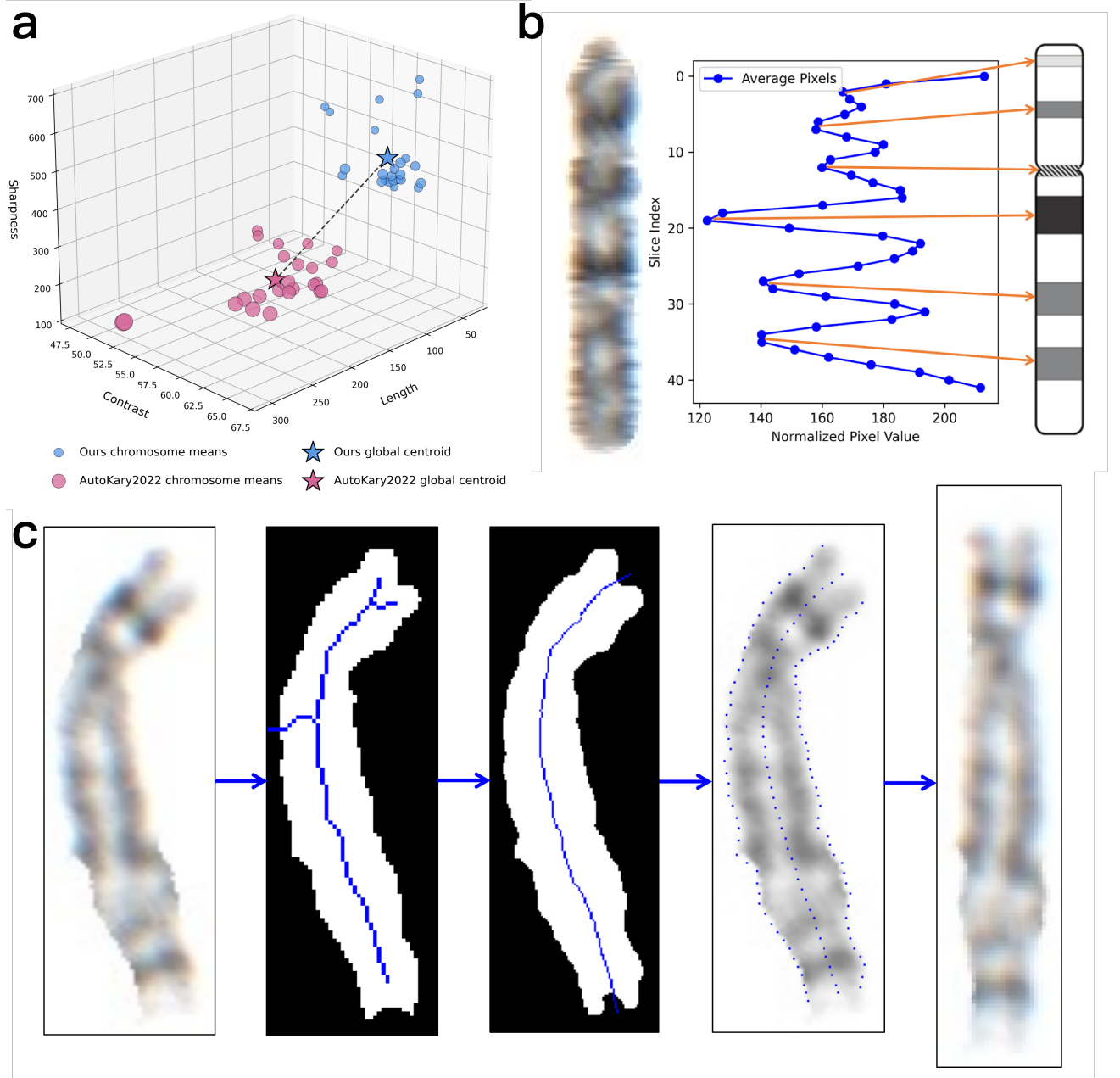
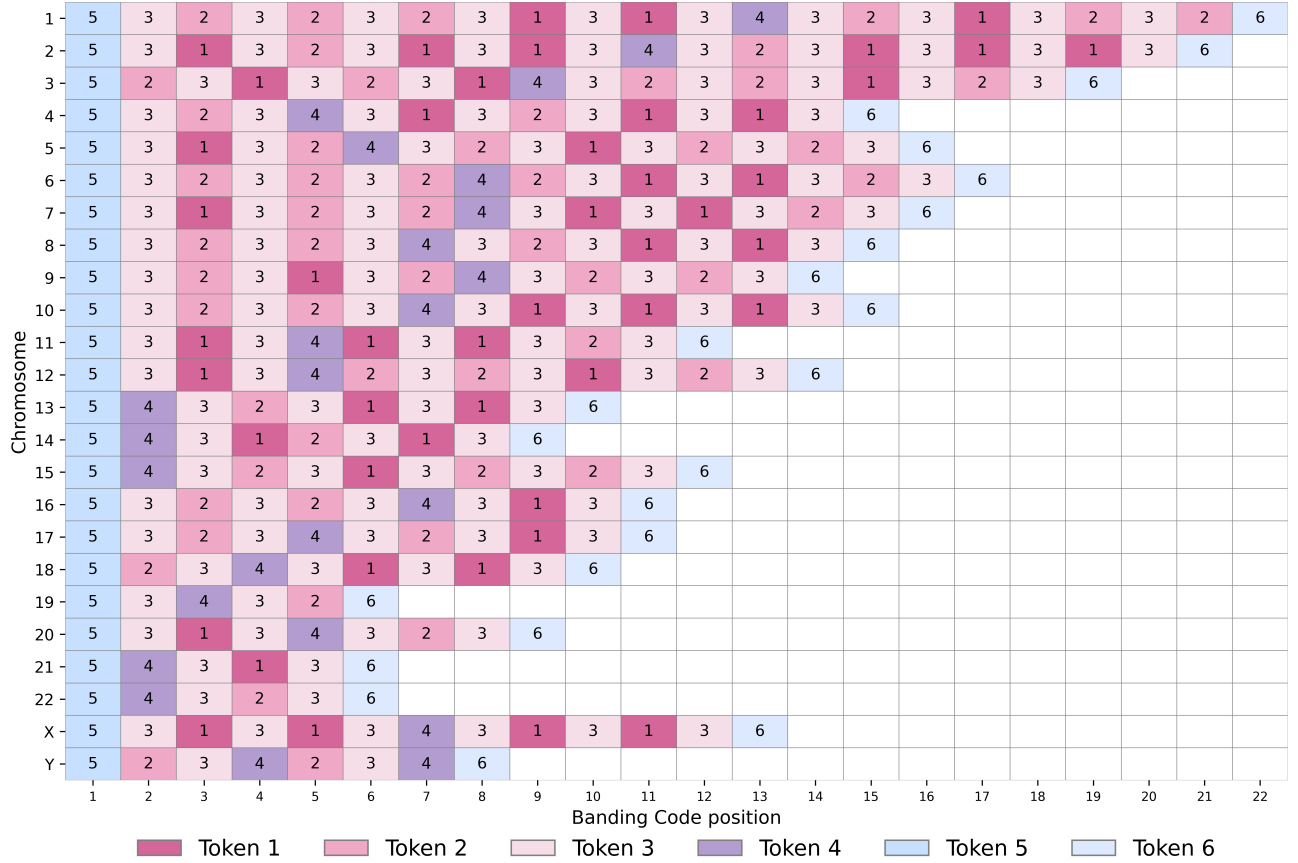


This Supplementary Information provides additional methodological and reference materials supporting the construction of Banding Code representations. Supplementary Fig. 1 summarizes dataset-level image characteristics, grayscale-to-banding correspondence, and the chromosome straightening workflow. Supplementary Fig. 2 presents the reference Banding Code sequences used for the 24 human chromosome classes.



Supplementary Figure 1: **Comparison of dataset images, grayscale mapping, and chromosome straightening workflow.** **a**, Comparison of chromosome features between the two datasets (internal dataset and AutoKary2022). Each point represents the statistical mean of a chromosome class. **b**, Correspondence between chromosome grayscale variations and the standard pattern diagram. The average pixel intensity along the longitudinal axis of each chromosome is mapped to the pattern diagram, showing the spatial correspondence between image intensity and Banding Code tokens. **c**, Schematic of the chromosome straightening workflow. Starting from the original curved chromosome image, the process sequentially applies skeleton extraction, branch pruning, normal intersection computation, and TPS geometric transformation, resulting in a morphology-standardized chromosome image for subsequent Banding Code generation and analysis.



Supplementary Figure 2: **Reference Banding Code sequences for the 24 human chromosome classes.** The heatmap shows the discrete Banding Code sequence for each chromosome class (1–22, X and Y), where the x-axis denotes the Banding Code position and the y-axis denotes the chromosome class. Each colored cell and its corresponding number indicate the token type at that position. Tokens 1–4 represent different band categories and the centromere, whereas Tokens 5 and 6 denote the start and end markers, respectively; blank cells indicate padding positions.