

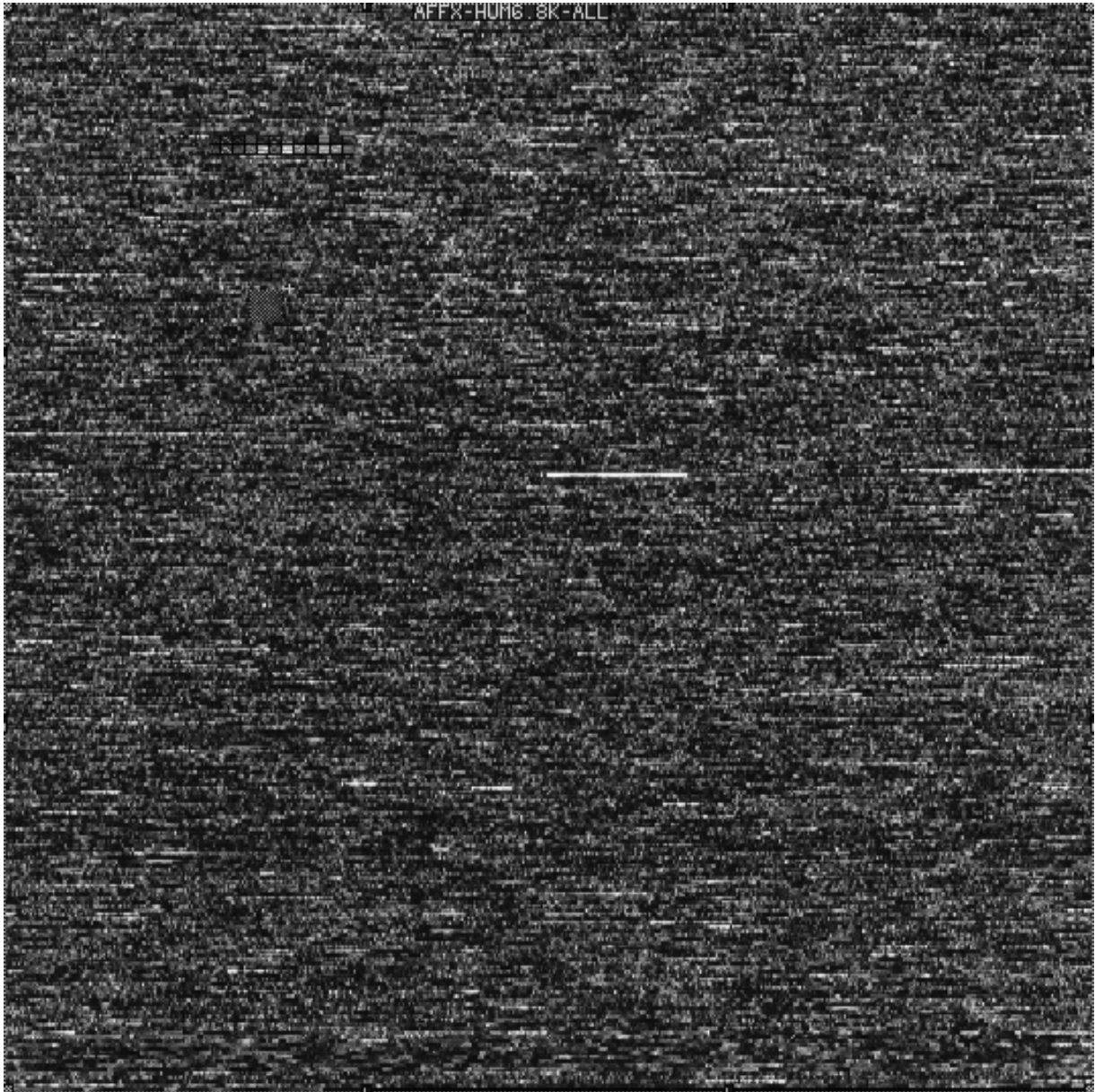
## Data from Affymetrix GeneChip Images

There are four levels of organization in an Affymetrix GeneChip raw data image file. Using the Affymetrix GeneChip Hu6800 as an example, the four levels of organization may be described as follows:

1. Raw pixels. There are  $4733^2 = 22,401,289$  pixels, which take on values in the range 0 to 65535.
2. Probe cells. There are  $536^2 = 287,296$  probe cells. Ideally, each probe cell should have 36 pixels for a total of 10,342,656 pixels. A small percentage of the probe cells have fewer than 36 pixels.
3. Probe pairs. There are approximately 142,580 probe pairs.
4. Probe sets. There are 7129 probe sets laid out on the image.

For the most part, there are 20 probe pairs per probe set and they are laid out in a contiguous row with the perfect match above the mis-match. Some probe sets are not contiguous. Each probe cell has a boundary approximately 1 pixel in width causing probe cells to be separated by 2 pixels.

APPR-HUMS 8K-ALL



## Affymetrix's Probe Set Summaries

After corrections for background illumination and overall image intensity.

Let  $x_1, \dots, x_{20}$  be the perfect matches.

Let  $y_1, \dots, y_{20}$  be the mismatches.

1. Average difference

$$\frac{1}{20} \sum_{j=1}^{20} (x_j - y_j)$$

2. Average log ratio

$$\frac{1}{20} \sum_{j=1}^{20} \log\left(\frac{x_j}{y_j}\right)$$

## Improvements

Although GeneChips measure the quantity of mRNA bound to oligonucleotide probes, Affymetrix's software does not try to estimate these quantities.

Issues/areas to direct improvements:

- Establishing the locations of probe cells.
- Use means when computing probe cell averages.
- Probe pairs are routinely discarded.
- Modelling background pixel intensities and noise.
- Estimating global illumination.
- A definition of repeatability.

