

**Figure 1: The HE CellProfiler pipeline workflow.** *A.* The original image (left) is masked and normalized using ExpressHist and slideNormalize in the slideToolKit workflow and used as input by CellProfiler; the graph (right) shows the tonal distribution in the digital whole-slide image on a RGB scale. *B.* The input image is converted to a gray scaled image (left); the graph (right) shows the tonal distribution in the gray scaled image. *C.* A Gaussian filter is applied to smoothen the image and reduce image artefacts (artifact size 20 pixels) and noise (left); the graph (right) shows the tonal distribution after smoothening. *D.* The gray scaled image is inverted, *i.e.* non-tissue will become black (left); the graph (right) shows the tonal distribution after inverting. *E.* The tissue area is identified, as demarcated by the green line in the left image; the total tissue area size is calculated in pixels (right image) and tabulated (table). F. The colors, *i.e.* stains, are unmixed using the original image (left): Hematoxylin (HE, middle), and Eosin (right). *G.* The HE-positive objects, *i.e.* nuclei, are identified, white areas in the left image; the HE-positive objects are demarcated by a green line in the middle image, areas that are excluded due to size (minimal size 12 pixels) are demarcated in magenta; the right image shows all the identified HE-positive objects in random colors; the total number of identified objects is calculated and tabulated (table). *H.* Finally the data for each tile are saved in a comma-separated table, including meta-data such as tile positions, image location, object counts (there could be multiple patches of stained areas or tissue). The original image (left) is used to overlay the HE-positive objects (demarcated with blue) and the tissue area (demarcated with green). Sample used: AE1468.UMC.HE.ndpi.