Project Acronym: MuG

Project title: Multi-Scale Complex Genomics (MuG)

Call: H2020-EINFRA-2015-1

Topic: EINFRA-9-2015

Project Number: 676556

Project Coordinator: Institute for Research in Biomedicine (IRB Barcelona)

Project start date: 1/11/2015

Duration: 36 months

Deliverable 6.5: Software tools on analysis of single cell experiments

Lead beneficiary: The University of Nottingham (UNOT)

Dissemination level: PUBLIC

Due date: 31/10/2018

Actual submission date: 12/12/2018

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556.

Document history

Version	Contributor(s)	Partner	Date	Comments
0.1	Satish Sati	CNRS-IGH	29/11/2018	First draft
0.2	Anna Montras	IRB Barcelona	29/11/2018	Additions in section 3
0.3	Isabelle Brun	IRB Barcelona	29/11/2018	Revision
0.4	Charlie Laughton	UNOT	05/12/2018	Revision
0.5	Anna Montras, Andrew Yates	IRB Barcelona, EMBL-EBI	10/12/2018	Update on availability of image data repositories
1.0			11/12/2018	Approved by Supervisory Board

Table of contents

Executive summary

Chromosome conformation capture assay results have the potential to be usefully complemented with results from DNA FISH assays. The former assay enables the user to aquire global interaction profiles (Hi-C) over a population, while the latter allows this to be done in a pairwise manner in a single cell specific manner. In this document we describe the challenges that hinder the integration of FISH analysis tools in the VRE. The main challenges are a) a lack of community-wide data standards in this experimental area; and b) a lack of community-accepted databases in which to store FISH assay raw data.

1 INTRODUCTION

3D genome organization plays a vital role in understanding the spatiotemporal regulation of gene expression (1, 2). There are two main ways of studying the 3D genome organization, one is microscopy based approach and the other is chromatin conformation capture based assay. Although microscopy based approaches have long been used to study nuclear architecture, recent progress in the chromatin conformation capture assays, such as Hi-C assay, have greatly enhanced our understanding of 3D genome architecture (3, 4). Independently both microscopy and conformation capture based assay have validated the occurrence of chromosome territories within nucleus (2). Further, capture based methods have dissected these territories into genomic compartments, which can be defined as gene rich or early replicating active (A) and gene poor or late replicating inactive (B) compartment (1,5). At a finer scale, chromatin was found to be organized into topologically associated domains (TADs), which are the basic unit of chromatin folding (6, 7). Advances in imaging methods such as super resolution microscopy have complemented these findings and have validated the presence of TADs and chromatin compartments in single cells (8).

Recently high resolution Hi-C maps were also able to display a variety of interactions such as promoter-enhancer, transcription start site (TSS-TSS), CTCF-CTCF, etc (9-11). However, a key limitation of conformation capture based assays is that, they display an average contact profile coming from millions of cells. Performing these assays on single cells is technically challenging. Thus, DNA FISH assay are generally employed to corroborate the findings of conformation capture assays.

Thus, the goal of this deliverable is to report the DNA FISH tools employed by the community and the feasibility to integrate them in the VRE

2 Overview

A simple quantitative DNA FISH experiment starts from sample or FISH probe preparation and culminates in image analysis (Fig 1). A few major points of consideration for DNA FISH are described in here.

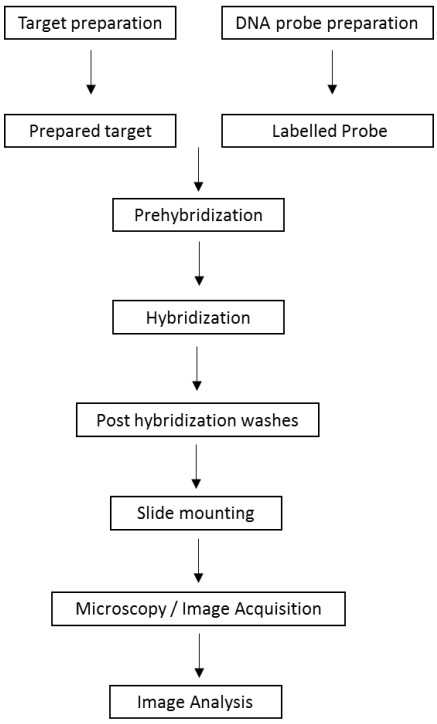

Figure 1. DNA FISH assay overvew

2.1 Experimental design

DNA FISH is generally employed to qualitatively and quantitatively study locus specific interactions, such as between enhancer and promoter. The goal of these assays is to measure the distance between the two loci and their localization within the nucleus (distance to periphery). Using special probes such as oligopaints (described below) and super resolution microscopy, one can also study the structure of a chromatin compartment or TADs.

At CNRS-IGH WP7.1 was involved in performing DNA FISH experiments. They employed the oligopaint based probes followed by confocal microscopy to measure the

- distance between the loci of interest
- distance of these locus from periphery (nuclear lamina)
- diameter of the signal, as a measure of the compaction (explained in deliverable 7.1)

2.2 Sample and Probe preparation

The specimen or the sample for the FISH can be a tissue section or a layer of cultured cells. As highlighted in fig1, the first step in the assay is the fixation of the sample. The samples are generally fixed with formaldehyde. Once fixed the samples can be stored under proper conditions for a limited period of time. DNA FISH results might vary if the storage conditions are not proper, leading to error in conclusions.

DNA probes for the FISH assay can be made in following ways

- nick translation and labelling of the PCR products
- Oligopaint based approach: Complex ssDNA libraries consisting of 43 bp unique stretch of genomic sequence (black lines) flanked by non-genomic regions (colored lines) containing primer sequences are amplified, labelled and processed to produce ssDNA probes (Figure 2).

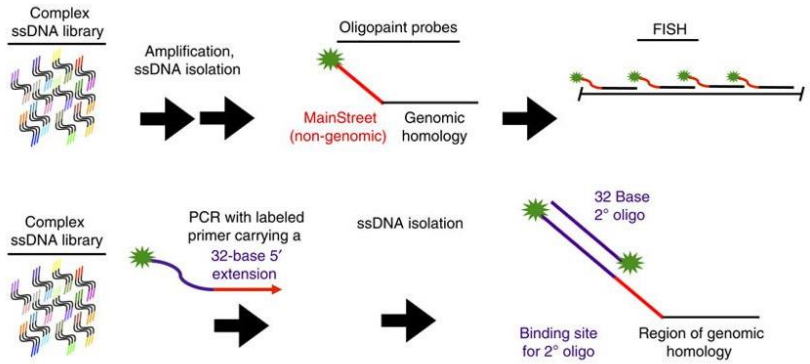

Figure 2. Oligopaint probe preparation. Modified from Beliveau et al., Nature Communications volume 6, Article number: 7147 (2015)

- Probes from Bac clones: Bac clones containing region of interested are amplified, labelled with biotin, digoxigenin or fluorophores and are directly used as probe.

2.3 Data collection

Once the samples are mounted on slides, image acquisition is performed. Depending on the aim and the type of the FISH experiment, different types of microscopy can be performed. The microscopes used in these assay are

- Conventional microscopes
- Confocal
- Super resolution

These microscopes produce digital images which are formed when optical image from the microscope is recorded by a detector. These images display two types of information: spatial (3D localization) and intensity (concentration of probes). A careful acquisition of images is a very crucial step as the intensity values of the image also contains background noise. Thus, a good FISH experiment is limited by the signal to noise ratio of the digital image.

2.4 Image processing and storage

The image processing and storage are very important for any quantitative microscopy experiment. An error in image processing and storage such as bit-depth conversion or image compression, can change the intensity values of the digital images making them a bad substrate for quantitative measurements. Thus, choosing an image-processing algorithm for a quantitative study is a very critical step. A list of the available software packages for quantitative image analysis were submitted as milestone MS24 (Fig 3, http://www.multiscalegenomics.eu/MuGVRE/image-analysis-tools-comparison/).

Tool	Function	Programming Language	GUI	OS (#)	Availability	Doc	Source code	Bio-Formats	Interactive / Flexibility	Macros / Plugins Available	Quantitative and Qualitative Analysis
Icy (*)	Standalone platform to visualize, annotate and quantify bioimaging data	Java	Yes	All	Open source	Yes	Yes	Yes	Yes	Yes	Yes
CellProfiler (†)	Standalone platform for quantitative analysis of biological images	Java, C++ and Python	Yes	All	Open source	Yes	Yes	Yes	Yes	Yes	Yes
Daime	Analysis and visualization program for microbiology and microbial ecology	C++	Yes	Linux, Windows	Open to Academia	Yes	NA	NA	NA	NA	Yes
CMEIAS (‡)	Software of the analysis of microbial ecology	NA	Yes	Windows	Free License	Yes	NA	NA	NA	NA	NA
ImageJ / Fiji	Open source image processing platform	Java	Yes	All	Open source	Yes	Yes	Yes	Yes	Yes	Yes
Volocity (§)	3D imaging software	NA	Yes	Mac OS, Windows	Commercial License	Yes	No	No	Yes	NA	Yes
LAS AF Lite (§)	Standalone tool for image Analysis	NA	Yes	Windows	Free License	Yes	No	No	No	NA	No
Imaris (§)	Standalone tool for visualization and analysis of 3D and 4D microscopy datasets	NA	Yes	Mac OS, Windows	Commercial License	Yes	No	Yes	Yes	NA	Yes
Omero (¶)	Cloud based software to view, organize, analyze and share image data.	NA	Yes	All	Open source	Yes	Yes	Yes	NA	NA	Yes
Bisque (**)	Software to store, visualize, organize and analyze images in the cloud	MATLAB, Python, Java+ImageJ	Yes	All	Open Source	Yes	Yes	Yes	Yes	NA	Yes
FISH-quant	Tool to analyze single molecule mRNA FISH data	MATLAB	Yes	All	Open source	Yes	Yes	NA	NA	NA	Yes
FISH Finder	For automated analysis of FISH images	MATLAB	Yes	All	Open source	Yes	Yes	NA	NA	NA	Yes
TANGO	Tool to high-throughput processing and analysis of 3D fluorescence images	Java, R (on FIJI)	Yes	All	Open source	Yes	Yes	NA	NA	NA	Yes
Nemo	Automatic or semiautomatic FISh and CT analysis	Java	Yes	All	Open source	Yes	NA	NA	NA	NA	Yes
Smart 3D-FISH	An imageJ plugin to spot segmentation and distance measurements in FISH	Java (FIJI)	Yes	All	Open source	Yes	NA	NA	NA	NA	Yes
IMACULAT	Software package for quantitative analysis of chromosome localization in nucleus	Perl (Plug-in)	NA	Linux	Open source	Yes	NA	NA	NA	NA	Yes

Figure 3: Available softwares or analysis tools for DNA FIS analysis (from milestone MS24)

3 Challenges in integrating microscopy based tools on VRE

The quantitative measurement of DNA FISH is susceptible to a variety of potential sources of error. Errors in measurements can be introduced by the specimen, during image acquisition, or the downstream image processing. One of the aims of the MUG project was to integrate DNA FISH based tools to the VRE. However, the microscopy field lacks standard guidelines for all the steps mentioned in section 2. To clarify, we can contrast the current situation for FISH data analysis with that for genomic data analysis. Integration of analysis pipelines for genomic data analysis was feasible since ENCODE consortium has given the guidelines for almost all genomic assays, but no such guidelines exists for the FISH based methods. Depending upon the aim of the project, researchers use either commercial image analysis platforms or in-house data analysis pipelines.

In addition, the analysis of the intensity values of an image should be done on raw images. This led us to another challenge, the **storage of the raw images**. At the current time there are no established databases that store raw image data. Again, we can contrast this with the situation for genomic data: for genomic assays there are databases such as SRA, GEO, EBI Array express etc., where experimentalists can submit their data to be used by the community to train or reanalyze these

datasets in their studies. Through partner EMBL-EBI, the consortium liaised with the BioSamples repository (https://www.ebi.ac.uk/biosamples/) team to look into the possibility of storing imaging data. We are also closely following up on new resources that are arising driven by demand, such as the Image Data Resource (https://idr.openmicroscopy.org/about/) as part of the partners' commitment towards VRE sustainability.

Furthermore, there is no consensus regarding the **processing of the raw images**, thus we can not usefully link any of the open source platforms such as ImageJ to the VRE. In future, if a consensus is reached regarding a specific open source platform recommended by the community, it will be integrated in the VRE.

In the case of WP7.1 the senescence project, we performed DNA FISH with the aim to quantify

- distance between the SAHDs (mentioned in deliverable 7.1)
- SAHD distance to periphery
- area of the SAHDs, as a measure of the compaction (mentioned in deliverable 7.1)

For this, oligopaint based FISH probes were used to perform DNA FISH. For the analysis of the data "IMARIS" software was used. However, the software is commercial and not open source, thus we were not able to integrate it in the VRE.

As part of the sustainability roadmap, MuG is committed to keeping the tool offer up to date with the latest advances. The MuG tool wrapping API will facilitate the inclusion of image analysis tools as they become available in the market and as their use is standardized. Driven by demand, technology companies are becoming progressively interested in the development of tools to fill this gap.

To this end, MuG has made contact with KML Vision OG (Austria, https://www.kmlvision.com/) and Wimasis (Spain, https://www.wimasis.com/en/) as potential candidates to integrate their tools into the VRE in the future.

4 REFERENCES

1. Bonev, B. et al. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. 17, 661–678.
2. Sexton, T. et al. (2015). The role of chromosome domains in shaping the functional genome. Cell. Mar 12;160(6):1049-59.
3. Sati, S., et al. (2017). Chromosome conformation capture technologies and their impact in understanding genome function. Chromosoma. Feb;126(1):33-44.
4. Spielmann, M. et al., (2018). Structural variation in the 3D genome. Nat Rev Genet. Jul;19(7):453-467.
5. Lieberman-Aiden, E., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. Oct 9;326(5950):289-93.
6. Sexton, T. et al. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. Feb 3;148(3):458-72
7. Dixon, J.R., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. Apr 11;485(7398):376-80.

8. Szabo, Q. et al. (2018). TADs are 3D structural units of higher-order chromosome organization in Drosophila. Sci Adv. Feb 28;4(2):eaar8082.
9. Rao, S.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. Dec 18;159(7):1665-80.
10. Bonev, B., et al. (2017) Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell. Oct 19;171(3):557-572.e24.
11. Ogiyama, Y., et al., (2018). Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. Mol Cell. Jul 5;71(1):73-88.e5.