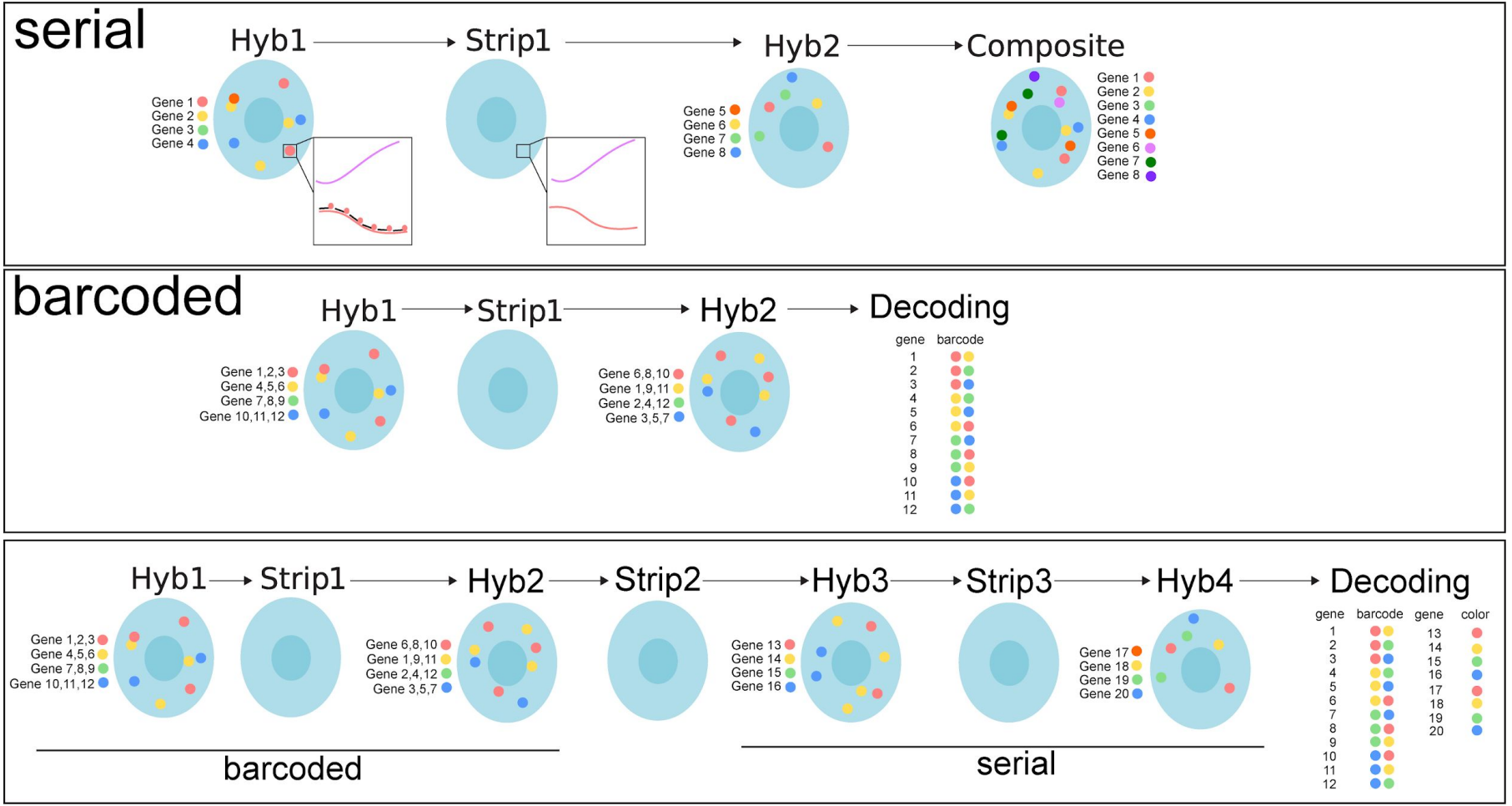
Metadata Acquisition in Cyclic smFISH Experiments

Simone Codeluppi
Linnarsson Lab
Karolinska Institutet

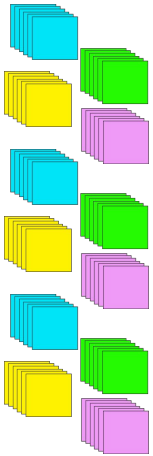
Experimental Approach



Analysis Steps

1

2

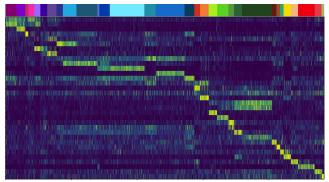


data mining

RNA molecules

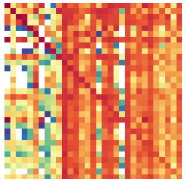
data analysis

clustering



atlassing

spatial relationships



Collected Metadata: Type of Content

Experiment Metadata

Sample Related Metadata

Data Acquisition Metadata

Data Mining Metadata

Data Analysis Metadata

Storage Metadata (Backup/Cold Storage)

Experiment Metadata

- **motivation**: why the dataset has been collected?
- **people responsible** (name, contact info, worked done)
- **institutions** (a different institution can be associated to different experimental steps)
- **information regarding experimental design**
  Example:

```yaml
experiment_name: oPool4
experiment_type: barcoded
rois:
  0:
  - 0
  - 635
image_type:
  Europium: registration_beads
channel_code: {}
total_rounds: 0
probe_pools:
  Europium:
    round_2:
      probe_set: Beads
    round_1:
      probe_set: Beads
extras: {}
```

Sample Related Metadata

Donor related info (examples):

- age
- sex
- genotype
- disease phenotype
- ----------

Tissue related info (examples):

- anatomical informations
- time of collection
- storage
- shipping condition
- travelling time
- ----------

Data Acquisition Metadata

Pre-imaging (example):

- sample cutting
- protocol sample preparation
  txt file vs. *forked version of protocol.io template* (ex. osmFISH)
- ----------

Imaging (example):

- logs acquired from the automated fluidic system
- logs from the microscope during acquisition
- image specific information
- ----------

```
round_2:
  default_tile_format: ZARR
  pixel_microns: 0.122222222222222
  shape:
    height: 2048
    width: 2048
    z: 11
  coords:
    x: 6248.900000000001
    y: 3106.1363636363635
    z: 3680.543636363636
```

```
preprocessing:
  fish:
    flat_field_kernel:
    - 2
    - 100
    - 100
    filtering_small_kernel:
    - 1
    - 8
    - 8
    filtering_laplacian_kernel:
    - 0.2
    - 0.5
    - 0.5
    large_obj_removal_percentile: 99
```

```
2018-04-10 18:08:28,064 - INFO: Connecting and initiating hardware.
2018-04-10 18:08:28,079 - INFO:     Finding device addresses.
2018-04-10 18:08:28,079 - INFO:     MX_Valve_1 Initialized.
2018-04-10 18:08:28,095 - INFO:     MX_Valve_2 Initialized.
2018-04-10 18:08:28,509 - INFO:     ThermoCube_1 Initialized.
2018-04-10 18:08:28,927 - INFO:     ThermoCube_2 Initialized.
2018-04-10 18:08:29,032 - INFO:     YoctoThermistor Initalized.
2018-04-10 18:08:48,248 - INFO:     Connecting with pump.
2018-04-10 18:08:57,666 - INFO:     Established connection with pump.
2018-04-10 18:08:57,666 - INFO:     Initiating pump.
2018-04-10 18:09:14,860 - INFO:     Pump Initialized.
2018-04-10 18:09:14,860 - INFO: Start parameters:
Hyb_time_2_B: 2.0
Start_date_2: 2018-04-10
Imaging_temperature: 20.0
Operator: Lars
Staining_temperature: 37.0
Chamber_EXP_2: Chamber2
Program: Test
Chamber_EXP_1: Chamber1
Readout_temperature: 30.0
Target_cycles_2: 12
EXP_number_2: EXP20180410LB_HumanMTG_Inhibitory
EXP_number_1: EXP20180410LB_HumanMTG_Excitatory
Target_cycles_1: 11
Heatshock_temperature: 75.0
Hyb_time_1_A: 6.0
Hyb_time_1_B: 2.0
Start_date_1: 2018-04-10
Hybmix_volume: 500
Hyb_time_2_A: 6.0

2018-04-10 18:18:12,823 - INFO: Start new experiment, start with chamber 1.
2018-04-10 18:18:12,838 - INFO: STARTING EXP20180410LB_HumanMTG_Excitatory, CYCLE: 1
2018-04-10 18:18:53,004 - INFO: Primed SB65 buffer after replacement.
2018-04-10 18:19:20,047 - INFO: Primed WB20 buffer after replacement.
2018-04-10 18:20:00,954 - INFO: Primed HYB buffer after replacement.
2018-04-10 18:20:42,182 - INFO: Primed IB buffer after replacement.
2018-04-10 18:21:51,591 - INFO:     Dispensed 1000ul of SSC2X to Chamber1.
2018-04-10 18:21:52,818 - INFO:     Chamber1 set to 37.0 degree Celcius.
2018-04-10 18:21:52,818 - INFO: Hybridization Chamber1, cycle 1, indirect A, start dispense
2018-04-10 18:36:40,428 - INFO:     Dispensed C1_01_A to Chamber1, start hybridization.
2018-04-10 18:47:03,271 - INFO:     Washed tubbing of "HYB01" 5 times with SSC2X.
2018-04-11 00:21:52,826 - INFO: Hybridization Chamber1, cycle 1, indirect A, finished
2018-04-11 00:21:52,826 - INFO: Stingency wash Chamber1 start
2018-04-11 00:23:04,135 - INFO:     Dispensed 1000ul of WB20 to Chamber1 with speed 3750ul/min. padding=True, same_buffer_padding=
2018-04-11 00:37:26,564 - INFO:     Dispensed 500ul of WB20 to Chamber1 with speed 3750ul/min. padding=False, same_buffer_padding=
2018-04-11 00:52:32,940 - INFO:     Dispensed 500ul of WB20 to Chamber1 with speed 3750ul/min. padding=False, same_buffer_padding=
2018-04-11 01:07:39,407 - INFO:     Dispensed 500ul of WB20 to Chamber1 with speed 3750ul/min. padding=False, same_buffer_padding=
2018-04-11 01:23:26,841 - INFO:     Dispensed 1000ul of SSC2X to Chamber1.
2018-04-11 01:24:36,284 - INFO:     Dispensed 1000ul of SSC2X to Chamber1.
2018-04-11 01:25:45,613 - INFO:     Dispensed 1000ul of SSC2X to Chamber1.
2018-04-11 01:25:45,613 - INFO: Stringency wash Chamber1. Washed 4 times with 500ul WB20 for 15 minutes
2018-04-11 01:25:46,819 - INFO:     Chamber1 set to 30.0 degree Celcius.
2018-04-11 01:25:46,819 - INFO: Hybridization Chamber1, cycle 1, indirect B, start dispense
2018-04-11 01:40:31,077 - INFO:     Dispensed C1_01_B to Chamber1, start hybridization.
2018-04-11 01:50:51,385 - INFO:     Washed tubbing of "HYB02" 5 times with SSC2X.
2018-04-11 03:25:46,826 - INFO: Hybridization Chamber1, cycle 1, indirect B, finished
2018-04-11 03:25:46,826 - INFO: Stingency wash Chamber1 start
2018-04-11 03:26:58,040 - INFO:     Dispensed 1000ul of WB20 to Chamber1 with speed 3750ul/min. padding=True, same_buffer_padding=
2018-04-11 03:41:20,506 - INFO:     Dispensed 500ul of WB20 to Chamber1 with speed 3750ul/min. padding=False, same_buffer_padding=
2018-04-11 03:56:26,939 - INFO:     Dispensed 500ul of WB20 to Chamber1 with speed 3750ul/min. padding=False, same_buffer_padding=
2018-04-11 04:11:33,363 - INFO:     Dispensed 500ul of WB20 to Chamber1 with speed 3750ul/min. padding=False, same_buffer_padding=
2018-04-11 04:27:20,779 - INFO:     Dispensed 1000ul of SSC2X to Chamber1.
2018-04-11 04:28:30,141 - INFO:     Dispensed 1000ul of SSC2X to Chamber1.
2018-04-11 04:29:39,530 - INFO:     Dispensed 1000ul of SSC2X to Chamber1.
2018-04-11 04:29:39,530 - INFO: Stringency wash Chamber1. Washed 4 times with 500ul WB20 for 15 minutes
2018-04-11 04:29:39,530 - INFO: Imaging buffer injecting into Chamber1
2018-04-11 04:29:40,740 - INFO:     Chamber1 set to 20.0 degree Celcius.
2018-04-11 04:30:36,979 - INFO:     Dispensed 500ul of IB to Chamber1 with speed 3750ul/min. padding=True, same_buffer_padding=Fal
2018-04-11 04:32:13,018 - INFO:     Chamber1 within range of target temperature 20.0C, allowed error 3C. Reached in 96 seconds aft
2018-04-11 04:32:13,018 - INFO: Imaging buffer. Injected 500ul of imaging buffer into Chamber1, temperature set to 20.0C
2018-04-11 09:25:16,942 - INFO: Imaging Experiment: EXP20180410LB_HumanMTG_Inhibitory Cycle: None done.
2018-04-11 09:25:16,942 - INFO: Start Imaging of Experiment: EXP20180410LB_HumanMTG_Excitatory Cycle: 1
```

Data Mining Metadata

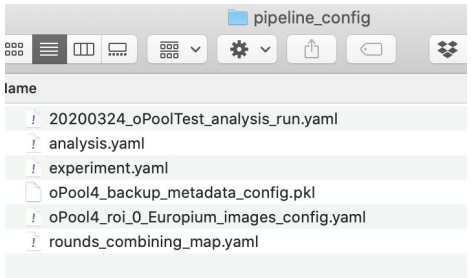


Image Processing Steps (example):

```yaml
analysis_name: oPoolTest
pipeline_steps:
  parse_nd2_files: reparsing
  backup_microscope_raw: false
  filtering: false
  raw_counting: false
  stitching: false
  consolidate_experiments: false
analysis_parameters:
  fov_analysis_parameters:
    rounds_registration:
      registrations:
      - to_reference
      reference_round: 1
      min_error: 3
      max_iterations: 20
      chunk_size: 200
      min_dots_chunk: 20
      n_neighbors: 2
      matching_radius: 2
      min_acceptable_distance: 3
      residual_threshold: 10
      min_samples: 3
      percent_padding: 0
      max_trials: 20
```

Specific FOV Processing Parameters (example):

```yaml
preprocessing:
  fish:
    flat_field_kernel:
    - 2
    - 100
    - 100
    filtering_small_kernel:
    - 1
    - 8
    - 8
    filtering_laplacian_kernel:
    - 0.2
    - 0.5
    - 0.5
    large_obj_removal_percentile: 99
    large_obj_removal_min_obj_size: 50
    large_obj_removal_selem: 3
counting:
  min_distance: 2
  min_obj_size: 2
  max_obj_size: 200
  num_peaks_per_label: 1
beads_registration: {}
```

Data Analysis Metadata

Same approach for the image processing metadata: code + parameters
Example:

- cell type clustering methods and parameters
- spatial clustering methods and parameters
- ----------

Storage Metadata (Backup/Cold Storage)

Maybe not relevant for this discussion but important for us because we have a ‘working’ copy of the data and a cold storage copy for backup

Discussion Points

- Sort the metadata in searchable vs non-searchable and where to store them

- standard for file format for the different type of metadata (ex. yaml/json/csv we are currently using .yaml)

- how to integrate analysis code or processing environment (docker images ex. https://hub.docker.com/r/simonecodeluppi/pysmfish_img, github repository hosted where?)