Computational Histopathology Analysis Supplementary Note

Data Aquisition and Pre-processing:
Digital Whole Slide Images (WSIs) of diagnostic H&E slides from IMRT were acquired using the Zeiss AxioScan.Z1 slide scanner. Slides from this cohort were scanned at a resolution of 0.11 µm/pixel, or an equivalent of a 40x magnification. WSIs from the PROMIS, used for initial training of the cell classifier, were scanned using a Hamamatsu Nanozoomer scanner. Slides from this cohort were scanned at a resolution of 0.22 µm/pixel, or an equivalent of a 40x magnification.

For interoperability with images from other scanners, a number of pre-processing steps are applied to the IMRT slide images. This is primarily to improve interoperability between images originating from different scanners. Each slide image is converted to a set of 2000x2000 pixel JPEG tiles, at a 0.44 µm/pixel resolution, spanning the entire image with no overlap. The change in resolution is made to match the resolution of the training dataset used to train the cell detector and classifier from PROMIS images. All subsequent image sizes are with respect to the resolution of these tiles, unless otherwise stated. Tiles are then digitally sharpened with unsharp masking, with radius = 5 and amount = 2.

Stain Normalisation:
Due to differences in tissue preparation for slides originating from many centres, there is significant stain variation across the IMRT cohort. This variation may reduce the accuracy of the deep learning analysis if the stain appearance differs from the images of the training dataset. Therefore, as a final pre-processing step, Reinhard colour transfer is applied to the JPEG tiles in order to adjust the colours of the H&E stains to match that of a reference whole slide image. Because Gleason Classification and Cell Detection/Classification were trained with distinct datasets, this step is applied separately for each model. Thus, for each stain normalisation a distinct reference whole slide image is selected, corresponding to a slide from the original training dataset. La*b* colour channel statistics for Reinhard colour transfer are computed using only pixels deemed as tissue by tissue segmentation. In addition, only pixels deemed as tissue will be adjusted during colour transfer.

Tissue Segmentation:
To improve processing speed, a segmentation algorithm is applied to each tile to separate the tissue section from the background glass. An entropy filter is applied to a grayscale version of the tile, which is thresholded with a fixed value of 3.5. Pixels greater than this value are grouped into connected components. Connected components with less than 500 pixels or a median grayscale intensity greater than 225 (based on a full intensity range of 0-255) are removed. Pixels in the remaining connected components form the final segmentation for the tile.

The rationale behind this method is that, unlike tissue, empty glass is predominantly textureless and thus is expected to have a very low value on an entropy filter. Additional checks for size and median intensity are included to remove small artefacts outside of the tissue. However larger artefacts, such as pen markings, may still be partially segmented as tissue using this method.

Tiles that are found to contain no tissue are not processed by the cell detection or Gleason gland segmentation models. In addition, the tissue segmentation is used to filter out results, such as detected cells or segmented glandular regions, that occur in regions with no segmented tissue.

Gleason Classification
The automated Gleason gland segmentation model takes, as input, two tissue images of size 500x500 pixels: an image at 20x magnification (0.44 microns/pixel) and another at a 10x magnification (0.88 microns/pixel). The two images are centred on the same position on the slide, and thus the 10x image can be considered a zoomed-out view of the region. Due to the use of pre-trained ResNet layers in the model, these images are first resized to 224x224 pixels before being processed by the model.

The output of the network is a segmentation map of the 20x magnification input image, showing the glandular grading of the tissue at each pixel. The model is capable of labelling the tissue with one of 6 labels: No Gland, Normal, Prostatic Intraepithelial Neoplasia (PIN), Gleason 3, Gleason 4, or Gleason 5. However, for this work, Normal and PIN labels were subsequently merged into a single "Benign" class. The raw output of the network is therefore a 224x224x6 array, corresponding to the 6 probability maps for the input 224x224 pixel image patch. To enable direct comparison with the input image patch, probability maps are first resized to the original 500x500 pixel size of the patch.

To scale up segmentation to a whole slide level, analysis is performed in a sliding window fashion. At 20x magnification, the image is divided into 500x500 pixel windows, with 250 pixels of overlap between them. The overlap is included to reduce the presence of border artefacts and improve overall segmentation accuracy. From the output probability maps of these patches, a set of 6 whole slide probability maps are synthesised by joining them together and taking the mean probability for overlapping regions. To avoid potential issues with border artefacts, a border of 10 pixels is discarded from each output probability map before integration into the whole slide probability map.

Upon completion of the whole slide analysis, a segmentation map is generated from the whole slide probability maps. At each pixel the label with the highest probability is selected. Pixels outside of the tissue segmentation are automatically labelled "No Gland". The resultant segmentation map is used in this work to determine the primary and secondary patterns of the slide for the traditional Gleason score, and for the computation of Gleason Morisita Index.

Gleason Model Description

The model is based upon a UNet architecture, with modifications (Figure S11). Convolutional layers on the downsampling half of the network are replaced by ResNet blocks, pre-trained on ImageNet data. Convolutional layers remain in place for the upsampling half of the network.

In addition, modifications have been made to accommodate multi-resolution input. For a given input patch, the N resolution level images are processed in parallel by N sets of downsampling layers. Each set of layers is architecturally identical, but will have its own distinct trainable parameters. The parallel outputs of these downsampling levels are concatenated and then passed through a 1x1 convolutional kernel, before being passed to their corresponding upsampling layer. Unlike the downsampling half of the network, the upsampling half is only composed of a single set of layers.

For our analysis $N=2$ was chosen for the network. This was based upon the observation that an improvement in overall performance could be attained by the addition of a 10x view of the tissue region. This view is perhaps able to provide context of the surrounding area, at the expense of resolution. The inclusion of additional lower resolution levels did not provide any measurable improvement in the accuracy of the segmentation and thus were not used in the final model.

Cell Detection

The cell detector uses an SCCNN architecture. The input to the network is a 31x31 pixel image patch of the tissue. The output of the network is an 11x11 pixel probability map with peaks at locations believed to be cell nuclei. A whole slide probability map is generated by applying the network to patches of the slide in a sliding window fashion. The whole slide probability map is converted to a set of detected cell coordinates by identifying the locations of peaks in the map with a maximum clique algorithm. These detected cell positions will generally correspond to the position of the cell's nucleus. Cells detected in regions outside of the tissue segmentation are filtered out, and the remaining cells are provided as input to the cell classification.

Cell Classification

The cell classifier uses the DenseNet-201 architecture. Convolutional layers in the network were pre-trained on ImageNet data. The input to the network is a 51x51 pixel image patch at a 20x magnification (0.44 microns/pixel), centred on the coordinate of the detection. Due to the use of pre-trained DenseNet layers, the

image patches are resized to 224x224 pixels to fit the required input size. The output to the network is a label for the image patch, identifying the cell type within the image. The model is capable of labelling cells with one of 5 labels: Epithelial cell, Stromal cell, Chronic Inflammatory cell, Acute Inflammatory cell, and Unknown. For this work, Chronic Inflammatory cell and Acute Inflammatory cell labels are merged into single "Immune" cell label. The raw output to the network is a set of 5 probabilities, corresponding to the 5 possible labels. The label with the maximum probability is chosen as the final label.

Gleason Morisita Index

Gleason Morsita Index is a quantification of the degree of heterogeneity of the Gleason patterns present within a section. To achieve this, cells identified as epithelial cells by the cell classifier are sub-classified by the Gleason grade of their associated gland. This was achieved by projecting the positions of the cells onto the automated Gleason segmentation and relabelling every cell according to the assessed Gleason pattern at its position(Figure S17).

Morisita Index is a pairwise metric, and thus Gleason Morisita Index is computed between the epithelial cells belonging to the primary and secondary patterns of the section, as assessed by the automated classifier. Polygons for the Morisita index are generated using Voronoi tessellation as described in [26]. If the primary and secondary patterns are the same (for example: 4+4) then the Gleason pattern is consider to be completely homogeneous for the section and thus the Gleason Morisita is set to 0.

Gleason Scoring from Segmentation Map

For comparison with pathologists assessment, the segmentation map produced by image analysis must be converted into a traditional primary + secondary pattern score. In a broad sense, the computed primary and secondary patterns for a slide correspond to the most frequent and second most frequent pattern in the corresponding segmentation map. However, small amounts of misclassification may interfere with the scoring if this approach is applied naively. In particular, single pattern scores, such as 4+4, would be very unlikely to be selected, as a single pixel of another pattern would be enough to add a second pattern to the map.

Thus, the following criteria is applied to the segmentation map to select the final primary and secondary patterns:

1. Partition segmentation map into **segmented regions**, connected components corresponding to the same Gleason pattern.

2. Discard all segmented regions of size <15,000 pixels (~2,904 sq microns)

3. Sum the remaining area for each grade pattern.

4. Discard grades with combined area <200 000 pixels (~38 720 sq microns) or <10% of total tumour area.

Following this criteria, the remaining grades with the largest and second largest area are selected as the slide's primary and secondary pattern. If only one grade remains, it is selected as both the primary and secondary pattern. If no grade remains, the section is considered benign and is graded as 0+0.

From the Gleason score, a grade group for the slide is determined by applying the ISUP 2014 criteria. A patient-level grade group is computed by taking a weighted mean of the individual slide grade groups and rounding down. The slides are weighted by the size of the tumour area when computing the mean.