

Group bin method

Brief overview

Group bin distance is a novel measure of image quality based on a comparison between observed chromosomes lengths in an image with chromosome lengths obtained from the human genome sequence. Observed chromosome lengths are approximated by computing the chromosome area in pixels as actual chromosome length is difficult to measure accurately in images. Chromosomes are then assigned to one of three categories based on seven chromosome groups¹⁾:

- Category AB (groups A and B)
- Category C
- Category DG (groups D, E, F, and G)

An ideal metaphase image will have 10 AB chromosomes, 16 C chromosomes and 20 DG chromosomes if the individual is female, or 10 AB chromosomes, 15 C chromosomes and 21 DG chromosomes if the individual is male. The morphological quality of a metaphase image can be measured by comparing its chromosome categorizing result to the female/male standard.

When images in a sample are sorted by chromosome group bin area measurement, a certain number of top ranked images can then be selected for dicentric chromosome analysis. Complex [image selection models](#) can be created by filtering images first with filters and then selecting a certain number of top scoring images.

Group bin distance of all images in a sample can be [plotted](#). Consult the [plots page](#) for instructions on how to do so.

Interpretation

Group bin distance values of a sample (min, median, max, 1st and 3rd quartile, at 250th image, and at 500th image) are shown in the group bin distance [plot](#). They can be used as guides to measure the overall image quality of a sample or rate the image quality of samples against one another. Note the exact integer values discussed in this section (defining good and bad ranges) are meant as guides and are based merely on observations made during internal testing. Images/samples within “good” or “bad” ranges are not guaranteed to be of high or low quality.

The group bin distance of an individual image can be found in the [metaphase image viewer](#). Square brackets enclose the group bin distance of each image within the dropdown box in the lower left of the metaphase image viewer. We have internally classified images with < 8 group bin distance as “good”, group bin distance between 8 and 10 as “borderline”, and group bin distance > 10 as “poor”. Images with group bin distance > 10 are quite likely to be of poor overall quality, while images in the “good” range can often still be filtered out by other image quality metrics within [image selection models](#). Therefore, group bin distance is best used as a method to exclude poor images. This can be accomplished by including only the top x images in a sample ranked by the group bin method, therefore excluding the rest.

Median, 1st quartile, and 3rd quartile are metrics used to judge the overall image quality of a sample.

“At 250th image” and “At 500th image” are the group bin distance of the 250th and 500th image respectively. Ideally, the lowest ranked image included by the [image selection model](#) should be in the “good” range described above. A significant drop-off between these two values may simply be indicative of a small sample. It may also indicate a wide range of image quality within the sample. In either case, it may be advisable to include only the top 200-300 images from the sample when designing an [image selection model](#) using group bin distance. Note the same image selection model should be applied to all relevant samples when creating a [calibration curve](#) and performing [dose estimation](#).

Detailed description

Image morphology is the primary consideration in assessing metaphase image quality. The most common problems in poor quality metaphase cells are severe sister chromatid separation, excessive chromosome overlap, fragments of chromosomes in image segmentation, and multiple cells or incomplete cells in the same image. They result in changes in either the number of objects or areas of objects. For instance, chromatid separation and chromosome fragments cause more objects to be present in an image while areas of some objects are smaller than normal. Chromosome-overlaps reduce the number of objects, but their areas exceed those of discrete chromosomes.

To derive this novel quality measure, we exploited the general property that the different chromosome lengths are approximately proportionate to the known base-pair counts of each complete human chromosome. By comparing the distribution of observed chromosome object lengths with the gold standard derived from the lengths obtained from the human genome sequence, we can assess the overall quality chromosome segmentation of each cell. This assumption sets aside chromosome abnormalities which result from radiation exposure, which will be distributed randomly among cells analyzed, because the cells are synchronized and harvested after a single division. The actual chromosome lengths are difficult to measure accurately in images, so instead, individual chromosomes are approximated according to their corresponding chromosome areas (in pixels). Therefore, the area of an object in a metaphase image is used as a surrogate for which chromosome it represents. Once noisy non-chromosomal objects, nuclei and large overlapped chromosome clusters are removed, areas of the remaining objects are then calculated based on their fractions to the total area of all chromosomes, as overlapping chromosomes and chromatid separation do not significantly affect the total area of objects in each metaphase image. We bin the chromosomes in metaphase cell into three categories corresponding to the known cytogenetic classification system: group A and B (AB), group C (C) and groups D, E, F, and G (DG). A chromosome in category AB contains more than 2.9% (determined by the shortest B group chromosome) of total base-pairs in the complete chromosome set. A chromosome in category C has less than 2.9% (determined by the longest C group chromosome) but more than 2% (determined by the shortest C group chromosome) of total base-pairs in the set. Any chromosome in category DC contains fewer than 2% (determined by the longest D group chromosome) of the total base-pairs. These thresholds 2.9% and 2% are acceptable for the X and Y chromosomes, respectively. We apply these thresholds to object areas to count the number of chromosomes in each category in a metaphase image. An ideal metaphase image will have 10 AB chromosomes, 16 C chromosomes and 20 DG chromosomes if the individual is female, or 10 AB chromosomes, 15 C chromosomes and 21 DG chromosomes if the individual is male. Images with chromosome overlap will tend to have increased AB chromosome counts, while images with sister chromatid separation will likely have elevated DG chromosome counts. The morphological quality of a metaphase image can be measured by comparing its chromosome categorizing result to the female/male standard. In practice, we treat the categorizing result of an image as a 3-element vector and calculate the Euclidean distance to the standard. A larger distance corresponds to a less

satisfactory image, and we find that this measurement is universal for metaphase images from different samples.

When images in a sample are sorted by chromosome group bin area measurement, a certain number of top ranked images can then be selected for dicentric chromosome analysis. Complex [image selection models](#) can be created by filtering images first with filters and then selecting a certain number of top scoring images.

¹⁾

http://www.rerf.jp/dept/genetics/giensa_4_e.html