 303 **CenSegNet reveals CA patterns linked to tumour subtype and progression**

304 Using CenSegNet-derived spatial quantifications, we compared dynamic shifts in Stru CA and Num  
305 CA between tumour edges and cores. Patients were classified based on whether each CA subtype  
306 was more abundant in the tumour core (T>E) or at the edge (E>T). Baseline clinicopathological  
307 features were similar between spatial groups (**Supplementary Fig. 9**). Stru CA<sup>T>E</sup> tumours were  
308 associated with lobular histology (15.8%), lower histological tumour grade (26.3% grade 3), and  
309 lower HER2<sup>+</sup> prevalence (15.8%). In contrast, Stru CA<sup>E>T</sup> tumours more frequently displayed high  
310 tumour grade morphology (59.1% grade 3), higher HER2<sup>+</sup> status (immunohistochemistry 3+ or FISH-  
311 confirmed) (31.8%), and higher differentiation (**Supplementary Fig. 10a, b**). Single-cell analysis  
312 revealed that in both Stru CA spatial groups, Num CA-positive cells consistently contained more  
313 centrosomes in the core than at the edge (**Fig. 5a, b**), indicating a conserved numerical asymmetry.  
314 Num CA spatial prevalence was also associated with aggressiveness: Num CA<sup>T>E</sup> tumours were  
315 more often grade 3 (40.6%), had greater nodal involvement (31.2%), and lower HER2<sup>+</sup> frequency  
316 (12.5%) compared with Num CA<sup>E>T</sup> tumours (28.6%, 7.1%, and 35.7%, respectively;  
317 **Supplementary Fig. 10c, d**). Mixed histology was more common in Num CA<sup>T>E</sup> tumours (12.5%),  
318 whereas Num CA<sup>E>T</sup> tumours were associated with lobular carcinomas (28.6% versus 6.3%), a  
319 subtype typically linked to slower growth and smaller size. Across both Num CA<sup>T>E</sup> Num CA<sup>E>T</sup>  
320 groups, centrosomes were significantly larger at the edge than in the core. Notably, Num CA-positive

321 cells with larger centrosomes harboured fewer of them (**Fig. 5c, d**), revealing a robust inverse  
322 relationship between centrosome size and number across all spatial groups (**Fig. 5e**). These findings  
323 highlight distinct spatial CA patterns associated with tumour subtype and behaviour. Tumours  
324 displaying a shift from fewer, larger centrosomes at the periphery to smaller, more numerous  
325 centrosomes in the core exhibit features of increased aggressiveness. These spatial patterns,  
326 resolved through CenSegNet, provide a proxy for tumour heterogeneity and may reflect the  
327 evolutionary trajectory of CA during tumour progression.

328

329 **CenSegNet integration and accessible deployment for high-throughput centrosome analysis**  
330 To support broad adoption and integration into diverse analytical workflows, we provide CenSegNet  
331 as both a lightweight application programming interface (API) and an interactive graphical user  
332 interface (GUI) (**Fig. 6a, b**). The GUI comprises three modules: a data upload panel for inputting  
333 whole-slide or high-resolution images; a prediction module for adjustable inference parameters  
334 including probability thresholds and region selection for optimised segmentation of centrosomes and  
335 epithelial compartments; and an export tool for structured outputs. Post-inference, users can retrieve  
336 per-cell pixel-resolved size estimates, spatial coordinates, and centrosome counts within tissue  
337 context (**Fig. 6b**). The interface supports batch processing and accepts both immunohistochemistry  
338 and immunofluorescence formats. Benchmarking shows that CenSegNet processes a 6,000 ×  
339 6,000-pixel immunohistochemistry image or a 2,048 × 2,048-pixel immunofluorescence image in <1  
340 min on standard GPUs, enabling rapid, scalable analysis across large datasets.

341

## 342 Discussion

343 We introduce CenSegNet, a modular deep learning framework for quantification of centrosomes in  
344 epithelial tissues at spatial and single-cell resolution. Unlike previous approaches that treat  
345 morphological analysis and segmentation as separate tasks, CenSegNet integrates centrosome  
346 detection and phenotyping, nuclear localisation, and epithelial boundary inference into a unified,  
347 multichannel pipeline. This enables context-aware segmentation of both structural (Stru CA) and  
348 numerical (Num CA) abnormalities across diverse imaging modalities. CenSegNet outperforms  
349 established models—including U-Net, SegNet, and DeepLabv3+—in accuracy, generalisability, and  
350 morphological fidelity, particularly in densely packed or morphologically heterogeneous tissue  
351 regions. The framework comprises three specialised modules: a YOLOv11-based detector trained  
352 on over 15,000 annotated centrosomes for robust localisation; a U-Net model for precise  
353 segmentation of centrosome area and morphology; and a StarDist-based cell segmentation module  
354 optimised for delineating epithelial boundaries in complex tissue architectures. This modular  
355 integration, coupled with uncertainty-aware postprocessing, enables systematic and standardised  
356 phenotyping of CA subtypes at single-cell and spatial resolution. Applied to 911 sample cores from  
357 127 patients, CenSegNet-based profiling of over 330,000 centrosomes reveals previously  
358 uncharacterised spatial trajectories and clinical correlates of Stru CA and Num CA, uncovering their  
359 mechanistic uncoupling, age-dependent modulation, and associations with tumour progression,  
360 hormone receptor status, HER2 expression, and genomic alterations.

361 Computational tools have been developed previously for CA quantification in epithelial cells. A semi-  
362 automated machine learning pipeline quantified PCM defects and numerical CA in  
363 immunofluorescence images of normal and breast cancer cells<sup>41</sup>, but the method offers limited  
364 spatial resolution and required extensive manual curation. Another semi-automated approach linked  
365 centrosome size and number to chromosomal instability in human breast cancer tissues<sup>17</sup>, yet lacked  
366 single-cell resolution and scalability. Another pipeline assessed centriole number and length,  
367 revealing structural defects arising from fragmentation and ectopic procentriole formation<sup>37</sup>, but

368 without spatially resolved quantification. An automated detection algorithm for high-throughput  
369 mapping of CA in ovarian cancer tissues, identified heterogeneous CA phenotypes associated with  
370 chromosomal instability and chemotherapy resistance<sup>44</sup>. However, this method did not distinguish  
371 between structural and numerical CA or provide single-cell semantic segmentation. CenSegNet  
372 addresses these limitations by integrating cellular, spatial, and clinical dimensions of CA across  
373 large-scale tissue cohorts. In doing so, it offers new insights into centrosome biology and the  
374 functional relevance of CA heterogeneity in breast cancer, with implications for risk stratification and  
375 precision oncology. Beyond breast cancer, CenSegNet has demonstrated generalisability across  
376 diverse epithelial tissues—including kidney, colon, and appendix—characterised by high stromal  
377 content and architectural complexity (**Supplementary Fig. 11**). To facilitate broad adoption, we  
378 provide CenSegNet as an open-source GUI enabling scalable extraction of structured, spatially  
379 anchored centrosome metrics. This will allow researchers and clinicians, regardless of computational  
380 expertise, to integrate centrosome profiling into histopathological and biomarker discovery  
381 workflows. Thus, CenSegNet extends beyond methodological innovation to practical application,  
382 accelerating systematic investigation of centrosome biology across anatomically and histologically  
383 diverse tissues, and enabling the identification of CA-driven vulnerabilities with potential therapeutic  
384 relevance.

385 Recent studies using a composite centrosome amplification score (CAS) that integrates both  
386 numerical and structural abnormalities<sup>7, 43</sup>, showed a progressive increase in CAS from normal  
387 breast tissue to invasive carcinoma. However, it did not distinguish the individual contributions of  
388 numerical and structural centrosome abnormalities. Our spatially resolved single-cell analysis  
389 demonstrates that Num CA and Stru CA represent distinct phenotypic axes of centrosome  
390 dysregulation. Although they frequently co-occur in tumour tissues, they are uncorrelated at both  
391 tissue and single-cell levels and exhibit unique spatial distributions: Stru CA is enriched at tumour  
392 edges, while Num CA predominates in tumour cores, suggesting that different regional pressures  
393 drive centrosome overduplication *versus* structural enlargement. Single-cell data further uncover a

394 robust inverse relationship between centrosome number and size—cells with multiple centrosomes  
395 tend to have smaller ones—indicating a compensatory constraint on total centrosome volume.  
396 Consistent with this, centriole over-elongation can induce CA *via* fragmentation and ectopic  
397 procentriole formation in breast cancer cells<sup>37</sup>, yet not all centrioles within a cell exhibit these  
398 changes, highlighting intra-cellular heterogeneity in elongation susceptibility. Collectively, these  
399 findings support a model in which Num CA and Stru CA are mechanistically uncoupled, evolving  
400 along orthogonal spatial gradients during tumour progression. This challenges the monolithic view  
401 of CA and instead portrays it as a dynamic, regionally modulated process shaped by local  
402 microenvironmental cues. The mechanistic decoupling of CA has important implications for  
403 understanding its functional heterogeneity in cancer and underscores the need for spatially informed  
404 biomarkers.

405 CenSegNet-based profiling reveals that aging exerts distinct and spatially patterned effects on Stru  
406 CA and Num CA in breast cancer. Num CA peaks between ages 50 and 70—overlapping with the  
407 menopausal transition and the most common window for breast cancer diagnosis<sup>55-57</sup>—whereas Stru  
408 CA accumulates progressively after age 70, indicating a later-life trajectory of centrosome  
409 architectural dysregulation. While CA overall increases with age—including in normal tissues—our  
410 data suggest that age is not a deterministic initiator but rather a factor that modulates the magnitude  
411 and spatial distribution of CA subtypes. In mid-life patients, Num CA is preferentially enriched in  
412 tumour cores—regions typically characterised by high proliferation—whereas Stru CA in older  
413 individuals extends more diffusely, often into tumour margins, likely reflecting age-associated  
414 changes in epithelial architecture, tissue repair dynamics, and microenvironmental stress. A study in  
415 prostate cancer has reported elevated CA in patients over 53 years of age<sup>7</sup>. CA increases with age,  
416 in normal breast epithelial cells derived from individuals aged 20-80, treated with DNA damage-  
417 inducing stimuli<sup>51</sup>. Chronic centrosome overduplication in aging mouse models of intestinal cancer  
418 drives aneuploidy and spontaneous tumorigenesis, supporting a role for age-associated CA in early  
419 malignant transformation<sup>18</sup>. Centrosome function deteriorates with age, evidenced by accumulation

420 of structural defects and impaired mitotic fidelity<sup>58, 59</sup>. Aging epithelial cells exhibit centrosome  
421 fragmentation, aberrant centriole elongation, and altered PCM composition<sup>59, 60</sup>, while broader  
422 declines in DNA repair and chromosomal segregation fidelity likely contribute to a permissive  
423 environment for CA<sup>59</sup>. Thus, aging tissues accumulating centrosome abnormalities, mitotic fidelity  
424 defects, and weakened genomic surveillance, may be more vulnerable to CA-driven tumorigenesis.  
425 Our study delineates the divergence of CA subtypes across the lifespan. The inverse scaling  
426 between centrosome number and size supports the idea that Num CA and Stru CA act as  
427 compensatory, rather than co-occurring, phenotypes. Notably, Stru CA in elderly patients is  
428 frequently associated with hormone receptor-negative tumours, implicating centrosome structural  
429 dysregulation in the biology of more aggressive or dedifferentiated late-life cancers. CA can induce  
430 breast cancer cell dedifferentiation and intrinsically drive high-grade tumours<sup>61</sup>. Age remains a key  
431 prognostic factor in breast cancer, influencing tumour subtype distribution, hormone receptor status,  
432 and genomic instability<sup>62, 63</sup>. Together, these results establish the first clinically and spatially resolved  
433 framework for understanding how aging modulates centrosome biology in human cancer. This  
434 framework lays the foundation for developing age-stratified biomarkers and clarifies why specific CA  
435 subtypes and their associated chromosomal instability may emerge more frequently or have greater  
436 clinical impact at distinct stages of life.

437 The spatial heterogeneity of CA and its clinical relevance for cancer remain poorly defined<sup>3, 7</sup>.  
438 CenSegNet systematically maps Stru CA and Num CA across tumour compartments and stratifies  
439 tumours into composite CA subtypes with distinct spatial, biological, and clinical profiles.  
440 Stru<sup>high</sup>Num<sup>high</sup> CA tumours are consistently associated with larger size, greater nodal involvement,  
441 and germline genomic alterations. Elevated CA in both core and edge regions suggests a cellular  
442 composition primed for proliferative expansion and invasive dissemination. Stru<sup>low</sup>Num<sup>low</sup> CA  
443 tumours often required nodal clearance, supporting evidence that even modest CA can drive  
444 aggressive behaviour in permissive genomic contexts—particularly when p53 surveillance is  
445 compromised<sup>7, 18, 64</sup>. CenSegNet also uncovers spatial discordance in CA subtypes, with

446 Stru<sup>high</sup>Num<sup>low</sup> and Stru<sup>low</sup>Num<sup>high</sup> CA profiles enriched at the invasive front. These spatial signatures  
447 are associated with enhanced metastatic potential, consistent with models in which centrosome  
448 abnormalities promote invasion through both cell-autonomous<sup>65-67</sup> and non-cell-autonomous<sup>68-70</sup>  
449 mechanisms. Structural centrosome defects have been shown to drive cell extrusion, facilitating  
450 invasion<sup>67, 71</sup>. Our findings corroborate longitudinal studies of tumour progression, such as in Barrett's  
451 oesophagus, where CA appears early in premalignant lesions and expands with p53 loss<sup>64</sup>,  
452 supporting a role for CA in tumour initiation rather than as a mere by-product of transformation.  
453 Similarly, CA was shown to increase from normal tissue to ductal carcinoma *in situ* (DCIS) to invasive  
454 carcinoma, and to correlate with recurrence and poor prognosis<sup>7, 61</sup>. Finally, our observation that Stru  
455 CA and Num CA spatial patterns are uncoupled from systemic physiological metrics reinforces CA  
456 as a tumour-intrinsic hallmark. Together, these results advance our understanding of centrosome  
457 biology in cancer and highlight the power of CenSegNet-driven integration of subcellular organelle  
458 features into spatial pathology, with direct implications for clinical decision-making, prognostic  
459 modelling, and therapeutic targeting.

460 By integrating CenSegNet-based centrosome phenotyping with spatially resolved hormone receptor  
461 profiling, we uncover compartment-specific associations between CA subtypes and ER/HER2  
462 expression. Num CA is selectively enriched in ER<sup>-</sup> tumours within the tumour core, whereas HER2<sup>+</sup>  
463 tumours display lower Num CA in the core despite elevated levels at the tumour edge. This spatial  
464 divergence suggests that hormone receptor signalling modulates centrosome number and structure  
465 in a regionally distinct manner. These observations corroborate previous studies linking CA to  
466 hormone receptor status and tumour aggressiveness<sup>37, 42, 66, 72</sup> and support the hypothesis that ER  
467 loss promotes CA through transcriptional or post-translational dysregulation of centriole biogenesis  
468 pathways. Basal-like ER<sup>-</sup>PR<sup>-</sup>HER2<sup>-</sup> breast cancers—characterised by genomic instability and poor  
469 prognosis<sup>73</sup>, frequently exhibit high CA, often with centriole fragmentation and ectopic procentriole  
470 formation driven by over-elongation<sup>37</sup>. These defects recruit excess pericentriolar material (PCM),  
471 generating supernumerary or structurally abnormal MTOCs that drive mitotic errors and

472 chromosomal instability<sup>74, 75</sup>. In contrast, HER2<sup>+</sup> tumours are more frequently enriched in discordant  
473 CA subtypes (Stru<sup>low</sup>Num<sup>high</sup> and Stru<sup>high</sup>Num<sup>low</sup>) at the tumour edge, suggesting that HER2 signalling  
474 may differentially regulate centrosome number and structure depending on spatial context. This is  
475 consistent with evidence that both PLK4 and AURKA expression—key regulators of centriole  
476 biogenesis and maturation, respectively—is differentially influenced by HER2 status<sup>72, 76</sup>. Despite  
477 high proliferative capacity, HER2<sup>+</sup> tumours exhibit lower overall CA, raising the possibility that these  
478 tumours suppress CA to preserve mitotic fidelity or evade immune detection. Together, our findings  
479 identify spatially distinct and mechanistically diverse relationships between hormone receptor  
480 signalling and centrosome biology. ER loss is associated with elevated Num CA and mitotic instability  
481 in tumour cores, while HER2 signalling appears to exert compartmentalised control over CA subtype  
482 distribution. Spatially resolved centrosome profiling thus provides a framework for identifying  
483 hormone-specific vulnerabilities that could inform targeted breast cancer therapies.

484 CenSegNet enables fine-grained stratification of tumours based on the relative abundance of  
485 structural (Stru CA) and numerical (Num CA) centrosome abnormalities, revealing distinct spatial–  
486 biological associations. Tumours with Stru CA enriched at the invasive edge (Stru CA<sup>E>T</sup>) are more  
487 proliferative, more frequently HER2<sup>+</sup>, and exhibit higher differentiation, whereas those with Stru CA  
488 more abundant in the core (Stru CA<sup>T>E</sup>) are more commonly associated with lobular histology and  
489 slower growth. Single-cell analyses further show that tumours with grade 3 histology, nodal  
490 involvement, and mixed subtypes—hallmarks of aggressive disease—are more prevalent in the Num  
491 CA<sup>T>E</sup> subgroup, while Num CA<sup>E>T</sup> tumours are enriched for lobular carcinomas, typically linked to  
492 indolent behaviour<sup>73</sup>. These spatially resolved patterns corroborate previous reports suggesting that  
493 numerical centrosome abnormalities increase with tumour progression and are more frequent in  
494 aggressive basal-like carcinomas<sup>7, 37</sup>. Our single-cell data also reveal an inverse relationship  
495 between centrosome number and size—where edge regions harbour fewer but larger  
496 centrosomes—supporting a model in which Stru CA at the periphery primes cells to acquire invasive  
497 behaviours, while Num CA in the core drives proliferation and genomic instability. This dynamic

498 interplay likely reflects microenvironmental influences on CA trajectories during tumour progression.  
499 These findings extend the “CA set point” concept<sup>77</sup>, which postulates that tumours maintain a  
500 context-dependent equilibrium of CA phenotypes to balance proliferation, invasion and survival.  
501 CenSegNet-based spatial profiling demonstrates that CA is not only subtype-specific but also  
502 spatially regulated, providing new insights into the architectural evolution and heterogeneity of breast  
503 cancer.  
  
504 In summary, CenSegNet delivers the first fully integrated and spatially resolved framework for  
505 profiling CA at single-cell resolution across large-scale human cancer tissues. By enabling precise,  
506 high-throughput quantification of both numerical and structural CA phenotypes, CenSegNet  
507 uncovers distinct mechanistic, temporal, and spatial trajectories of centrosome dysregulation in  
508 breast cancer. The discovery that numerical and structural CA are decoupled numerical and  
509 structural CA are decoupled—not only in cellular architecture but also in their associations with  
510 clinical features such as tumour grade, hormone receptor status, germline mutation, and patient  
511 age—advances our understanding of how centrosome abnormalities contribute to intratumoral  
512 heterogeneity, tumour progression, with considerable implications for therapeutic resistance.  
513 Importantly, the identification of compensatory dynamics between centrosome number and size, and  
514 their divergent distributions across tumour cores and margins, points to context-specific roles in  
515 modulating local tumour ecology and genomic instability. These insights challenge the longstanding  
516 view of CA as a uniform driver of malignancy, instead positioning CA subtypes as distinct functional  
517 modules in tumour evolution. CenSegNet thus provides a foundation for developing CA-based  
518 biomarkers to stratify patients by tumour subtype, age, and aggressiveness, and opens opportunities  
519 for therapies targeting CA-driven vulnerabilities. Given the availability of PLK4, AURKA, and HSET  
520 inhibitors<sup>78-82</sup>, spatial CA maps could guide personalised strategies, particularly in tumours with  
521 discordant amplification phenotypes. Future studies integrating CenSegNet-based CA profiling with  
522 transcriptomic and proteomic analyses will be essential to uncover the molecular drivers of spatial

523 CA dynamics and clarify their roles in tumour progression, metastatic dissemination, and therapy  
524 resistance.

525 **Methods**

526

527 **Ethics and human breast tissues**

528 The study participants were a subgroup of women diagnosed with early breast cancer who were  
529 recruited to a single-centre prospective observational cohort study at University Hospitals  
530 Southampton, "Investigating Outcomes from Breast Cancer: Correlating Genetic, Immunological and  
531 Nutritional Predictors ([BeGIN](#))<sup>83, 84</sup>." All procedures performed in studies involving human participants  
532 were in accordance with the ethical standards of the institutional and/or national research committee  
533 and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.  
534 All participants in BeGIN gave written informed consent. The research ethics committee approved  
535 the study (Research Ethics Committee (REC) - Cambridgeshire and Hertfordshire reference number:  
536 14/EE/1297). Women were eligible for the BeGIN study if they were aged >18 years and diagnosed  
537 with invasive breast cancer or DCIS at University Hospital Southampton after May 2015. Linked  
538 anonymised patient information, including patient characteristics, tumour characteristics and clinical  
539 management, were extracted from the hospital electronic record system. Body composition  
540 parameters were measured using Bioelectrical Impedance Analysis (BIA) with a phase-sensitive, 8-  
541 electrode device (Seca mBCA515)<sup>85</sup>. To conduct this study, 911 cores from 127 breast cancer  
542 patients were used. The TMAs were constructed from formalin fixed paraffin embedded (FFPE)  
543 histopathology tissue blocks from surgical treatment surplus to diagnostic requirements. Colon,  
544 kidney, and appendix tissue were incorporated into breast tissue TMAs to facilitate orientation during  
545 sectioning and analysis. Colon and kidney samples consisted of histologically normal tissue, the  
546 status of which was independently verified by a board-certified pathologist. A summary of the  
547 clinicopathological information linked to the human breast samples used in this study is included in  
548 **Supplementary Table 1**. Anonymous data from the BeGIN study is available for request to  
549 researchers who provide a completed Data Sharing request form that describes a methodologically  
550 sound proposal, for the purpose of the approved proposal. Proposals will be reviewed by the study

551 steering committee. Data will be shared once all parties have signed relevant data sharing  
552 documentation, covering the study steering committee conditions for sharing and if required, an  
553 additional Data Sharing Agreement from the Sponsor.

554

### 555 **Ethics and mice**

556 BALB/c HER-2/neu transgenic mice (referred to as BALB-NeuT)<sup>86</sup> carrying the transforming rat *Her-2/neu* oncogene under control of a MMTV-LTR were used. All experimental procedures involving  
557 mice were approved by the University of Southampton Local Ethics Committee and registered with  
558 the Ethics and Research Governance Online II (ERGO II; ID: 65385). All animal work was conducted  
559 in accordance with UK Home Office regulations, adhering to the principles of the 3Rs (Replacement,  
560 Reduction, Refinement) and the ARRIVE (Animal Research: Reporting of In Vivo Experiments)  
561 guidelines to minimise animal suffering throughout the study. Mice were housed in a specific  
562 pathogen-free (SPF) facility under controlled environmental conditions, including regulated  
563 temperature and humidity, with a 12-hour light/dark cycle. Animals had ad libitum access to standard  
564 chow and water.

566

### 567 **Cell culture**

568 MCF10A is a non-transformed human mammary epithelial cell line (ATCC® CRL-10317). The  
569 MCF10A-PLK4 cell line is a genetically engineered derivative of MCF10A that enables inducible  
570 overexpression of Polo-like kinase 4 (PLK4), a master regulator of centrosome duplication whose  
571 upregulation induces centrosome amplification<sup>66</sup>. The MCF10A-PLK4 cell line was kindly provided  
572 by Dr. Susana Godinho (The Barts Cancer Institute, Queen Mary University of London). Both  
573 MCF10A and MCF10A-PLK4 cells were cultured in DMEM/F12 medium (Invitrogen), supplemented  
574 with 10% donor horse serum (Gibco, 31331028), 20 ng/ml human epidermal growth factor (EGF;  
575 Sigma, E9644), 10 µg/ml insulin (Sigma, I1882), 100 µg/ml hydrocortisone (Sigma, H0888), 1 ng/ml  
576 cholera toxin (Sigma, C8052), and 50 U/ml penicillin with 50 µg/ml streptomycin (Life Technologies).

577 Cells were maintained at 37 °C in a humidified incubator with 5% CO<sub>2</sub>. To induce PLK4  
578 overexpression, cells were treated with doxycycline (Sigma, D9891) at 2 µg/ml for 48 hours.

579

580 **Tissue microarray construction and immunohistochemistry**

581 Tissue microarrays (TMAs) were constructed using formalin-fixed, paraffin-embedded (FFPE) tissue  
582 samples obtained following breast cancer surgery from 127 patients diagnosed with primary invasive  
583 breast carcinoma at University Hospital Southampton between July 9, 2015, and January 31, 2019,  
584 and participating in the BeGIN study. All patients underwent standardised treatment at a single  
585 institution, consisting of surgery, followed by adjuvant treatments according to local and national  
586 protocols. Pathological evaluation of hormone receptor and HER2 expression was conducted  
587 according to established national and international guidelines<sup>87, 88</sup>. Immunohistochemistry was used  
588 to determine hormone receptor status, and in situ hybridization (ISH) was employed to confirm HER2  
589 positivity for tumours with an IHC score of 2+. All procedures were performed within the standard  
590 clinical diagnostic pathway. A total of 911 sample cores were systematically sampled from three  
591 distinct, pathologically classified regions for each patient: tumour tissue (Tumour), tumour margin  
592 (Edge), and tumour-free tissue (Normal). Each patient represented by three technical replicate cores  
593 from Tumour, Edge, and Normal regions, were procured for analysis. Each 0.6 mm in diameter and  
594 5 µm in thickness were extracted from formalin-fixed, paraffin-embedded specimens and arrayed  
595 into recipient tissue microarray (TMA) blocks. This tri-regional, triplicate-core sampling strategy was  
596 designed to provide a comprehensive and robust representation of the tissue heterogeneity within  
597 and around the tumour microenvironment. TMA sections were mounted onto TOMO® adhesion  
598 microscope slides. Immunohistochemistry was performed using the Dako Autostainer Link 48  
599 automated platform. Endogenous peroxidase activity was quenched using EnVision FLEX blocking  
600 reagent (Dako), followed by a 30-minute incubation with the primary antibody against pericentrin  
601 (1:500 dilution; Abcam, ab4448). Signal amplification and enzymatic detection were achieved using  
602 EnVision FLEX HRP (Dako, 20 minutes) and Rabbit Link (Dako, 15 minutes). Slides were

603 counterstained with haematoxylin following three 5-minute washes in 3-amino-9-ethylcarbazole  
604 (AEC).

605

606 **Assessment of hormone receptor status**

607 ER or PR expression were evaluated by immunohistochemistry and scored using the Allred system  
608 (range 0–8). Tumours with an Allred score  $\geq 3$  were classified as ER or PR positive. HER2 status  
609 was determined by immunohistochemistry and scored as 0, 1+, 2+, or 3+. Tumours with a score of  
610 3+ were classified as HER2 positive. Cases with a score of 2+ were considered equivocal were  
611 further assessed by fluorescence in situ hybridization (FISH). Tumours with FISH amplification were  
612 designed as HER2 positive, defined as a HER2/CEP17 ratio  $\geq 2.0$  or average HER2 copy number  $\geq$   
613 6.0 signals per cell.

614

615 **Immunofluorescence**

616 The following primary antibodies were used: anti-GT335 (1:800; Adipogen, AG-20B-0020-C100),  
617 anti-pericentrin (1:250; Abcam, ab4448), and anti-keratin 8/18 (KRT8/18; 1:300; Origene, BP5007).  
618 Secondary antibodies (Life Technologies) included goat anti-mouse (A-32723), anti-rabbit (A-11037  
619 and A-11008), and anti-guinea pig (A-21450), each conjugated to Alexa Fluor 488, Alexa Fluor 594,  
620 or Alexa Fluor 647, and used at a final concentration of 5  $\mu$ g/ml.

621 OCT-embedded mammary gland sections (30  $\mu$ m thick) from BALB-NeuT mice were cryosectioned,  
622 air-dried for 30 minutes, and fixed in 4% paraformaldehyde (PFA) for 20 minutes at room  
623 temperature. Sections were permeabilised for 45 minutes with 0.1% Triton X-100 in PBS, then  
624 blocked for 2 hours in a solution containing 2% bovine serum albumin (BSA), 5% foetal bovine serum  
625 (FBS; Gibco), and 0.1% Triton X-100 in PBS. Sections were incubated overnight at 4 °C with primary  
626 antibodies against pericentrin, GT335, and KRT8, followed by washing and incubation with the  
627 appropriate secondary antibodies for 2 hours at room temperature. Nuclei were counterstained with  
628 DAPI using Fluoroshield mounting medium (Sigma, F6057).

629 MCF10A and MCF10A-PLK4 cells were fixed in anhydrous methanol at -20 °C for 10 minutes,  
630 followed by permeabilisation with 0.1% Triton X-100 in PBS for 2 minutes. Cells were then washed  
631 three times for 5 minutes each with 0.1% Triton X-100 in PBS. Blocking was performed using 3%  
632 BSA in 0.1% Triton X-100 in PBS for 1 hour at room temperature. Cells were incubated overnight at  
633 4 °C with primary antibodies against pericentrin and GT335. After washing, cells were incubated with  
634 the appropriate secondary antibodies for 1 hour at room temperature and counterstained with DAPI  
635 using Fluoroshield.

636

### 637 **Microscopy and image annotation**

638 Immunohistochemistry images were acquired using a Zeiss Axio Imager Z1 upright microscope  
639 (Zeiss), equipped with an AxioCam MRc5 colour camera. Image capture was performed using Zeiss  
640 ZEN imaging software, following a predefined whole-slide brightfield scanning protocol (Whole Slide  
641 [WS] Brightfield [BF] fold-light) with a 20× objective lens. This configuration enabled high-resolution  
642 imaging of tissue sections with consistent illumination and contrast across samples.

643 Immunofluorescence images were captured using an inverted Leica STELLARIS 5 laser scanning  
644 confocal microscope (Leica Microsystems), equipped with a 40× oil immersion objective (HC PL  
645 APO 40×/1.30 Oil CS2). Z-stacks were acquired at 16-bit depth with a 0.2 μm step size, using fields  
646 of view (FOVs) ranging from 1024 × 1024 to 2048 × 2048 pixels. All cells from all FOVs obtained  
647 during the experiments were included in the evaluation of model performance. Image processing  
648 was performed using Fiji software (<https://imagej.net/software/fiji/>)<sup>89</sup>.

649 Manual annotations were performed by Jiaoqi Cheng (see **Supplementary Fig. 12**), who received  
650 training and supervision from Constantinos Savva. Centrosome size quantification in  
651 immunohistochemistry images was performed using calibrated spatial resolution, wherein each pixel  
652 corresponded to an area of 0.0483 μm<sup>2</sup>. For each centrosome, pixel-level segmentation masks were  
653 used to compute total pixel occupancy, which was then multiplied by the calibrated pixel area to  
654 derive centrosome size in μm<sup>2</sup>.

655 To generate the immunohistochemistry training dataset, images from 78 patients were included,  
656 comprising human breast cancer tissue, normal breast tissue, normal liver, and normal kidney  
657 samples. Within these images, annotations were made for centrosomes ( $n = 14,679$ ), epithelial  
658 compartments ( $n = 2,486$ ), and stromal compartments ( $n = 108$ ), using QuPath v0.5.1  
659 (<https://qupath.github.io/>)<sup>90</sup>. All centrosome annotations were performed under 200 $\times$  magnification  
660 for each region of interest (ROI). The manual annotation process required over 200 person-hours.  
661 Annotations were exported as GeoJSON files.

662 For the immunofluorescence training dataset, only 15 high-resolution image z-stacks of MCF10A  
663 and MCF10A-PLK4 cell lines were required. Annotation was performed using Cellpose 3.0<sup>91, 92</sup>.  
664 Datasets were first maximum-intensity projected, then split into contrast-adjusted single-channel  
665 images, with boundaries defined by edge features visible in the blue channel. The manual annotation  
666 process took approximately 16 person-hours. All annotations were exported as SVG files.

667

### 668 **Image processing and model training**

669 **Image processing.** For immunohistochemistry image processing in the CenSegNet pipeline, raw  
670 microscopy images were pre-processed to enhance contrast and suppress background noise,  
671 thereby improving the visibility of cellular structures. Each image was cropped into overlapping  
672 patches of 256  $\times$  256 pixels with a stride of 300 pixels. Patches exhibiting artefacts or poor quality  
673 were manually excluded to ensure dataset integrity. The final immunohistochemistry dataset  
674 comprised 1,122 annotated patches, each containing labelled information on centrosome location  
675 and size. For immunofluorescence images, RGB channels were converted to greyscale to reduce  
676 dimensionality and simplify the training process. This transformation allowed the model to remain  
677 invariant to colour information while improving computational efficiency.

678 **Model architecture and training.** CenSegNet employs a modular three-step architecture  
679 comprising approximately 40 million trainable parameters. The first step consists of a detection head  
680 that localises candidate ROIs, while the second stage performs fine-grained segmentation within

681 these ROIs to achieve precise spatial delineation. This decoupled design enables task-specific  
682 optimisation and improves memory efficiency and training stability. In the detection stage, we fine-  
683 tuned YOLOv11-seg model<sup>45</sup> (<https://github.com/ultralytics/ultralytics>) using our own training set,  
684 guided by a composite loss function comprising box loss, segmentation loss, classification loss, and  
685 distribution focal loss. These components respectively optimise object localisation, foreground-  
686 background separation, class prediction, and robustness to complex spatial distributions. Training  
687 was performed using the AdamW optimiser (learning rate = 0.002, momentum = 0.9) with a batch  
688 size of 16 over 300 epochs. The segmentation step 2 employed a U-Net architecture<sup>46</sup> with three  
689 input channels (RGB) and one output channel. Given the small size of centrosomes, segmentation  
690 was formulated as a binary classification task. The model was trained for 100 epochs using the  
691 Binary Cross Entropy with Logits loss function (BCEWithLogitsLoss). Optimisation was performed  
692 using RMSprop (learning rate =  $1 \times 10^{-4}$ , weight decay =  $1 \times 10^{-8}$ , momentum = 0.9), which stabilised  
693 training and improved convergence for small target structures. Training was conducted in Python  
694 (v3.10) using PyTorch (v2.1). We applied data augmentations, such as HSV colour jittering and  
695 geometric transformations, which can be easily implemented using the torchvision.transforms  
696 module in PyTorch. The detection head was trained independently for approximately 24 hours,  
697 followed by an additional 12 hours of training for the segmentation head. All experiments were run  
698 on four NVIDIA A100 GPUs (40 GB VRAM each) using PyTorch's Distributed Data Parallel (DDP)  
699 framework with NVIDIA Collective Communications Library (NCCL) backend for gradient  
700 synchronisation. Input batches were evenly partitioned across GPUs, with local gradient computation  
701 and synchronisation via NCCL's optimised collective communication, achieving near-linear scaling  
702 in throughput. Batch sizes were dynamically adjusted to maximise GPU utilisation while maintaining  
703 training stability.

704 **Comparative segmentation models.** To benchmark CenSegNet, we compared its performance  
705 against established segmentation models including U-Net<sup>46</sup>, SegNet<sup>48</sup>, and DeepLabv3+<sup>49</sup>. Each  
706 model was trained on uniformly sized  $256 \times 256$  image patches cropped from the original dataset.

707 Official implementations were used without architectural modifications. Comparative experiments  
708 were conducted with and without ImageNet pretraining to assess the impact of transfer learning.  
709 Model outputs were evaluated against ground truth segmentation masks using a composite loss  
710 function combining weighted binary cross-entropy and Dice loss, balancing pixel-wise accuracy with  
711 regional overlap. All models were trained under identical conditions: Adam optimiser (initial learning  
712 rate =  $1 \times 10^{-4}$ ), cosine learning rate decay, L2 regularisation (weight decay =  $1 \times 10^{-8}$ ), batch size of  
713 16, and early stopping based on validation loss plateauing. A total of 108 immunohistochemistry  
714 images were manually annotated and divided into training (n=80), validation (8), and test (n=20)  
715 datasets. An additional private dataset comprising 25 images was generated and manually curated  
716 for evaluation.

717 **Cell segmentation.** For cell nucleus segmentation step 3, we integrated StarDist<sup>47</sup> with the QuPath  
718 platform. StarDist offers state-of-the-art performance in dense and noisy biological environments.  
719 Following accurate nucleus segmentation, cell membrane boundaries were estimated using the  
720 spatial coordinates of nuclei as anchor points. Expansion thresholds of 3, 4, 5, and 6  $\mu\text{m}$  were tested  
721 and validated against KRT8-stained images to determine the optimal value for our dataset.

722

### 723 **Evaluation metrics**

724 **Centrosome segmentation performance.** Centrosome segmentation performance was evaluated  
725 on a test set comprising 25 immunohistochemistry and 17 immunofluorescence images. Metrics  
726 included precision, recall, intersection over union (IoU), and F1 score. IoU was defined as the  
727 common area between the predicted segmentation and the ground truth:

$$728 \quad IoU = \frac{|\text{Prediction} \cap \text{Groundtruth}|}{|\text{Prediction} \cup \text{Groundtruth}|}$$

729 The F1 score, representing the harmonic mean of precision and recall, was calculated as:

$$730 \quad F1 \text{ score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

731 With:

$$732 \quad Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

733 Where, TP, FP, and FN denote true positives, false positives, and false negatives, respectively. Both  
734 IoU and F1 scores range from 0 to 1, with higher values indicating superior segmentation  
735 performance.

736 These metrics provide a comprehensive assessment of spatial overlap, agreement, and pixel-wise  
737 accuracy, with higher values indicating better performance in medical image segmentation.

738 **Epithelial segmentation performance.** To assess the performance of epithelial segmentation by  
739 CenSegNet, we quantified the  $F1^{\text{IoU}50}$  metric<sup>93</sup>, a widely adopted benchmark in biomedical image  
740 analysis. IoU50 means the IoU threshold was set to 0.5 (or 50%). A predicted bounding box was  
741 considered a correct detection if: IoU with ground truth  $\geq 0.5$ .

742

### 743 Statistical analysis

744 The exact n is stated in the corresponding figure legend. GraphPad Prism 10.3.1 (GraphPad  
745 Software) was used to perform statistical significance analysis. Normality and lognormality were  
746 assessed prior to statistical analysis. For datasets exhibiting a normal distribution, comparisons  
747 between two groups were conducted using an unpaired t-test, while comparisons across multiple  
748 groups were performed using one-way analysis of variance (ANOVA), followed by Tukey's post hoc  
749 test for multiple comparisons or two-way ANOVA, followed by Tukey's post hoc test for multiple  
750 comparisons. For datasets that deviated from normality, non-parametric testing was employed, using  
751 the Kruskal-Wallis test for multiple group comparisons, followed by the Mann-Whitney U test as a  
752 post hoc analysis. For correlation analysis, a two-tailed Pearson test was used followed by simple  
753 linear regression for graphical representation for correlation analysis. To compare the composition  
754 of groups based on categorical variables (e.g., histological tumour grade, T stage, histological  
755 tumour types, hormone status, nodal status, nodal clearance, HER2 status), Fisher's exact test was

756 employed. This non-parametric test was chosen for its appropriateness with count data and its ability  
757 to provide accurate p-values when expected cell counts are low (<5). All comparisons were two-  
758 sided. All values were presented as mean  $\pm$  s.e.m. For all statistical tests, \* $P \leq 0.05$ , \*\* $P \leq 0.01$ ,  
759 \*\*\* $P \leq 0.001$  and \*\*\*\* $P \leq 0.0001$  were considered significant.

760

761 **Assistance with manuscript preparation**

762 During the preparation of this manuscript, ChatGPT (OpenAI, version 4) was used to assist with  
763 stylistic and grammatical refinement. All AI-generated content was critically reviewed and edited by  
764 the authors, who take full responsibility for its accuracy and integrity.

765

766 **Data availability**

767 Examples of immunohistochemistry and immunofluorescence image datasets used for  
768 benchmarking and for testing CenSegNet within the demo version of the pipeline are available on  
769 <https://github.com/SKELab/CenSegNet/> and <https://zenodo.org/records/17131573>. All other  
770 relevant data supporting the key findings of this study are available within the article and its  
771 Supplementary Information files or upon reasonable request. A summary of the human breast  
772 samples used in this study are included in Supplementary Table 1. For participants on the BeGIN  
773 study, further donor anonymised clinicopathological information is available upon reasonable  
774 request, provided all relevant ethics approvals are in place (see “Ethics and human breast tissues”  
775 section for further details). The source data that support the findings in all Figures and  
776 Supplementary Figures are provided as a Source Data file within the paper. All reagents generated  
777 in this study are available upon reasonable request.

778

779 **Code availability**

780 The source code and software CenSegNet as a ready-to-use executable with a quickstart guide,  
781 example datasets and step-by-step procedures are freely available at  
782 <https://github.com/SKELab/CenSegNet/> and <https://zenodo.org/records/17131573>.

783 .

784

785 **Acknowledgements**

786 We would like to thank Dr. Mark Willett at the Imaging and Microscopy Centre for valuable assistance  
787 with fluorescence microscopy; Monette Lopez at the Research Histology Facility, University Hospital  
788 Southampton for performing immunohistochemistry staining and imaging; Dr Margaret Ashton-Key  
789 for classifying breast tumour compartments (normal, edge, core) and providing advice on  
790 immunohistochemistry staining conditions; Professor Susana Godinho for providing the MCF10A  
791 and MCF10A-PLK4 cell lines. In addition, we thank Dr. Xi Jia, Dr. Colinda Scheele, Dr. Marcin  
792 Przewloka, and Dr. Kif Liakath-Ali, for critically reviewing the manuscript. Cartoons in Figs. 1a, 1c  
793 and Supplementary Fig. 12 were created using <https://www.freepik.com/>. This work was supported  
794 by an MRC New Investigator Research Grant (MR/R026610/1) awarded to S.E. and an Institute for  
795 Life Sciences Pilot Grant awarded to R.C., S.B. and S.E. JC was supported by the Gerald Kerkut  
796 Charitable Trust and a China Scholarship Council (CSC) PhD studentship. The BeGIN study was  
797 supported by Breast Cancer Now.

798 **Author contributions**

799 J.C. designed and performed experiments, analysed, interpreted the data, and wrote the manuscript.  
800 K.F. trained models and developed CenSegNet. C.S. M.B and X.D. supervised centrosome  
801 annotation and performed transfer annotations from GeoJSON Files to PNG files and data analysis.  
802 M.G. performed immunofluorescence experiments. R.C. provided ethically approved BeGIN breast  
803 cancer tissues. S.B. supervised J.C. and interpreted the data. X.C. supervised K.F. on CenSegNet  
804 development. S.E. conceived and designed the project, analysed and interpreted the data,  
805 supervised J.C., wrote the manuscript, and acquired funding. All the authors provided intellectual  
806 input, edited, and approved the final manuscript.

807 **Competing interests**

808 C.S. reports involvement in a research collaboration with Proteotype and BioNTech. R.C. reports the  
809 following declaration: 1. Grants or Contracts: AstraZeneca, awarded an educational grant to the  
810 University of Southampton in November 2021 for the long-term follow-up of the POSH study. 2.  
811 Leadership or Fiduciary Roles: NICE, acts as a Breast Cancer Topic Advisor, contributing to guideline  
812 development. Association of Breast Surgery, Member of the clinical practice and standards  
813 committee. 3. Receipt of Equipment, Materials, etc.: SECA provided equipment for measuring body  
814 composition to University Hospital Southampton under a model industry collaborative agreement.

815

## 816 References

- 817 1. Conduit, P.T., Wainman, A. & Raff, J.W. Centrosome function and assembly in animal cells.  
818 *Nat Rev Mol Cell Biol* **16**, 611-624 (2015).
- 819 2. Kim, S. & Dynlacht, B.D. Assembling a primary cilium. *Curr Opin Cell Biol* **25**, 506-511  
820 (2013).
- 821 3. Godinho, S.A. & Basto, R. Centrosomes and cancer: balancing tumor-promoting and  
822 inhibitory roles. *Trends Cell Biol* **35**, 515-526 (2025).
- 823 4. Kiermaier, E., Stotzel, I., Schapfl, M.A. & Villunger, A. Amplified centrosomes-more than just  
824 a threat. *EMBO Rep* **25**, 4153-4167 (2024).
- 825 5. Nigg, E.A., Schnerch, D. & Ganier, O. Impact of Centrosome Aberrations on Chromosome  
826 Segregation and Tissue Architecture in Cancer. *Cold Spring Harb Symp Quant Biol* **82**,  
827 137-144 (2017).
- 828 6. Raff, J.W. & Basto, R. Centrosome Amplification and Cancer: A Question of Sufficiency.  
829 *Dev Cell* **40**, 217-218 (2017).
- 830 7. Mittal, K. et al. Centrosome amplification: a quantifiable cancer cell trait with prognostic  
831 value in solid malignancies. *Cancer Metastasis Rev* **40**, 319-339 (2021).
- 832 8. Boveri, T. Concerning the origin of malignant tumours by Theodor Boveri. Translated and  
833 annotated by Henry Harris. *J Cell Sci* **121 Suppl 1**, 1-84 (2008).
- 834 9. Godinho, S.A. & Pellman, D. Causes and consequences of centrosome abnormalities in  
835 cancer. *Philos Trans R Soc Lond B Biol Sci* **369** (2014).
- 836 10. Hsu, L.C., Kapali, M., DeLoia, J.A. & Gallion, H.H. Centrosome abnormalities in ovarian  
837 cancer. *Int J Cancer* **113**, 746-751 (2005).
- 838 11. Lingle, W.L., Lutz, W.H., Ingle, J.N., Maihle, N.J. & Salisbury, J.L. Centrosome hypertrophy  
839 in human breast tumors: implications for genomic stability and cell polarity. *Proc Natl Acad  
840 Sci U S A* **95**, 2950-2955 (1998).
- 841 12. Pihan, G.A. et al. Centrosome defects and genetic instability in malignant tumors. *Cancer  
842 Res* **58**, 3974-3985 (1998).
- 843 13. Sato, N. et al. Centrosome abnormalities in pancreatic ductal carcinoma. *Clin Cancer Res*  
844 **5**, 963-970 (1999).
- 845 14. Giehl, M. et al. Centrosome aberrations in chronic myeloid leukemia correlate with stage of  
846 disease and chromosomal instability. *Leukemia* **19**, 1192-1197 (2005).
- 847 15. Kramer, A., Neben, K. & Ho, A.D. Centrosome aberrations in hematological malignancies.  
848 *Cell Biol Int* **29**, 375-383 (2005).
- 849 16. Basto, R. et al. Centrosome amplification can initiate tumorigenesis in flies. *Cell* **133**, 1032-  
850 1042 (2008).
- 851 17. Lingle, W.L. et al. Centrosome amplification drives chromosomal instability in breast tumor  
852 development. *Proc Natl Acad Sci U S A* **99**, 1978-1983 (2002).
- 853 18. Levine, M.S. et al. Centrosome Amplification Is Sufficient to Promote Spontaneous  
854 Tumorigenesis in Mammals. *Dev Cell* **40**, 313-322 e315 (2017).
- 855 19. Chan, J.Y. A clinical overview of centrosome amplification in human cancers. *Int J Biol Sci*  
856 **7**, 1122-1144 (2011).
- 857 20. Shekhar, M.P., Lyakhovich, A., Visscher, D.W., Heng, H. & Kondrat, N. Rad6  
858 overexpression induces multinucleation, centrosome amplification, abnormal mitosis,  
859 aneuploidy, and transformation. *Cancer Res* **62**, 2115-2124 (2002).
- 860 21. Meraldi, P., Honda, R. & Nigg, E.A. Aurora-A overexpression reveals tetraploidization as a  
861 major route to centrosome amplification in p53-/- cells. *EMBO J* **21**, 483-492 (2002).
- 862 22. Fujiwara, T. et al. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-  
863 null cells. *Nature* **437**, 1043-1047 (2005).
- 864 23. Duensing, A., Chin, A., Wang, L., Kuan, S.F. & Duensing, S. Analysis of centrosome  
865 overduplication in correlation to cell division errors in high-risk human papillomavirus  
866 (HPV)-associated anal neoplasms. *Virology* **372**, 157-164 (2008).

- 867 24. Duelli, D.M., Hearn, S., Myers, M.P. & Lazebnik, Y. A primate virus generates transformed  
868 human cells by fusion. *J Cell Biol* **171**, 493-503 (2005).
- 869 25. Khodjakov, A. et al. De novo formation of centrosomes in vertebrate cells arrested during S  
870 phase. *J Cell Biol* **158**, 1171-1181 (2002).
- 871 26. La Terra, S. et al. The de novo centriole assembly pathway in HeLa cells: cell cycle  
872 progression and centriole assembly/maturation. *J Cell Biol* **168**, 713-722 (2005).
- 873 27. Davoli, T. & de Lange, T. Telomere-driven tetraploidization occurs in human cells  
874 undergoing crisis and promotes transformation of mouse cells. *Cancer Cell* **21**, 765-776  
875 (2012).
- 876 28. Borel, F., Lohez, O.D., Lacroix, F.B. & Margolis, R.L. Multiple centrosomes arise from  
877 tetraploidy checkpoint failure and mitotic centrosome clusters in p53 and RB pocket  
878 protein-compromised cells. *Proc Natl Acad Sci U S A* **99**, 9819-9824 (2002).
- 879 29. Karakaya, K. et al. Overexpression of EVI1 interferes with cytokinesis and leads to  
880 accumulation of cells with supernumerary centrosomes in G0/1 phase. *Cell Cycle* **11**, 3492-  
881 3503 (2012).
- 882 30. Dwivedi, D., Harry, D. & Meraldi, P. Mild replication stress causes premature centriole  
883 disengagement via a sub-critical Plk1 activity under the control of ATR-Chk1. *Nat Commun*  
884 **14**, 6088 (2023).
- 885 31. Loncarek, J., Hergert, P. & Khodjakov, A. Centriole reduplication during prolonged  
886 interphase requires procentriole maturation governed by Plk1. *Curr Biol* **20**, 1277-1282  
887 (2010).
- 888 32. Difilippantonio, M.J. et al. Nucleation capacity and presence of centrioles define a distinct  
889 category of centrosome abnormalities that induces multipolar mitoses in cancer cells.  
890 *Environ Mol Mutagen* **50**, 672-696 (2009).
- 891 33. Hut, H.M. et al. Centrosomes split in the presence of impaired DNA integrity during mitosis.  
892 *Mol Biol Cell* **14**, 1993-2004 (2003).
- 893 34. Loffler, H., Fechter, A., Liu, F.Y., Poppelreuther, S. & Kramer, A. DNA damage-induced  
894 centrosome amplification occurs via excessive formation of centriolar satellites. *Oncogene*  
895 **32**, 2963-2972 (2013).
- 896 35. Loncarek, J., Hergert, P., Magidson, V. & Khodjakov, A. Control of daughter centriole  
897 formation by the pericentriolar material. *Nat Cell Biol* **10**, 322-328 (2008).
- 898 36. Starita, L.M. et al. BRCA1-dependent ubiquitination of gamma-tubulin regulates  
899 centrosome number. *Mol Cell Biol* **24**, 8457-8466 (2004).
- 900 37. Marteil, G. et al. Over-elongation of centrioles in cancer promotes centriole amplification  
901 and chromosome missegregation. *Nat Commun* **9**, 1258 (2018).
- 902 38. Tang, C.J., Fu, R.H., Wu, K.S., Hsu, W.B. & Tang, T.K. CPAP is a cell-cycle regulated  
903 protein that controls centriole length. *Nat Cell Biol* **11**, 825-831 (2009).
- 904 39. Schmidt, T.I. et al. Control of centriole length by CPAP and CP110. *Curr Biol* **19**, 1005-1011  
905 (2009).
- 906 40. Burgy, L. et al. CenFind: a deep-learning pipeline for efficient centriole detection in  
907 microscopy datasets. *BMC Bioinformatics* **24**, 120 (2023).
- 908 41. Sankaran, D.G., Stemm-Wolf, A.J., McCurdy, B.L., Hariharan, B. & Pearson, C.G. A semi-  
909 automated machine learning-aided approach to quantitative analysis of centrosomes and  
910 microtubule organization. *J Cell Sci* **133** (2020).
- 911 42. Patel, N. et al. Integrated genomics and functional validation identifies malignant cell  
912 specific dependencies in triple negative breast cancer. *Nat Commun* **9**, 1044 (2018).
- 913 43. Mittal, K. et al. A Quantitative Centrosomal Amplification Score Predicts Local Recurrence  
914 of Ductal Carcinoma In Situ. *Clin Cancer Res* **26**, 2898-2907 (2020).
- 915 44. Sauer, C.M. et al. Molecular landscape and functional characterization of centrosome  
916 amplification in ovarian cancer. *Nat Commun* **14**, 6505 (2023).
- 917 45. Khanam, R. & Hussain, M. YOLoV11: An Overview of the Key Architectural Enhancements.  
918 *arXiv e-prints*, arXiv:2410.17725 (2024).

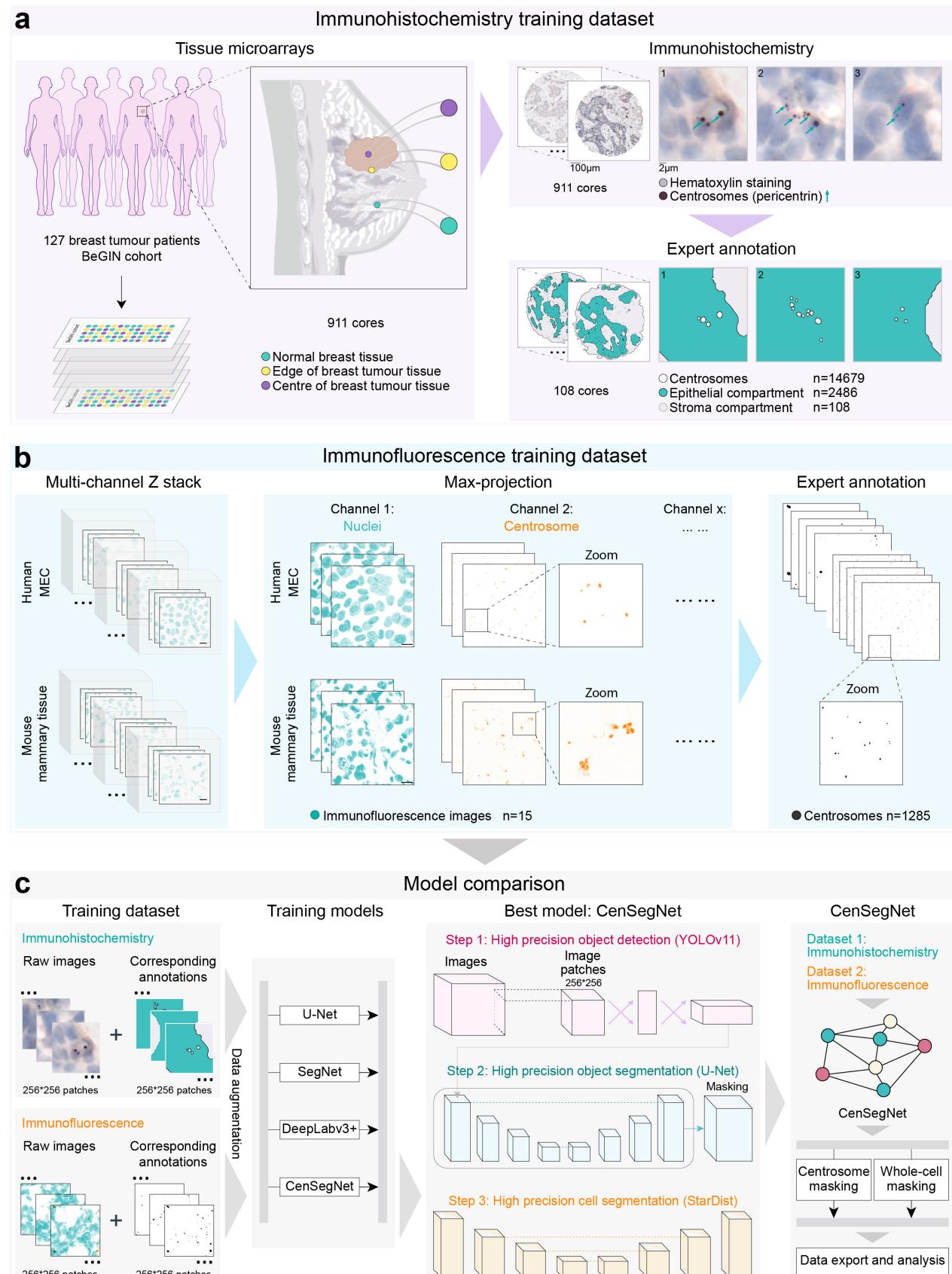
- 919 46. Ronneberger, O., Fischer, P. & Brox, T. in *Medical Image Computing and Computer-*  
920 *Assisted Intervention – MICCAI 2015*. (eds. N. Navab, J. Hornegger, W.M. Wells & A.F.  
921 Frangi) 234-241 (Springer International Publishing, Cham; 2015).
- 922 47. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. in *Medical Image Computing and*  
923 *Computer Assisted Intervention – MICCAI 2018*. (eds. A.F. Frangi, J.A. Schnabel, C.  
924 Davatzikos, C. Alberola-López & G. Fichtinger) 265-273 (Springer International Publishing,  
925 Cham; 2018).
- 926 48. Badrinarayanan, V., Kendall, A. & Cipolla, R. SegNet: A Deep Convolutional Encoder-  
927 Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and*  
928 *Machine Intelligence* **39**, 2481-2495 (2017).
- 929 49. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. arXiv:1802.02611 (2018).
- 930 50. He, L., Zhou, Y., Liu, L., Zhang, Y. & Ma, J. Application of the YOLOv11-seg algorithm for  
931 AI-based landslide detection and recognition. *Sci Rep* **15**, 12421 (2025).
- 932 51. Sridharan, D.M., Enerio, S., LaBarge, M.A., Stampfer, M.M. & Pluth, J.M. Lesion complexity  
933 drives age related cancer susceptibility in human mammary epithelial cells. *Aging (Albany*  
934 *NY)* **9**, 665-686 (2017).
- 935 52. Shimomura, A., Miyoshi, Y., Taguchi, T., Tamaki, Y. & Noguchi, S. Association of loss of  
936 BRCA1 expression with centrosome aberration in human breast cancer. *Journal of Cancer*  
937 *Research and Clinical Oncology* **135**, 421-430 (2009).
- 938 53. Agnoletto, A. & Brisken, C. Hormone Signaling in Breast Development and Cancer, in *A*  
939 *Guide to Breast Cancer Research: From Cellular Heterogeneity and Molecular*  
940 *Mechanisms to Therapy*. (eds. T. Sørlie & R.B. Clarke) 279-307 (Springer Nature  
941 Switzerland, Cham; 2025).
- 942 54. Loibl, S., Poortmans, P., Morrow, M., Denkert, C. & Curigliano, G. Breast cancer. *The*  
943 *Lancet* **397**, 1750-1769 (2021).
- 944 55. Menarche, menopause, and breast cancer risk: individual participant meta-analysis,  
945 including 118 964 women with breast cancer from 117 epidemiological studies. *The Lancet*  
946 *Oncology* **13**, 1141-1151 (2012).
- 947 56. Type and timing of menopausal hormone therapy and breast cancer risk: individual  
948 participant meta-analysis of the worldwide epidemiological evidence. *The Lancet* **394**,  
949 1159-1168 (2019).
- 950 57. Heer, E. et al. Global burden and trends in premenopausal and postmenopausal breast  
951 cancer: a population-based study. *The Lancet Global Health* **8**, e1027-e1037 (2020).
- 952 58. Delaval, B. & Doxsey, S.J. Pericentrin in cellular function and disease. *J Cell Biol* **188**, 181-  
953 190 (2010).
- 954 59. Jiang, T. et al. Drivers of centrosome abnormalities: Senescence progression and tumor  
955 immune escape. *Semin Cancer Biol* **110**, 56-64 (2025).
- 956 60. Nigg, E.A. & Raff, J.W. Centrioles, centrosomes, and cilia in health and disease. *Cell* **139**,  
957 663-678 (2009).
- 958 61. Denu, R.A. et al. Centrosome amplification induces high grade features and is prognostic of  
959 worse outcomes in breast cancer. *BMC Cancer* **16**, 47 (2016).
- 960 62. Van Herck, Y. et al. Is cancer biology different in older patients? *The Lancet Healthy*  
961 *Longevity* **2**, e663-e677 (2021).
- 962 63. Benz, C.C. Impact of aging on the biology of breast cancer. *Critical Reviews in*  
963 *Oncology/Hematology* **66**, 65-74 (2008).
- 964 64. Lopes, C.A.M. et al. Centrosome amplification arises before neoplasia and increases upon  
965 p53 loss in tumorigenesis. *J Cell Biol* **217**, 2353-2363 (2018).
- 966 65. Monteiro, P., Yeon, B., Wallis, S.S. & Godinho, S.A. Centrosome amplification fine tunes  
967 tubulin acetylation to differentially control intracellular organization. *EMBO J* **42**, e112812  
968 (2023).
- 969 66. Godinho, S.A. et al. Oncogene-like induction of cellular invasion from centrosome  
970 amplification. *Nature* **510**, 167-171 (2014).

- 971 67. LoMastro, G.M. & Holland, A.J. The Emerging Link between Centrosome Aberrations and  
972 Metastasis. *Dev Cell* **49**, 325-331 (2019).
- 973 68. Rhys, A.D. *et al.* Loss of E-cadherin provides tolerance to centrosome amplification in  
974 epithelial cancer cells. *J Cell Biol* **217**, 195-209 (2018).
- 975 69. Ganier, O. *et al.* Structural centrosome aberrations promote non-cell-autonomous  
976 invasiveness. *EMBO J* **37** (2018).
- 977 70. Garcia-Carpio, I. *et al.* Extra centrosomes induce PIDD1-mediated inflammation and  
978 immunosurveillance. *EMBO J* **42**, e113510 (2023).
- 979 71. Ganier, O., Schnerch, D. & Nigg, E.A. Structural centrosome aberrations sensitize polarized  
980 epithelia to basal cell extrusion. *Open Biol* **8** (2018).
- 981 72. Kahl, I. *et al.* The cell cycle-related genes RHAMM, AURKA, TPX2, PLK1, and PLK4 are  
982 associated with the poor prognosis of breast cancer patients. *J Cell Biochem* **123**, 581-600  
983 (2022).
- 984 73. Toft, D.J. & Cryns, V.L. Minireview: Basal-like breast cancer: from molecular profiles to  
985 targeted therapies. *Mol Endocrinol* **25**, 199-211 (2011).
- 986 74. Cosenza, M.R. *et al.* Asymmetric Centriole Numbers at Spindle Poles Cause Chromosome  
987 Misseggregation in Cancer. *Cell Rep* **20**, 1906-1920 (2017).
- 988 75. Ganem, N.J., Godinho, S.A. & Pellman, D. A mechanism linking extra centrosomes to  
989 chromosomal instability. *Nature* **460**, 278-282 (2009).
- 990 76. Ingebriktsen, L.M. *et al.* Elevated expression of Aurora-A/AURKA in breast cancer  
991 associates with younger age and aggressive features. *Breast Cancer Res* **26**, 126 (2024).
- 992 77. Wong, Y.L. *et al.* Cell biology. Reversible centriole depletion with an inhibitor of Polo-like  
993 kinase 4. *Science* **348**, 1155-1160 (2015).
- 994 78. Melichar, B. *et al.* Safety and activity of alisertib, an investigational aurora kinase A inhibitor,  
995 in patients with breast cancer, small-cell lung cancer, non-small-cell lung cancer, head and  
996 neck squamous-cell carcinoma, and gastro-oesophageal adenocarcinoma: a five-arm  
997 phase 2 study. *Lancet Oncol* **16**, 395-405 (2015).
- 998 79. Borisa, A.C. & Bhatt, H.G. A comprehensive review on Aurora kinase: Small molecule  
999 inhibitors and clinical trial studies. *Eur J Med Chem* **140**, 1-19 (2017).
- 1000 80. Veitch, Z.W. *et al.* Safety and tolerability of CFI-400945, a first-in-class, selective PLK4  
1001 inhibitor in advanced solid tumours: a phase 1 dose-escalation trial. *Br J Cancer* **121**, 318-  
1002 324 (2019).
- 1003 81. Lei, Q. *et al.* YLT-11, a novel PLK4 inhibitor, inhibits human breast cancer growth via  
1004 inducing maladjusted centriole duplication and mitotic defect. *Cell Death Dis* **9**, 1066  
1005 (2018).
- 1006 82. Watts, C.A. *et al.* Design, synthesis, and biological evaluation of an allosteric inhibitor of  
1007 HSET that targets cancer cells with supernumerary centrosomes. *Chem Biol* **20**, 1399-1410  
1008 (2013).
- 1009 83. Blyth, R.R.R. *et al.* Development and characterisation of a novel 3D in vitro model of  
1010 obesity-associated breast cancer as a tool for drug testing. *NPJ Breast Cancer* **11**, 50  
1011 (2025).
- 1012 84. Daly, A. *et al.* Assessment of body composition in breast cancer patients: concordance  
1013 between transverse computed tomography analysis at the fourth thoracic and third lumbar  
1014 vertebrae. *Front Nutr* **11**, 1366768 (2024).
- 1015 85. Bosy-Westphal, A. *et al.* What makes a BIA equation unique? Validity of eight-electrode  
1016 multifrequency BIA to estimate body composition in a healthy adult population. *Eur J Clin  
1017 Nutr* **67 Suppl 1**, S14-21 (2013).
- 1018 86. Rovero, S. *et al.* DNA vaccination against rat her-2/Neu p185 more effectively inhibits  
1019 carcinogenesis than transplantable carcinomas in transgenic BALB/c mice. *J Immunol* **165**,  
1020 5133-5142 (2000).
- 1021 87. Rakha, E.A. *et al.* UK recommendations for HER2 assessment in breast cancer: an update.  
1022 *J Clin Pathol* **76**, 217-227 (2023).

- 1023 88. Ellis, I.O., Rakha, E.A., Tse, G.M. & Tan, P.H. An international unified approach to reporting  
1024 and grading invasive breast cancer. An overview of the International Collaboration on  
1025 Cancer Reporting (ICCR) initiative. *Histopathology* **82**, 189-197 (2023).
- 1026 89. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods*  
1027 **9**, 676-682 (2012).
- 1028 90. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci  
1029 Rep* **7**, 16878 (2017).
- 1030 91. Stringer, C. & Pachitariu, M. Cellpose3: one-click image restoration for improved cellular  
1031 segmentation. *Nat Methods* **22**, 592-599 (2025).
- 1032 92. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for  
1033 cellular segmentation. *Nat Methods* **18**, 100-106 (2021).
- 1034 93. Ong, H.T. *et al.* Digitalized organoids: integrated pipeline for high-speed 3D analysis of  
1035 organoid structures using multilevel segmentation and cellular topology. *Nature Methods*  
1036 **22**, 1343-1354 (2025).

1037

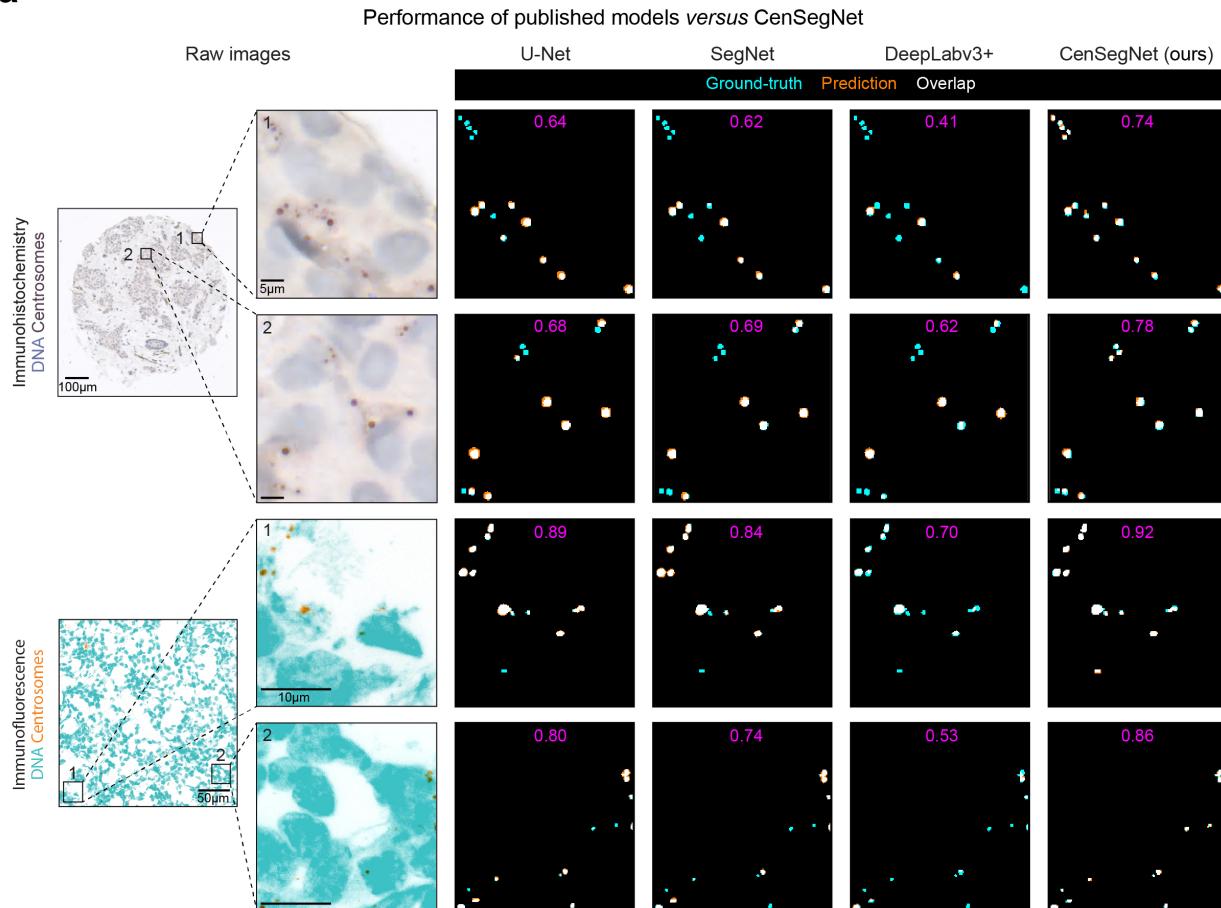
**Fig. 1**



**Fig. 1 Development and benchmarking of CenSegNet for centrosome segmentation.** **a** Left: workflow for generating tissue microarrays (TMAs) comprising 911 breast tissue sample cores from normal breast tissue, breast tumours, and adjacent non-tumour tissue, collected from 127 breast cancer patients in the BeGIN cohort. Figure 1a is schematic for illustrative purposes to demonstrate sampling strategy for the TMA. Cores were taken from FFPE blocks following breast cancer surgery. Top right: immunohistochemical staining of all ROIs with pericentrin (centrosome marker) and haematoxylin (nuclear counterstain). Bottom right: random selection of 108 ROIs used to construct a training dataset containing 2,486 epithelial compartments, 108 stromal compartments and 14,679 annotated centrosomes. **b** Left and middle: representative confocal images of human mammary epithelial cells (MECs) and mouse mammary tumour tissues exhibiting normal or amplified centrosomes. Cells and tissues were stained for pericentrin (orange) and counterstained with DAPI (DNA, teal). Scale bars, 10 µm. Right: training dataset derived from these images, comprising 1,285 annotated centrosomes. **c** Left: 256 × 256 pixel cropped patches from immunohistochemistry and immunofluorescence datasets. Middle left: training of existing segmentation architectures (U-Net, SegNet, DeepLabv3+) and the proposed CenSegNet using these datasets. Middle right: CenSegNet performance across both immunohistochemistry and immunofluorescence test sets. CenSegNet operates in three sequential phases: (1) object detection to generate bounding-box predictions for individual centrosomes; (2) pixel-level segmentation of detected objects; (3) StarDist-based whole-cell segmentation.

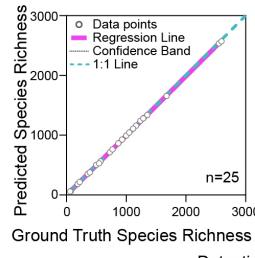
**Fig. 2**

**a**

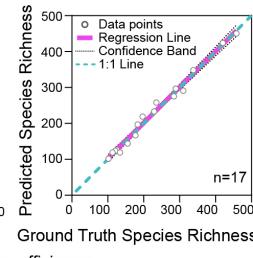


**b**

Immunohistochemistry  
 $R^2=0.9999; p<0.0001$

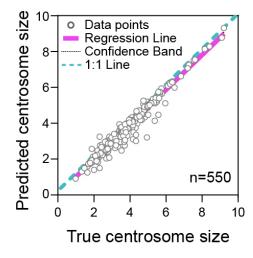


Immunofluorescence  
 $R^2=0.9873; p<0.0001$



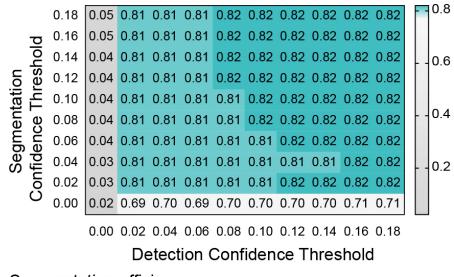
**d**

Immunohistochemistry  
 $R^2=0.9613; p<0.0001$



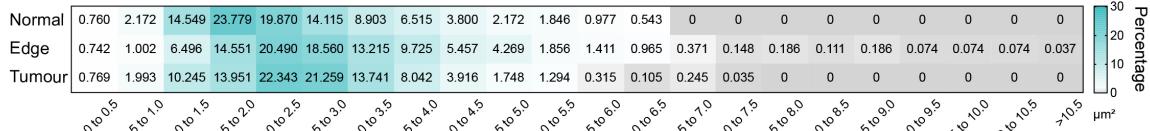
**e**

Overall F1 score (0.82) of Immunohistochemistry and Immunofluorescence datasets



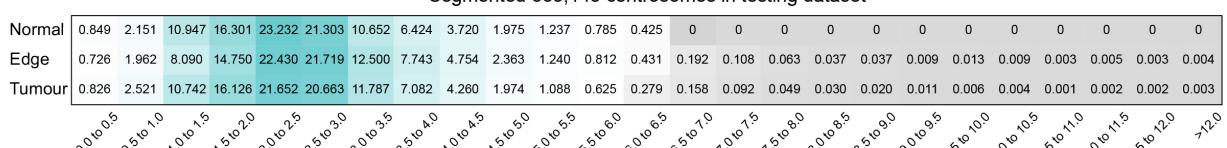
**f**

Annotated 6,475 centrosomes in testing dataset



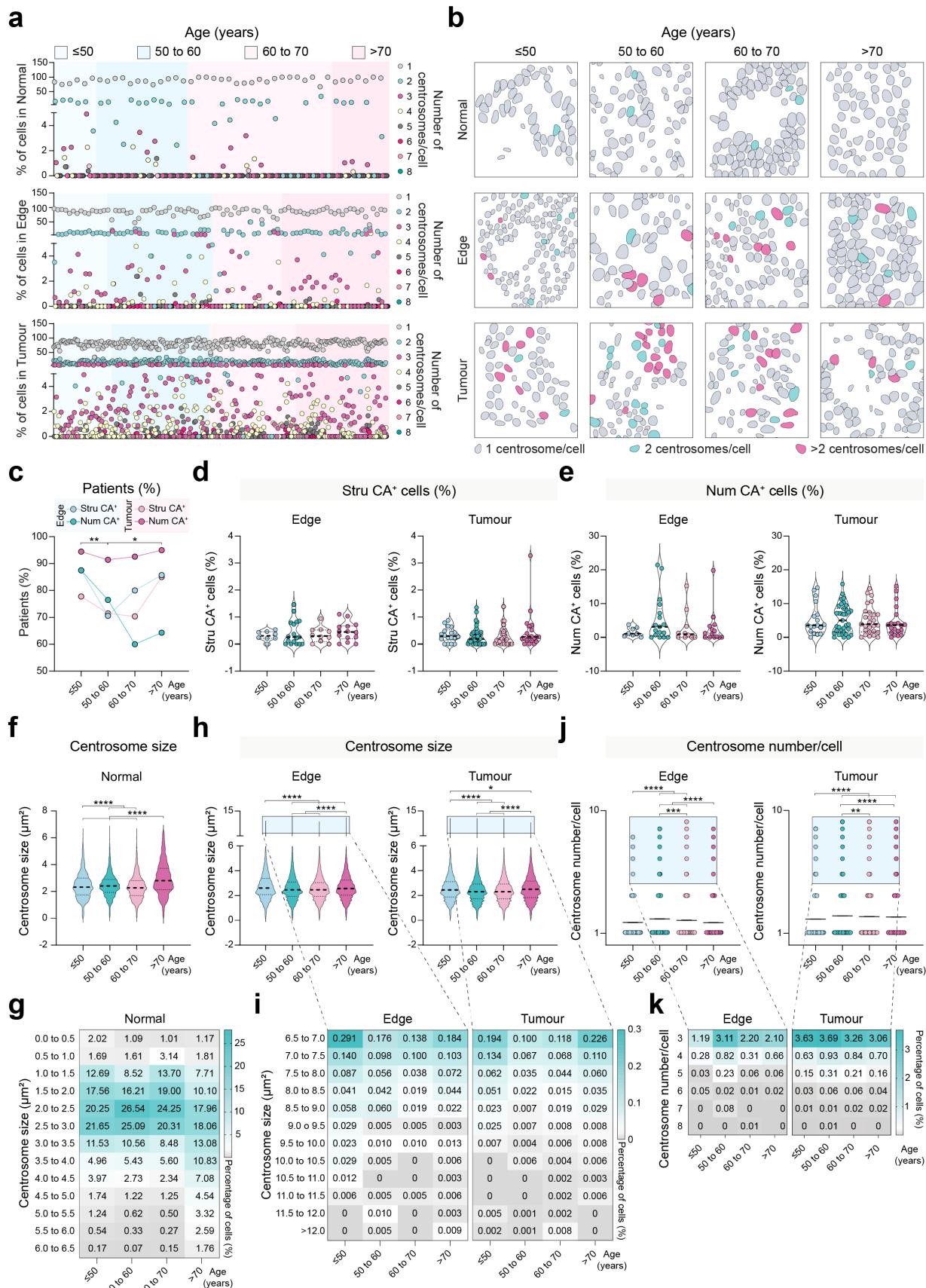
**g**

Segmented 333,148 centrosomes in testing dataset



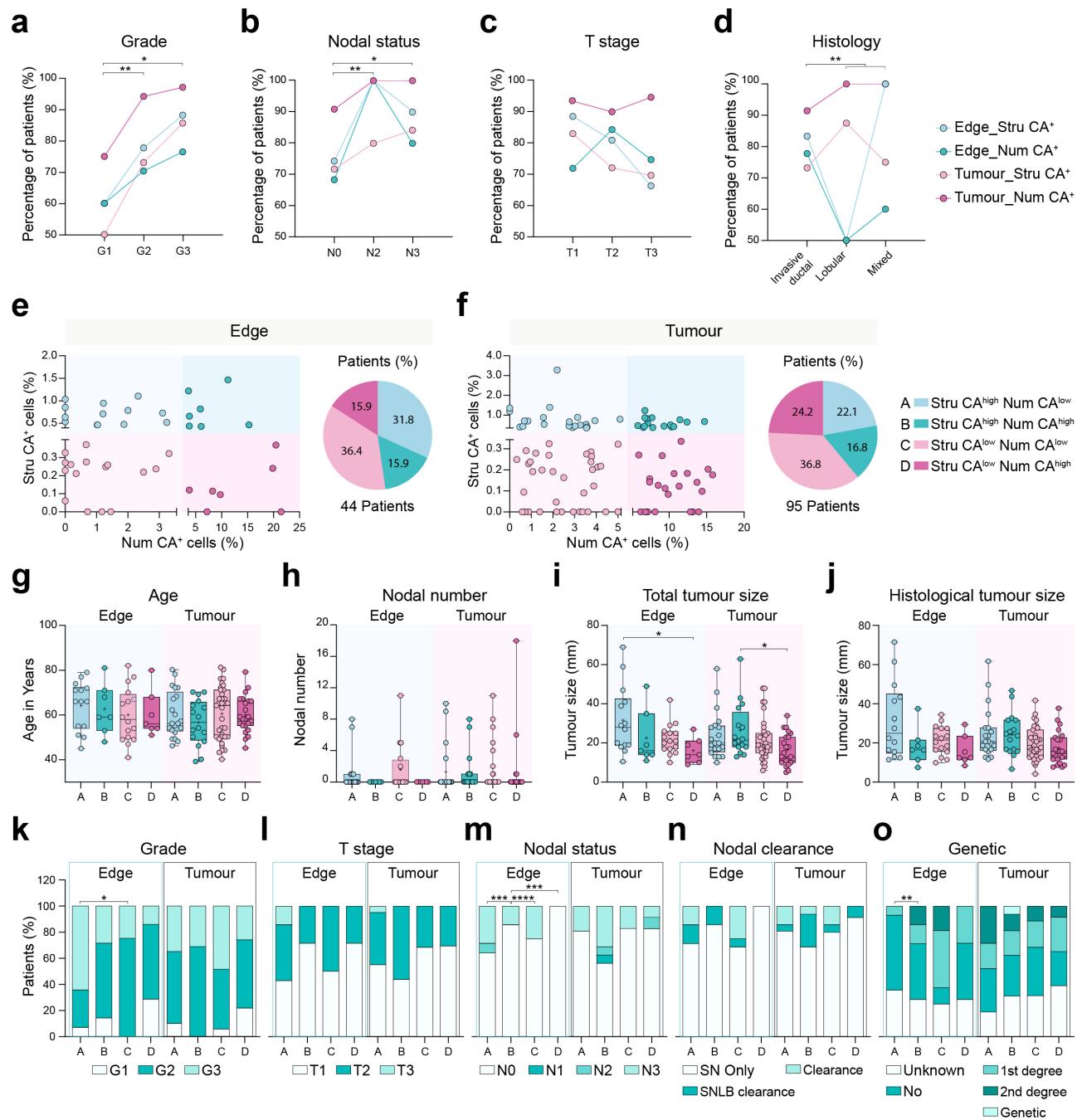
**Fig. 2 Validation of CenSegNet performance against alternative models and human annotations.** **a** Left: representative immunohistochemistry and immunofluorescence images used for centrosome segmentation. Right: colour overlays showing predictions from U-Net, SegNet, DeepLabv3+, and CenSegNet (teal: ground-truth annotations; orange: predicted segmentation; white: overlap between predictions and ground-truth). Magenta text indicates the F1 score for corresponding images of each method (0: complete disagreement; 1: complete concordance). Yellow arrow marks the overprediction. **b** Pearson correlation between centrosome counts predicted by CenSegNet and ground truth across 25 immunohistochemistry images. Grey circles represent individual patients, purple line represents regression line, Black dot line represents confidence band, and teal dot line represents 1:1 line. Two-tailed Pearson correlation test followed by simple linear regression for visualisation, \*\*\* $P < 0.0001$ . **c** Pearson correlation between centrosome counts predicted by CenSegNet and ground truth across 17 immunofluorescence images. Two-tailed Pearson correlation test, followed by simple linear regression for graphical representation, \*\*\* $P < 0.0001$ . **d** Pearson correlation between centrosome size predicted by CenSegNet and measured by annotators ( $n = 550$  centrosomes from  $>200$  ROIs across 20 images). Two-tailed Pearson correlation test, followed by simple linear regression for graphical representation, \*\*\* $P < 0.0001$ . **e** Heatmap of overall F1 scores (maximum 0.82) from combined immunohistochemistry and immunofluorescence test datasets. The x-axis represents the detection confidence threshold, and the y-axis represents the segmentation confidence threshold. Each cell represents the mean F1 score for that threshold pair; darker cyan indicates higher values. Optimal performance ( $F1 = 0.82$ ) was achieved across a broad range of threshold combinations, indicating model robustness to parameter variation. **f, g** Manual annotations of 6,475 centrosomes from normal, edge, and tumour regions—stratified by size ( $0.5\text{--}1.0 \mu\text{m}^2$  to  $>10.5 \mu\text{m}^2$ ) (**f**) were compared with automated segmentations from the full dataset of 333,148 centrosomes (**g**) ( $0\text{--}0.5 \mu\text{m}^2$  to  $>12.0 \mu\text{m}^2$ ). CenSegNet segmentation achieved performance comparable to expert human annotation. Data are presented as individual data points. Source data are provided as Source Data file.

**Fig. 3**



**Fig. 3 Age-related gradients of centrosome aberrations across breast tissue regions. a** Percentage of cells with different centrosome numbers in normal, edge, and tumour regions. **b** Representative cell segmentation masks of normal, edge, and tumour regions across patient age groups (Light grey: cells with one centrosome; purple: cells with two centrosomes; cyan: cells with more than two centrosomes). **c** Percentage of patients with structural CA (Stru CA) and Num CA across age groups in edge ( $\leq 50$  years: n = 8 patients; 50–60 years: n = 17 patients; 60–70 years: n = 10 patients; >70, n = 14 patients) and Tumour ( $\leq 50$  years: n = 18 patients; 50–60 years: n = 35 patients; 60–70 years: n = 27 patients; >70 years: n = 20 patients) regions. Two-way ANOVA with Tukey's test, \*P = 0.0477, \*\*P = 0.0053. Data are presented as mean  $\pm$  s.e.m. **d** Percentage of cells with Stru CA in edge and tumour regions across age in Edge and Tumour regions. One-way ANOVA (Stru CA<sup>+</sup> cells: Edge, P = 0.7637; Tumour, P = 0.4489 with Tukey's test. absence of asterisks indicates no statistically significant correlation. **e** Percentage of cells with Num CA in edge and tumour regions across age. One-way ANOVA (Num CA<sup>+</sup> cells: Edge, P = 0.3427; Tumour, P = 0.9086) with Tukey's test, absence of asterisks indicates no statistical significance. Data are presented as mean  $\pm$  s.e.m. **f, g** Centrosome segmentation results from normal regions across age groups, stratified by size (0.5–1.0  $\mu\text{m}^2$  to 6.0–6.5  $\mu\text{m}^2$ ). One-way ANOVA (\*\*\*\*P < 0.0001) with Tukey's test, \*\*\*\*P < 0.0001. Data are presented as violin plots showing the distribution of values; dashed lines indicate median and interquartile ranges. **h, i** Centrosome segmentation results from edge and tumour regions across age groups, stratified by size (6.5–7.0  $\mu\text{m}^2$  to 12.0  $\mu\text{m}^2$ ). One-way ANOVA (\*\*\*\*P < 0.0001) with Tukey's test, Edge: \*\*\*\*P < 0.0001; Tumour: \*P = 0.0484, \*\*\*\*P < 0.0001. Data are presented as violin plots showing the distribution of values; dashed lines indicate median and interquartile ranges. **j, k** Centrosome number per cell in edge and tumour regions across age groups, stratified by centrosome number (1–8). One-way ANOVA (Edge, \*\*\*\*P < 0.0001; Tumour, \*\*\*\*P < 0.0001) with Tukey's test, Edge: \*\*\*P = 0.0006, \*\*\*\*P < 0.0001; Tumour: \*\*P = 0.0054, \*\*\*\*P < 0.0001, absence of asterisks indicates no statistical significance. Data are presented as individual data points and mean  $\pm$  s.e.m. Source data are provided as Source Data file.

**Fig. 4**

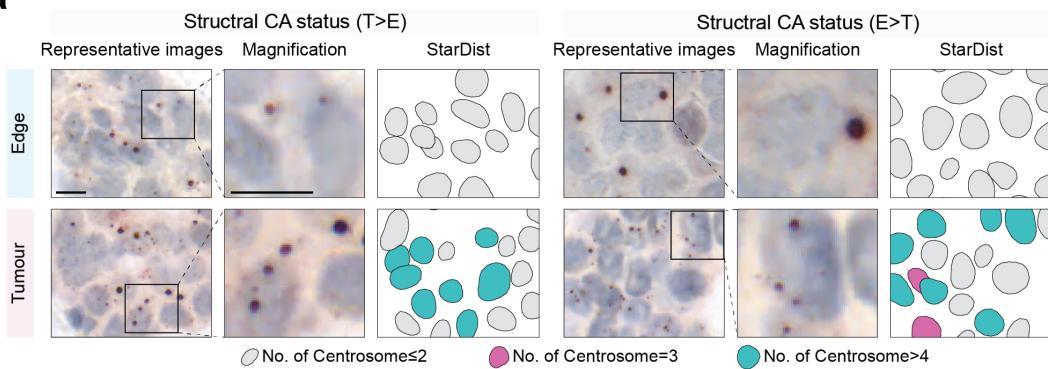


**Fig. 4 Patient stratification by CA burden and correlation with clinicopathological features. a-d** Percentage of patients with Stru CA and Num CA across histological tumour grade, nodal status, histological tumour size, and histological tumour type in Edge [Grade: G1 (n= 5), G2 (n=27), G3 (n=17); Nodal status: N0 (n=38), N2 (n=1), N3 (n=10); T stage: T1 (n=18), T2 (n=16), T3 (n=15); Histological type: Invasive ductal (n=36), Lobular (n=8), Mixed (n=5)], and Tumour [Grade: G1 (n=12), G2 (n=52), G3 (n=35); Nodal status: N0 (n=77), N2 (n=4), N3 (n=18); T stage: T1 (n= 30), T2 (n= 29), T3 (n= 40); Histological type: Invasive ductal (n=82), Lobular (n=8), Mixed (n=8)] regions. Two-way ANOVA (Grade, \*\*P = 0.0020; Nodal status, \*\* P = 0.0030; T stage, P = 0.4024; Histological

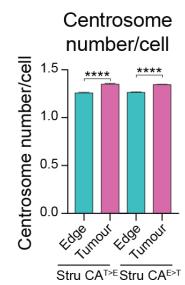
type, \*\* $P = 0.0037$ ) with Tukey's test, Grade: \* $P = 0.0130$ , \*\* $P = 0.0018$ ; Nodal status: \* $P = 0.0113$ , \*\* $P = 0.0030$ ; Histological type: (Invasive ductal *versus* Lobular) \*\* $P = 0.0063$ , (Invasive ductal *versus* Mixed) \*\* $P = 0.0061$ . Data are presented as mean  $\pm$  s.e.m. **e, f** Patients classified by composite CA burden into Stru<sup>high</sup>Num<sup>low</sup> (A), Stru<sup>high</sup>Num<sup>high</sup> (B), Stru<sup>low</sup>Num<sup>low</sup> (C), and Stru<sup>low</sup>Num<sup>high</sup> (D) groups in Edge (A: n = 14 patients; B: n = 7 patients; C: n = 16 patients; D: n = 7 patients) and Tumour (A: n = 20 patients; B: n = 16 patients; C: n = 35 patients; D: n = 23 patients) regions. Histograms and pie charts show percentages of patients in each group in edge and tumour regions. **g–j** Comparative analysis of patient characteristics, including age, number of involved nodes, and total and histological tumour size, across composite CA groups A–D in edge and tumour regions. One-way ANOVA Age (Edge,  $P = 0.7719$ ; Tumour,  $P = 0.6469$ ) Nodal number (Edge,  $P = 0.3164$ ; Tumour,  $P = 0.9944$ ); Total tumour size (Edge, \* $P = 0.0491$ ; Tumour, \* $P = 0.0318$ ); Histological tumour size (Edge, \* $P = 0.0403$ ; Tumour,  $P = 0.0732$ ) with Tukey's test, Total tumour size (Edge, \* $P$  (A vs D) = 0.0484; Tumour, \* $P$  (B vs D) = 0.0296). Absence of asterisks indicates no statistical significance. Data are presented as individual data points and as box and whiskers plots showing the distribution of values, median and quartiles. **k–o** Percentage of patients with different tumour characteristics (histological tumour grade, T stage, nodal status, nodal clearance, genetic subtype) across composite CA groups A–D in edge and tumour regions. Fisher's exact test, left: \* $P = 0.0261$ ; middle: \*\*\* $P = 0.0005$  (middle left), \*\*\* $P = 0.0006$  (middle right) and \*\*\*\* $P < 0.0001$ ; right: \*\* $P = 0.0091$ . **o** 1st degree, family history with 1<sup>st</sup> degree relative; 2<sup>nd</sup> degree, family history with 2<sup>nd</sup> degree relative. Genetic, pathogenic variant in breast cancer-related gene. Source data are provided as Source Data file.

## Fig. 5

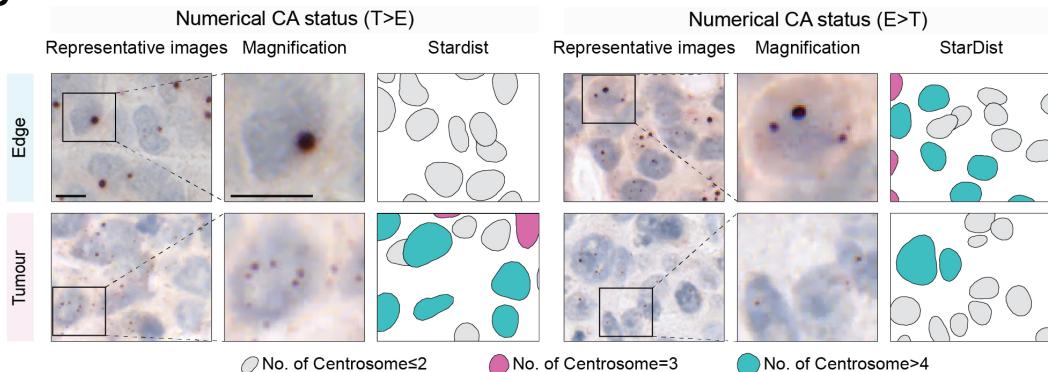
a



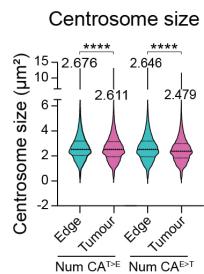
b



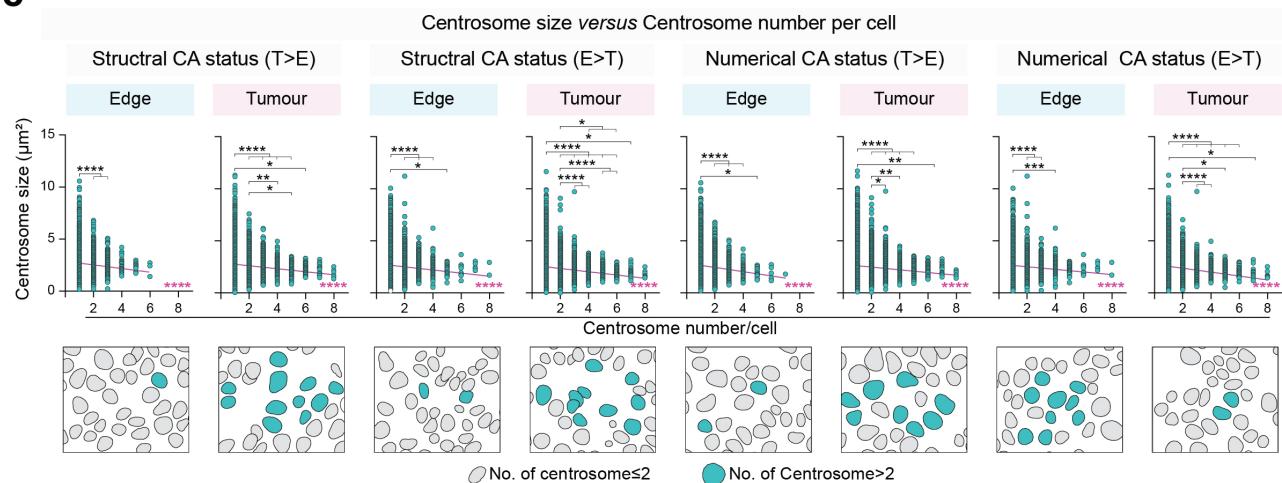
c



d



e



**Fig. 5 Interplay between Stru CA and Num CA drives centrosome defect spatial heterogeneity in breast edge and tumour regions.** a Representative immunohistochemistry images of human breast tumour tissues (stained for pericentrin, counterstained with haematoxylin) and corresponding StarDist cell masks showing differences in Stru CA between edge and tumour regions. Left: Stru CA higher in Tumour than Edge. Right: Stru CA lower in Tumour than Edge (teal: cells with  $\geq 4$  centrosomes; purple: cells with three centrosomes; grey: cells with  $\leq 2$  centrosomes). Scale bars, 10  $\mu\text{m}$ . b Comparative analysis of centrosome number per cell in edge and tumour regions for Stru CA<sup>T>E</sup> and Stru CA<sup>E>T</sup> groups. Comparisons between Edge and Tumour were performed separately for each condition using two-sided unpaired *t*-test, left: \*\*\*\* $P < 0.0001$ ; right: \*\*\*\* $P < 0.0001$ . Data

are presented as mean  $\pm$  s.e.m. **c** Representative immunohistochemistry images of human breast tumour tissues (stained for pericentrin, counterstained with haematoxylin) and corresponding StarDist cell masks showing differences in Stru CA and Num CA between edge and tumour regions. Left: Num CA higher in Tumour than Edge. Right: Num CA lower in tumour than edge (teal: cells with  $\geq 4$  centrosomes; purple: cells with three centrosomes; grey: cells with  $\leq 2$  centrosomes). **d** Comparative analysis of centrosome size in Edge and Tumour regions for Num CA<sup>T>E</sup> and Num CA<sup>E>T</sup> groups. Comparisons between Edge and Tumour were performed separately for each condition using two-sided unpaired *t*-test. Left: \*\*\*\*P < 0.0001; right: \*\*\*\*P < 0.0001. Data are presented as violin plots showing the distribution of values; dashed lines indicate median and interquartile ranges. **e** Top: correlation between centrosome number and mean centrosome size at the single-cell level in Stru CA<sup>T>E</sup>, Stru CA<sup>E>T</sup>, Num CA<sup>T>E</sup>, and Num CA<sup>E>T</sup> groups. One-way ANOVA (\*\*\*\*P < 0.0001 across Stru CA<sup>T>E</sup>, Stru CA<sup>E>T</sup>, Num CA<sup>T>E</sup>, and Num CA<sup>E>T</sup> groups) with Tukey's test: Stru CA<sup>T>E</sup> [Edge: \*\*\*\*P < 0.0001; Tumour: (top) \*P = 0.0318, (bottom) \*P = 0.0416, \*\*P = 0.0056, \*\*\*\*P < 0.0001]; Stru CA<sup>E>T</sup> [Edge: \*P = 0.0139, \*\*\*\*P < 0.0001; Tumour: (top) \*P (2 vs 5) = 0.0362, (top) \*P (2 vs 6) = 0.0144, \*\*\*\*P < 0.0001]; Num CA<sup>T>E</sup> [Edge: \*P = 0.0259, \*\*\*\*P < 0.0001; Tumour: \*P = 0.0421, (top) \*\*P = 0.0081, (bottom) \*\*P = 0.0025, \*\*\*\*P < 0.0001]; Num CA<sup>E>T</sup> [Edge: \*\*\*P = 0.0003, \*\*\*\*P < 0.0001; Tumour: (top) \*P = 0.0340, (bottom) \*P = 0.0115, \*\*\*\*P < 0.0001]. Bottom: Corresponding representative StarDist masks showing cells with Stru CA and Num CA in Edge and Tumour regions. Data are presented as individual data points and mean  $\pm$  s.e.m. Source data are provided as Source Data file.

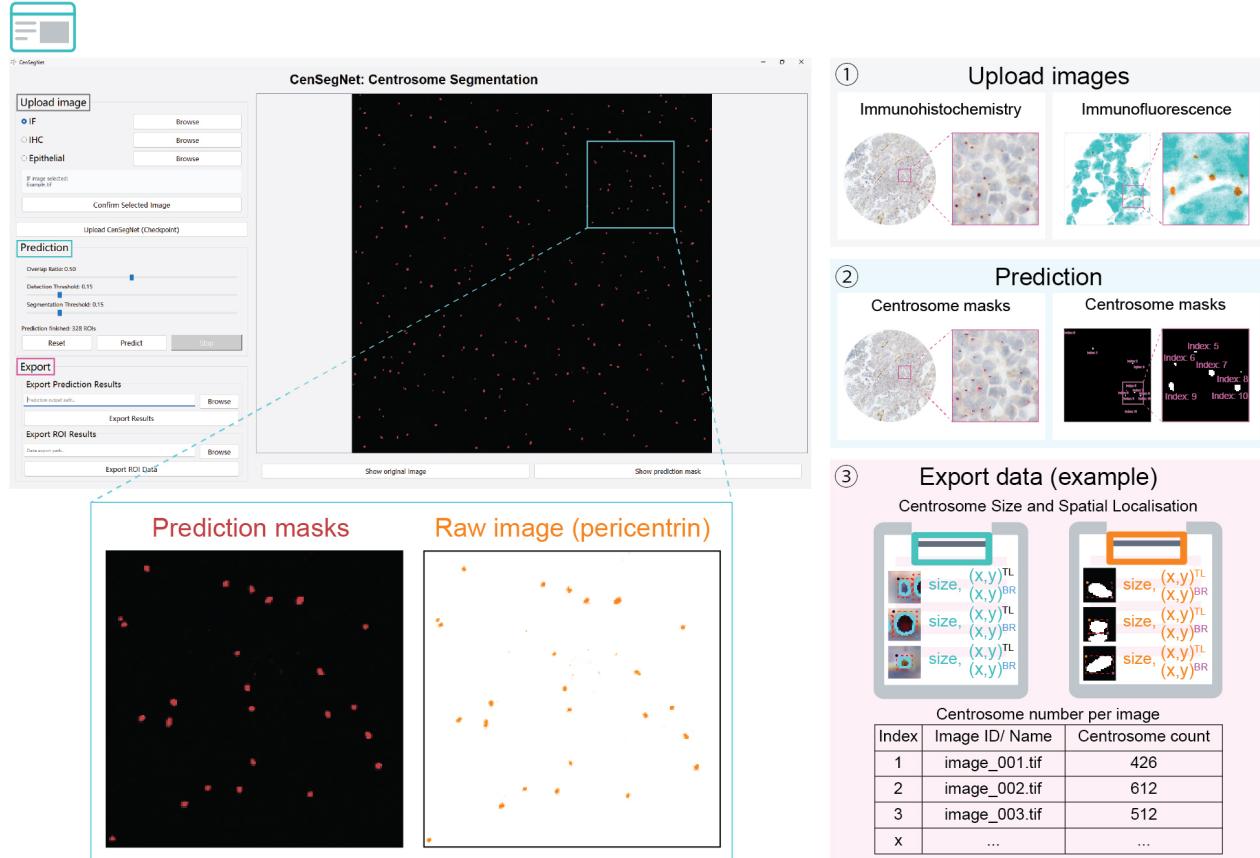
## Fig. 6

a



```
model = initialize_CenSegNet()  
predicted_mask, roi = model.predict(img)
```

b



**Fig. 6 Open accessibility and integration of CenSegNet for broad adoption.** **a** A CenSegNet can be accessed via a Python application programming interface (API). **b** Top: graphical user interface (GUI) of CenSegNet designed for a streamlined workflow, supporting image upload, model-based prediction, and data export. Bottom: interface functions for both immunohistochemistry and immunofluorescence images, enabling users to upload images, apply the relevant prediction models, and export quantitative data including pixel-level centrosome size, localisation, and count.