High throughput NGS sequence data in-vitro, in color

Will Trimble

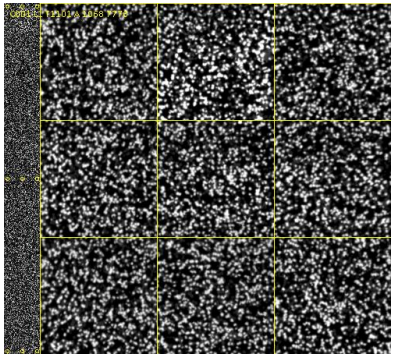
Argonne National Laboratory

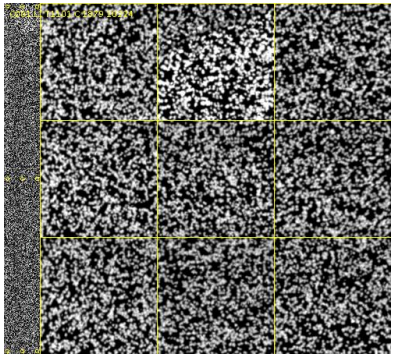
Oct 8, 2013

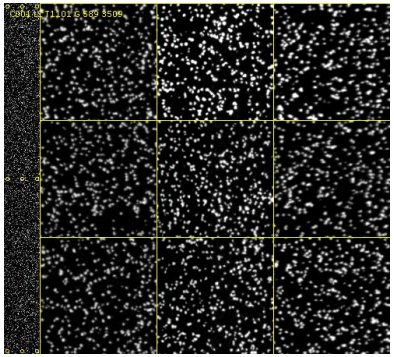
What are thumbnails?

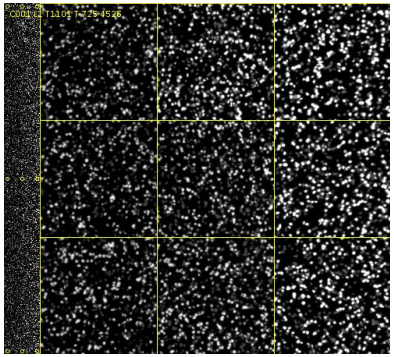
- Illumina instruments record low-resolution images of the flowcell to assist the instrument operators in correcting problems. These are saved in the `Thumbnail_images` directory.
- Scripts to generate false-color composites of entire flowcells are in http://github.com/wltrimbl/thumbnailpolish
- Google search for “illlumina thumbnailpolish”

Color compositing







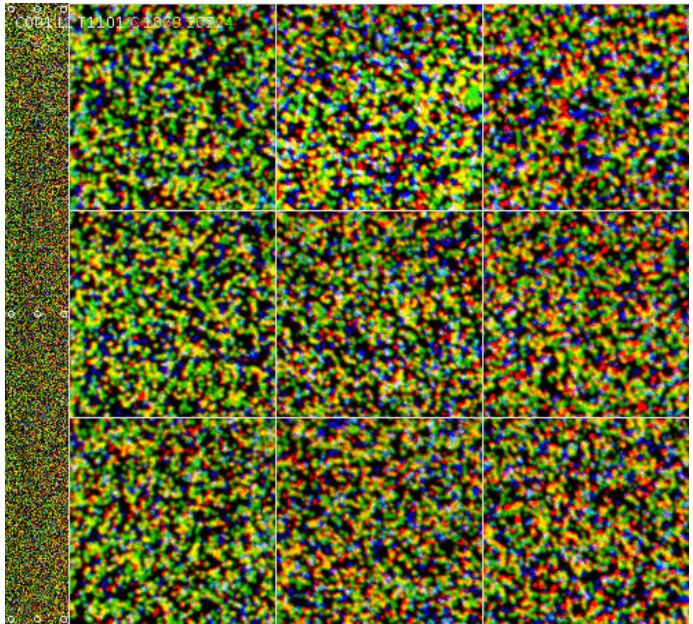


A image

C image

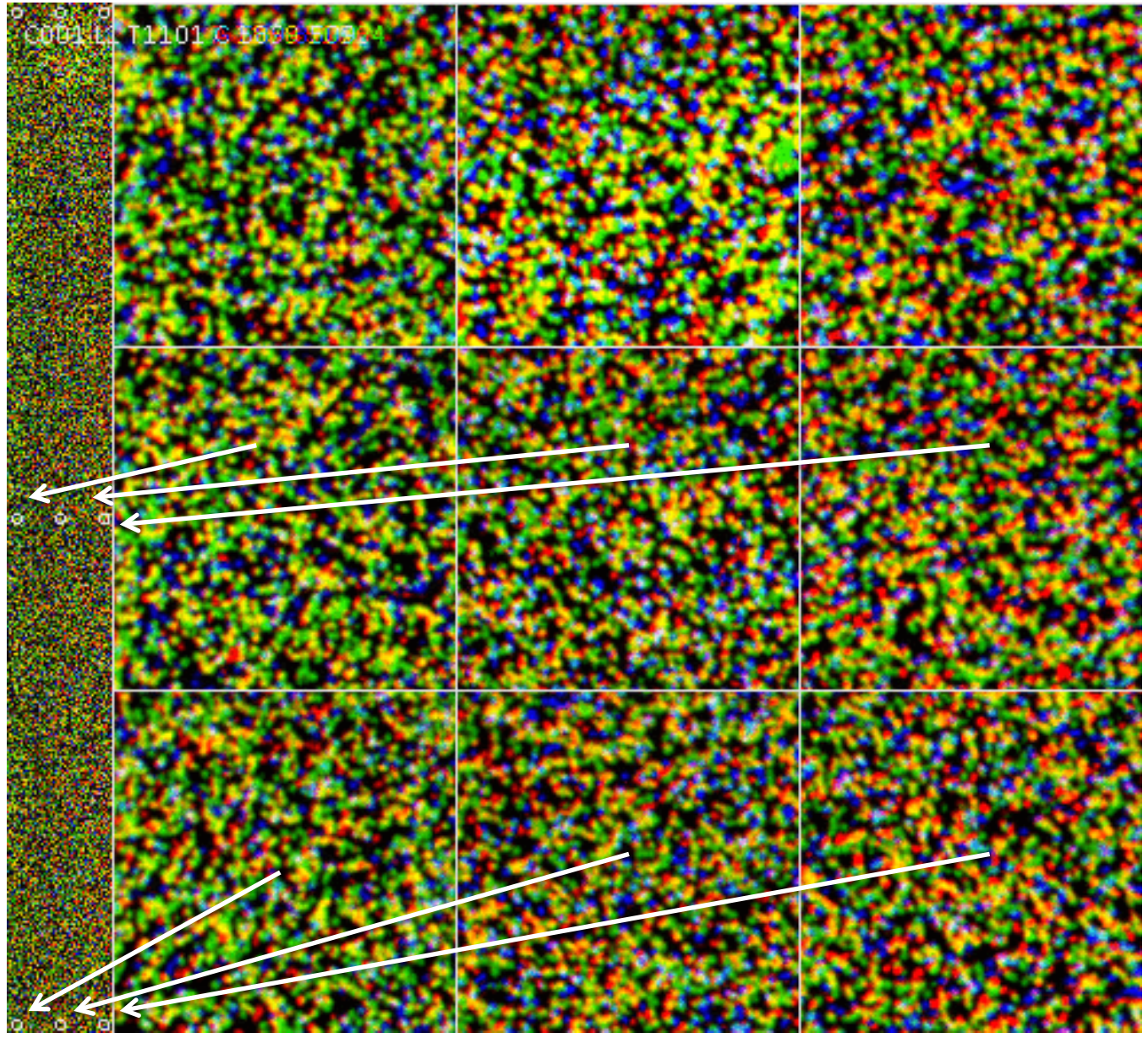
G image

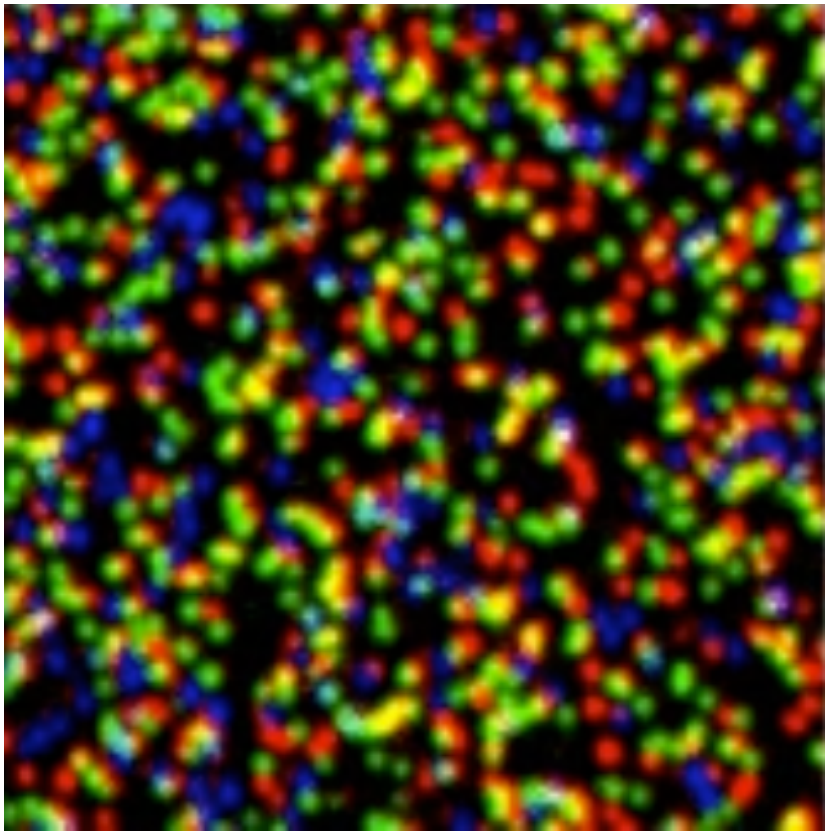
T image



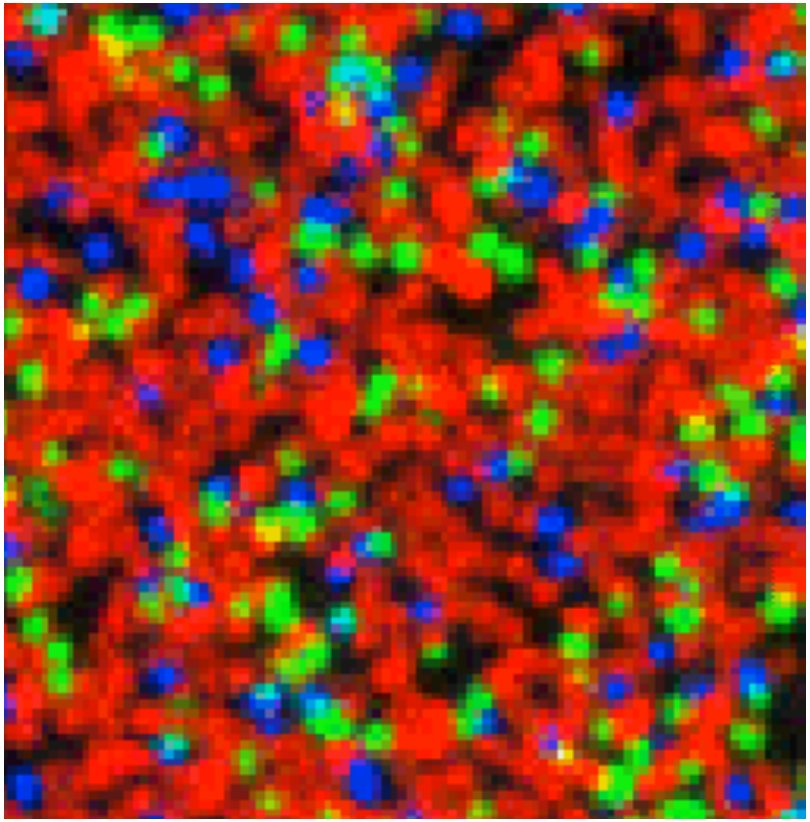
(These don't correspond to the actual colors of the four organic dyes, but what matters is that the camera can adequately distinguish all four.

Color composite


Text
C001.1
T1101
cycle
and
tile
number
Wide shot
(image of
one tile)
on left
4mm
Close-ups:
Nine boxes
with cluster-
resolved
images.
50 µm
750 µm
50 µm
Note: this is not the actual data,
but a low-resolution visualization of it.



Shotgun: about a quarter of the clusters are each of A, C, G, and T for each cycle



Amplicon: because of sequence conservation, most of these spots are T in this cycle.
The clusters that are not red reflect 10% of the sequences that are shotgun.

Flowcell map of one lane

Tiles

1108	1107	1106	1105	1104	1103	1102	1101
1208	1207	1206	1205	1204	1203	1202	1201
1308	1307	1306	1305	1304	1303	1302	1301
2108	2107	2106	2105	2104	2103	2102	2101
2208	2207	2206	2205	2204	2203	2202	2201
2308	2307	2306	2305	2304	2303	2302	2301

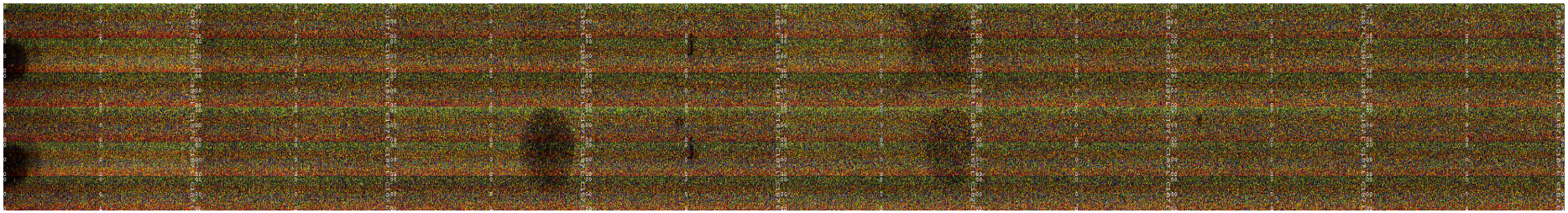
top (?)

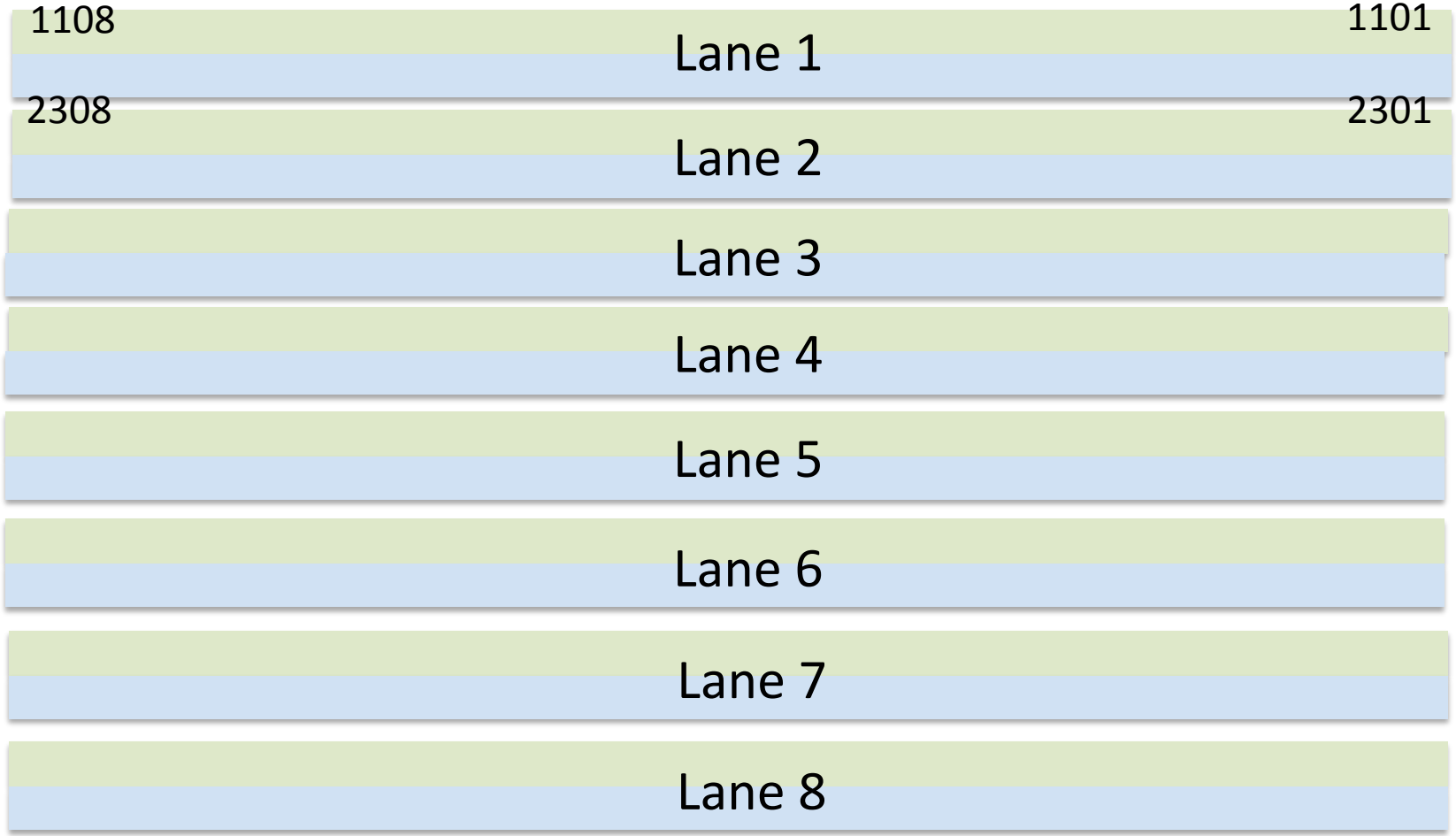
bottom

Tiling the tiles

Each lane has 48 or 96 “tiles” that are the units for image analysis; these have four-digit names.

Stitching the images of the tiles together, we can get an image of the whole lane:

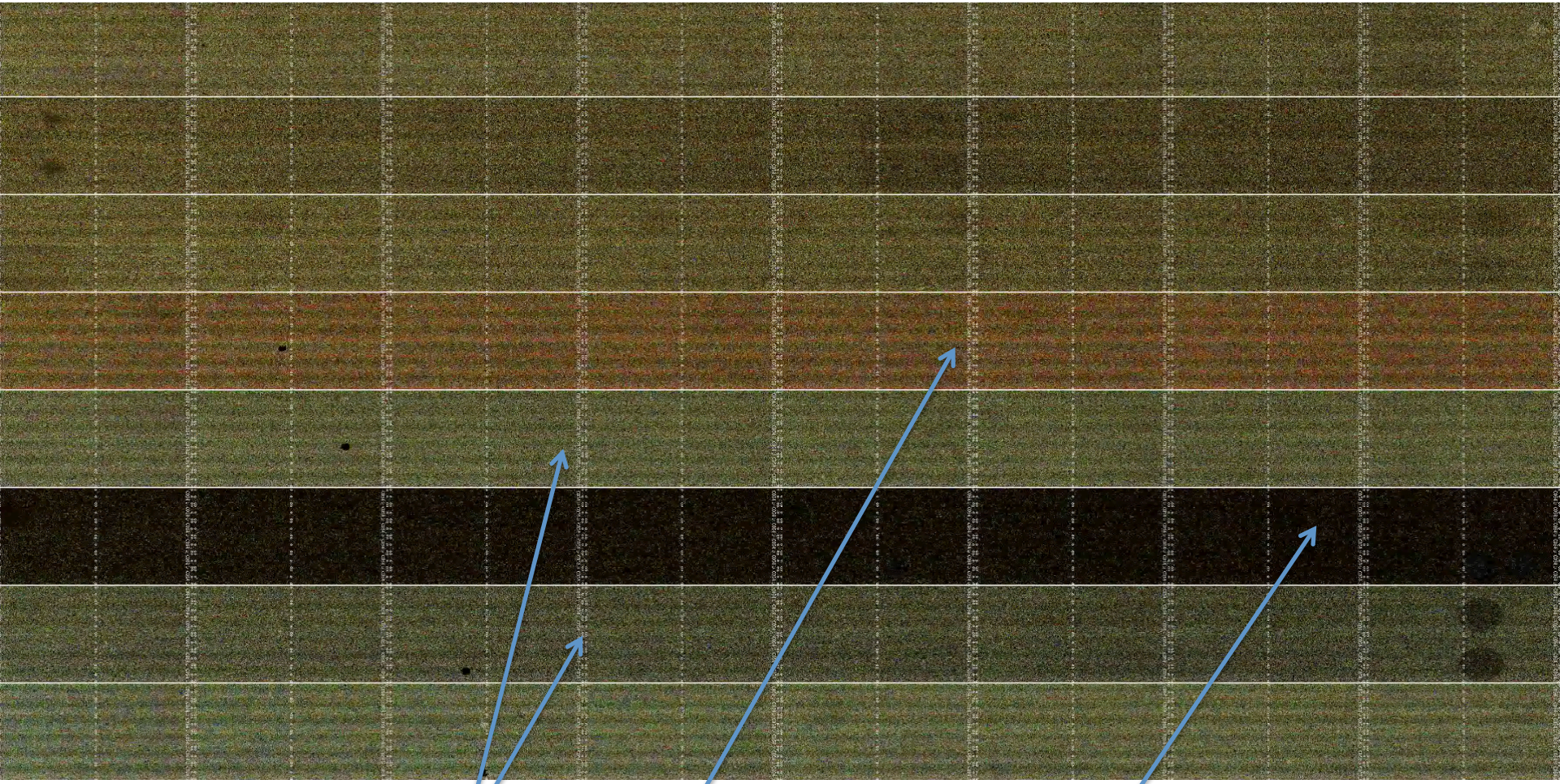




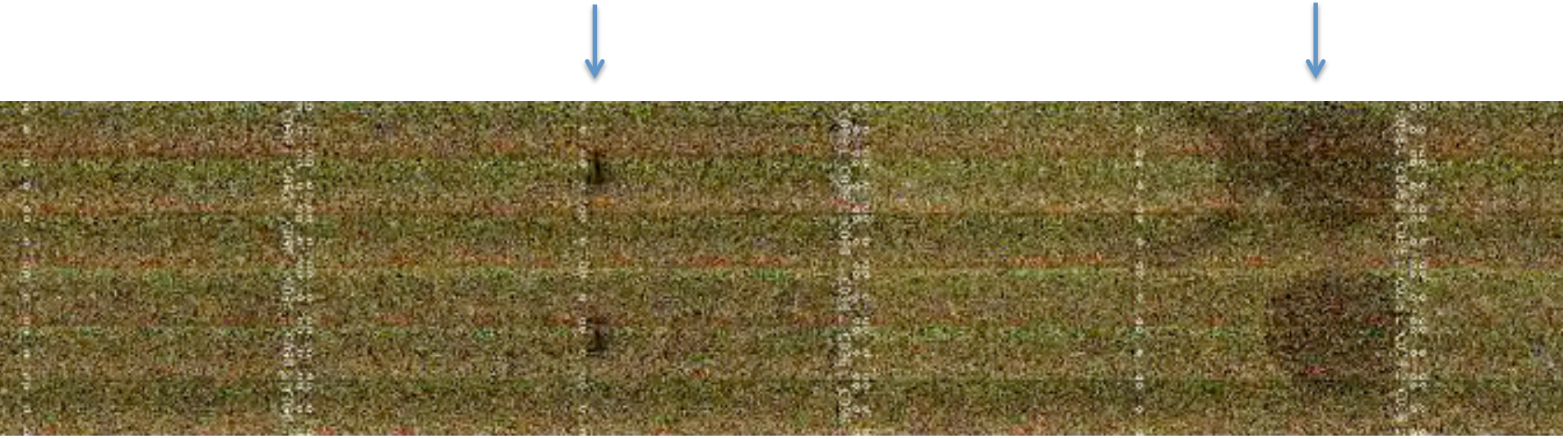
Hiseq flowcells have eight lanes, so these can be composited together to make an image of an entire flowcell.

The Hiseq instruments can recognize about a billion spots in one data-acquisition session.

One flowcell (8 lanes): A C G T



Soft spots

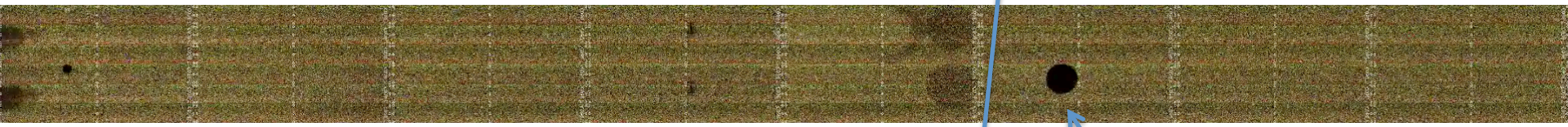


There are spots on most flowcells that are dimmer than average.
Locally low cluster density or low cluster brightness.
Some of these persist for the entire run -> flowcell defects, bubbles in cbot reagents.
Some appear or disappear at the beginning of new reads: R1, index read, R2 -> bubbles in primer / polymerase buffer.

Soft spots

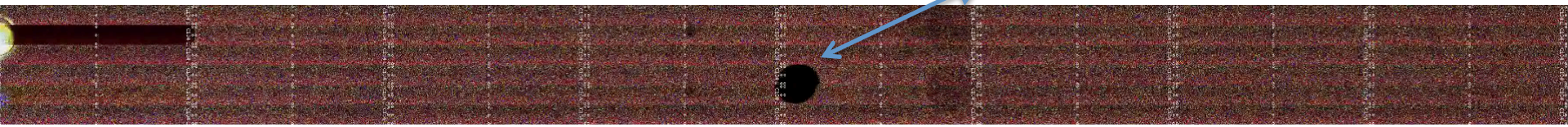
Persistent soft spot

End of R1:



Transient bubbles

Middle of index read:



New soft spot following R2 priming persists for all of R2.

Beginning of R2:

Dark bubbles

Many flowcell images have black circles. These look like flattened-out air bubbles. These bubbles give rise to "N" basecalls.
These bubbles do not, however, keep the reagents from the clusters. These are bubbles that were present when the camera was digitizing the images of the flowcell.
The chemistry happens when the camera is parked. The bubbles that could impede the sequencing-by-synthesis reagents are flushed away before the camera arrives.

Inference: The clusters are there, and have the appropriate dyes, but they can't be photographed because of the bubbles in the imaging buffer. Consequently knowing the positions of these bubbles doesn't really tell you anything useful for sequence-analysis jujitsu that you didn't know by taking account of the Ns.

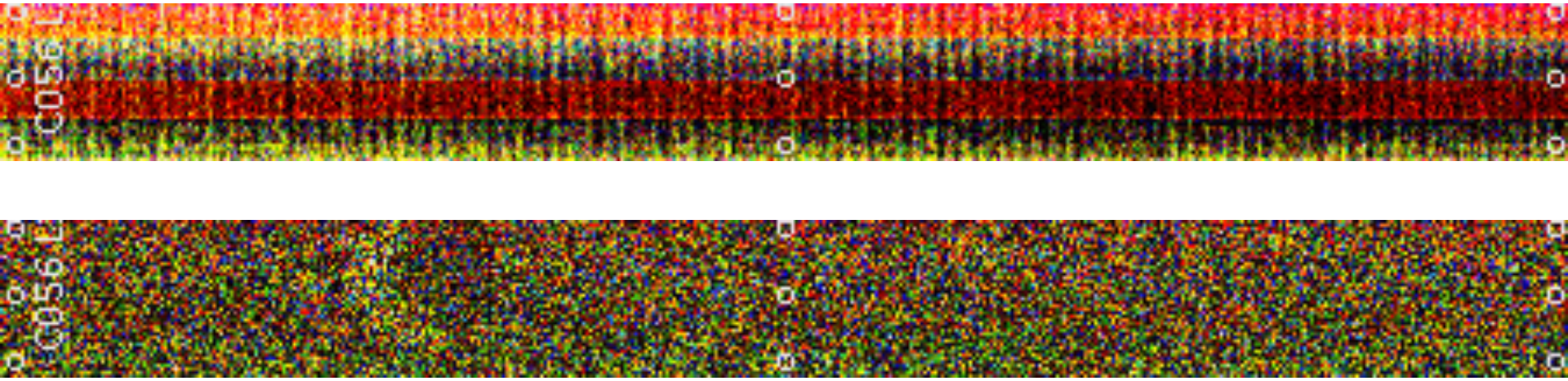
Dark bubbles

How much do these bubbles hurt me?

This bubble is about as large as a tile, so it causes 2% of the clusters on this lane to have an “N” basecall on this cycle.
There are 200 cycles for this run, so this bubble made 10^{-4} of the basecalls on the lane to be N’s.

Some tiles have higher error rates than others. (Cox et al. PMID 20875133)
This is one of the mechanisms for error-prone tiles.

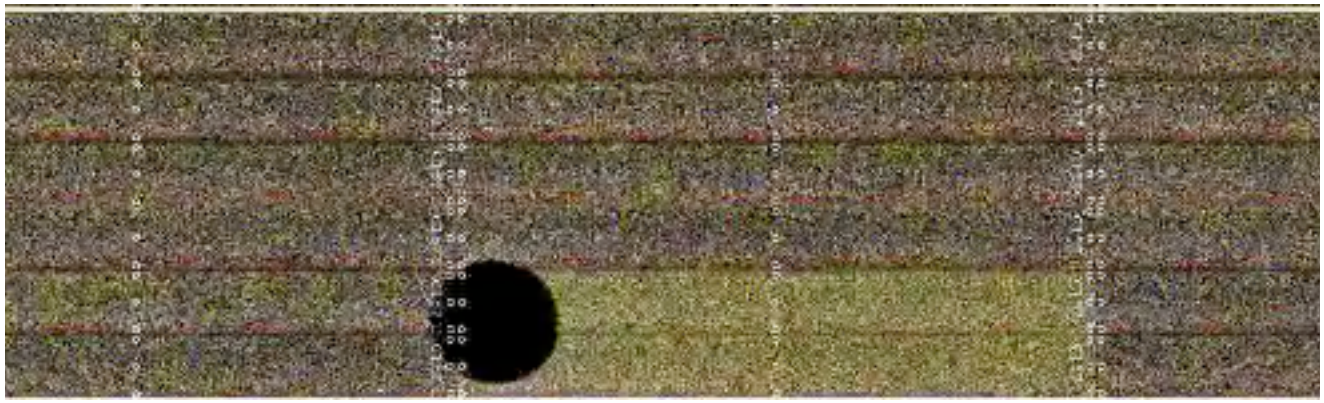
Striped tiles



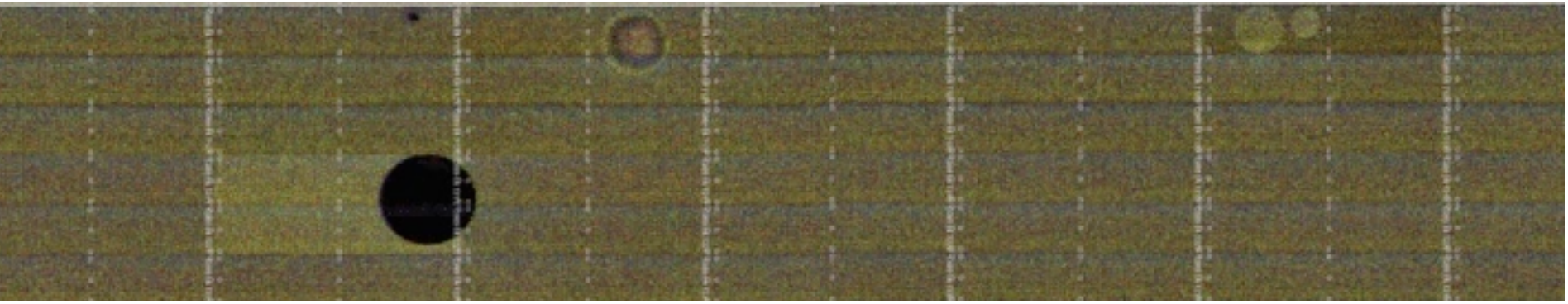
- When cluster yields are very low or when bubbles are present, stripe-shaped artifacts appear in the thumbnails.
- Probably the result of color-correction in the camera going awry. Sometimes results in whole tiles being excluded from basecalling; this eats into yield but is otherwise not fatal.

Top and bottom bubbles

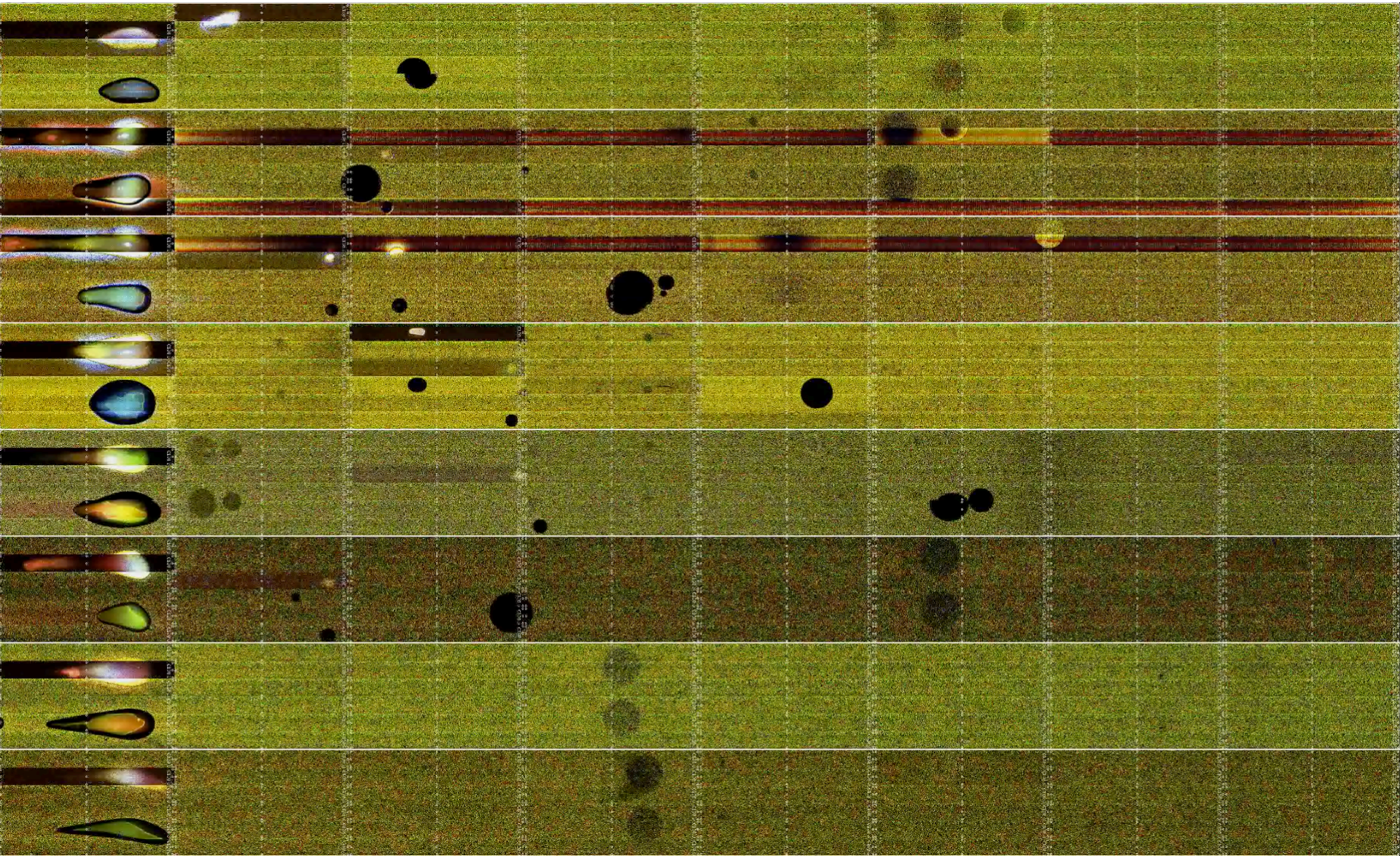
Sometimes, bubbles upset the color balance in the rest of the tile:



Other times (on the lower surface?) bubble-shaped ghosts appear bright:



Sometimes the flow cell has a bad day.



Gallery

- http://www.mcs.anl.gov/~trimble/nodi/hiseq-example.mp4 (contains the close-up of individual spots)
  http://www.mcs.anl.gov/~trimble/nodi/C1KU8ACXX.mp4 (entire cell)
  http://www.mcs.anl.gov/~trimble/nodi/130626_SN1035_0154_AC26L9ACXX-movie-lg.mp4 (more cells)

- http://www.mcs.anl.gov/~trimble/nodi/sarah/Flowcellvideos.html

  The repository with the scripts to produce the movies is here:
  https://github.com/wltrimbl/thumbnailpolish