ENSURE LEGIBILITY AND FOCUS ON THE MESSAGE

Create legible visualizations with a strong message. Make elements large enough to be resolved comfortably. Bin dense data to avoid sacrificing clarity.

Distinguish between exploration and communication.

Use exploratory tools (e.g. genome browsers) to discover patterns and validate hypotheses. Avoid using aesthetics from these applications for communication – they are typically too complex and cluttered with navigational elements to be an effective static figure.

Do not exceed resolution of visual acuity.

Our acuity is ~50 gyres/degree or about 0.3 pt at 10 inches. Ensure the reader can comfortably see detail by limiting resolution to no more than 50% of acuity. Where possible, elements that require visual separation should be at least 1 pt part.

Use no more than ~500 scale intervals.

Ensure data elements are at least 1 pt on a two-column Nature figure (6.22 mm), 4 pixels on a 920 pixel horizontal resolution display, or 2 pixels on a typical LCD projector. These restrictions become challenges for large genotypes.

Show variation with statistics.

Data on large genomes must be downsampled. Depart variations with minima pits and consider hiding it when it is within noise levels. Help the reader notice significant outliers.

Do not draw small elements to scale.

Map size of elements onto clearly legible symbols. Legibility and clarity are more important than precise positioning and scaling. Deterministic sizes and positions to facilitate making meaningful comparisons.

Aggregate data for focused theme.

A strong visual message has no uncertainty in its interpretation. Focus on a single theme by aggregating unnecessary detail.

Show density maps and outliers.

Establishing context is helpful when emergent patterns in the data provide a useful perspective on the message. When data sets are large, it is difficult to maintain detail in the context layer because the density of points can visually overwhelm the area of interest. In this case, consider showing only the outliers in the data set.

Consider whether showing the full data set is useful.

The reader's attention can be focused by increasing the salience of interesting patterns. Different complex data sets, such as networks, are shown more effectively when context is carefully edited or removed.

Use effective visual encodings to organize information.

Match the visual encoding to the hypothesis. Use encodings specific and sensitive to important patterns. Dense annotations should not be independent of the core data in distinct visual layers.

Use the simplest encoding.

Choose concise encodings over elaborate ones.

Help the reader judge accurately.

Accuracy and speed in detecting differences in visual form depend on how much information is presented. We judge relative lengths more accurately than areas, particularly when elements are aligned and adjacent. Our judgment of area is poor because we use length as a proxy, which causes us to systematically underestimate.

Use perceptual palettes.

Selecting perceptually favorable colors is difficult because most software does not support the required color spaces. Brewer palettes [BRE11] exist for the full range of colors to help us make useful choices. Quantitative palettes have no perceived order of color. Sequential palettes are suitable for heat maps because they have a natural order and the perceived difference between adjacent colors is constant. Twin hue diverging palettes are useful for two-sided quantitative encodings, such as immunceninence and copy number.

Crop scale to reveal fine structure in data.

Biological data sets are typically high-resolution changes at half-pixel level can mean changes, square (or circumference of changes) can be orders of magnitude greater than the affected area and connect distant regions by adjacency relationships (gene fusion and other rearrangements). It is difficult to take these properties into account on a fixed linear scale, the kind used by traditional genome browsers. To mitigate this, crop and order axis segments arbitrarily and apply a scale adjustment to a segment or portion thereof.

Use encodings that are robust and comparable.

In addition to being transparent and predictable, visualizations must be robust with respect to the data. Changes in the data set should be reflected by proportional changes in the visualization. Be wary of incorrectly network layouts, which have too many auto-correlation. In general, these are neither sensitive nor specific to patterns of interest.

Use effective design principles to emphasize and communicate patterns.

Well-designed figures illustrate complex concepts and patterns that may be difficult to express concisely in words. Figures that are clear, concise and attractive are effective – they form a strong connection with the reader and communicate with immediacy. These qualities can be achieved with methods of graphic design, which are based on theories of how we perceive, interpret and organize visual information.

Use consistent alignment. Center on theme.

Establish equivalence using consistent alignment. Avoided callouts can be avoided if elements are logically placed.

Respect natural hierarchies.

When the data set embodies a natural hierarchy, use an encoding that highlights it clearly and memorably. The use hierarchy in layout (e.g. tabular form) and encoding can significantly improve a multi-faceted figure.