Pràcticum I. Final Report

Nom	Albert Torelló Pérez	**Data matriculació PI**	6-9-2017
Grau	Ciències Biomèdiques	**Grup de recerca**	Genome Archtecture; Gene Regulation, Stem Cells and Cancer Programme; Centre for Genomic Regulation
Període de realització	1st Semester 2017	**Correu electrònic**	alberttorelloperez@gmail.com
Data d'entrega		**Tutor extern**	Dr. Guillaume Filion
Tutor intern	Dr. Ferran Reverter		

Contents

1. Introduction

1.1 Mission of the CRG

According to its website, the Centre for Genomic Regulation (CRG) is an international biomedical research institute of excellence, created in December 2000. It is a non-profit foundation funded by the Catalan Government through the Departments of Economy & Knowledge and Health, the Spanish Ministry of Economy, Industry and Competitiveness, the "la Caixa" Banking Foundation, and includes the participation of Pompeu Fabra University.

They also stated that the mission of the CRG is to discover and advance knowledge for the benefit of society, public health and economic prosperity. The CRG believes that the medicine of the future depends on the ground-breaking science of today. This requires an interdisciplinary scientific team focused on understanding the complexity of life from the genome to the cell to a whole organism and its interaction with the environment, offering an integrated view of genetic diseases.

1.2 Genome Architecture Lab

The Genome Architecture Laboratory was established in 2012 by Dr Guillaume Filion, who also supervised the work presented in this report.

During this period, he has established several research lines, but they have in common one thing: they are all focused on the understanding the chromatin and topology effects on gene expression and nuclear biology. As the sequence is not enough to explain the variation in gene expression across time and space, there is a need to identify the other factors involved in transcription regulation.

The most important research lines, among others not named, are:

- Position-dependent effects on housekeeping promoters in *Drosophila*.
- Position effects on HIV integration and expression.
- Polymer physics of chromatin and properties of transcription factors mechanics.
- Nuclear organization driven by progesterone receptor and oestrogen receptor in breast cancer cells elucidated using super-resolution light microscopy.

1.3 Purpose of the project

The aim of the project (not only of this practicum) is to understand the factors involved in the expression of a housekeeping promoter integrated into heterochromatin regions (Corrales *et al.* , 2017).

To do that, I made use of TRIP technology. Briefly, TRIP consists in the use of a library of barcoded reporters (reporters with a 20-mer random sequence) that are integrated randomly in the genome using a transposase. This system allows us to relate the expression of each unique reporter to a location in the genome. Using this technology enable us to study the position effects in a high-throughput way (Akhtar *et al.* , 2013, 2014).

To address the question above, I cloned a barcoded before the poly-A tail. I amplified the library using electrocompetent *E. coli* cells and cotransfected it into Kc167 cells (a *Drosophila* cell line) with the *Sleeping Beauty* transposase (Mátés *et al.* , 2009). After growing the cells, I FACS selected them and seeded 1 cell/well in a 96-well plate and grow them. The goal of this process is to get isogenic populations (cells that come from a single cell progenitor) that putatively contain at least 1 integration.

After growing these clones, I harvested the cells and extracted the genomic DNA in order to map the barcodes in the genome. The will of that is to get a clone with at least 2 mapped integrations, 1 in euchromatin and the other one in heterochromatin.

The workflow is summarized in figure 1.

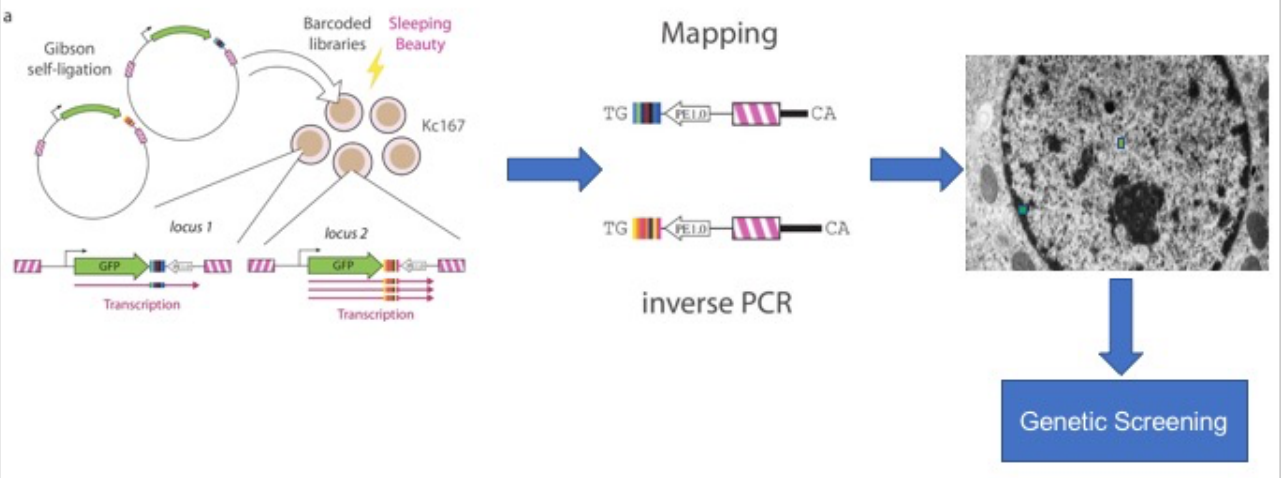


Figure 1: Workflow of the project. The green squares represent the GFP integrations

This will allow us to screen this cells for factors involved in the heterochromatic expression of this promoter of interest.

1.4 Position effects variegation

As this practicum aims to disentangle the mechanisms of a position effects event, it is important to properly describe it.

Defined by one of the most important contributors of the field Dr Sarah Elgin, position effects variegation (PEV) results when a gene normally in euchromatin is juxtaposed with heterochromatin rearrangement or transposition. When heterochromatin packaging spreads across the heterochromatin/euchromatin border, it causes transcriptional silencing in a stochastic pattern.

PEV is intensely studied in *Drosophila* using the *white* gene (Elgin & Reuter, 2013). Muller in 1930 described a phenotype in *Drosophila* in which the eye was variegating with some white facets altering the typical red pigmentation. This observation drove to the conclusion that *white* gene was not mutated. After the examination of the polytene chromosomes, he confirmed that this phenotype was due to an inversion or rearrangement with one breakpoint within the pericentromeric heterochromatin (Elgin & Reuter, 2013).

Traditionally, the definition of heterochromatin was mainly cytological. Heterochromatin was defined as such regions that appeared electrodense in transmission electron microscopy experiments (Rego *et al.* , 2008). The figure 2 exemplifies this fact.

The screening of factors involved in PEV using the *white* gene and a transposon led to the discovery of proteins and histone modifications that define heterochromatin by its biochemical features. As result of that, there were characterized the suppressors and the enhancers of variegation; proteins, if they were mutated, suppress or enhance the variegated phenotype. As a consensus, heterochromatin is composed of chromatin domains enriched with Su(var)3-9, HP1, and the HP1-interacting proteins LHR and HP6, and marked by H3K9me2 (Elgin & Reuter, 2013; Filion *et al.* , 2010).

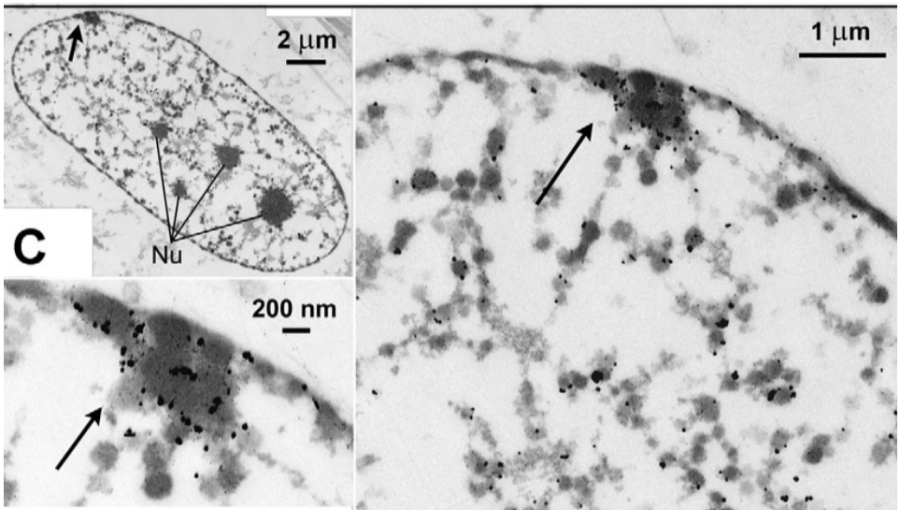


Figure 2: H3K3mK27 (rabbit antibody) immunogold labelling decorating the Xi heterochromatin periphery. Arrowheads and arrows point to the Barr body; Nu, nucleolus. Picture adapted from Rego *et al.* , 2008.

1.5 The TRIP technology

Following the thread exposed above (section 1.3), TRIP was developed to assay the position effects but in a genome-wide and parallel way. TRIP was, fundamentally, a combination of the traditional transgene reporter assay with random barcoding technology and high-throughput sequencing (Akhtar *et al.* , 2013). As far as TRIP can be used to study diverse regulatory elements (promoters, enhancers, insulators) and DNA-binding proteins, here I used TRIP to study a promoter.

TRIP is based on a large library of reporter constructs. These reporters include a short DNA sequence of 20 bp long, which I refer to as a barcode. These reporters are randomly integrated throughout the genome in a pool of cells of interest, and subsequent reporter activity is determined by the chromatin environment at the site of integration. The barcode associated with each reporter serves as a UMI (unique molecular identifier) for tracking of the reporter location and activity. The location is determined using an inverse PCR (iPCR)-based method, and the expression is quantified by Illumina sequencing of barcodes (as the barcode is integrated before the poly A tail). The count for each barcode in RNA is normalized by the corresponding count in genomic DNA (Akhtar *et al.* , 2014).

2. Objectives and activities

2.1 Aim of the practicum

The main goal of this practicum is to get a complete barcoded library with enough complexity and integrate the reporters in Kc167 cells. Complexity is a term to refer to the diversity of barcodes, in other words, how many different barcodes contain the library.

The first step of the library construction is to clone the promoter of interest upstream the reporter gene, in this case, GFP (Green Fluorescent Protein). After that, the barcode is added to the vector backbone by PCR using custom primers which one of them contains a 20-mer random (the primer sequences are in section 5.1) sequence obtaining a linearised sequence. Then, the template is destroyed by digesting it with DpnI. After purifying the PCR products, the plasmid is re-ligated with Gibson cloning Gibson *et al.* (2009). Products are purified again

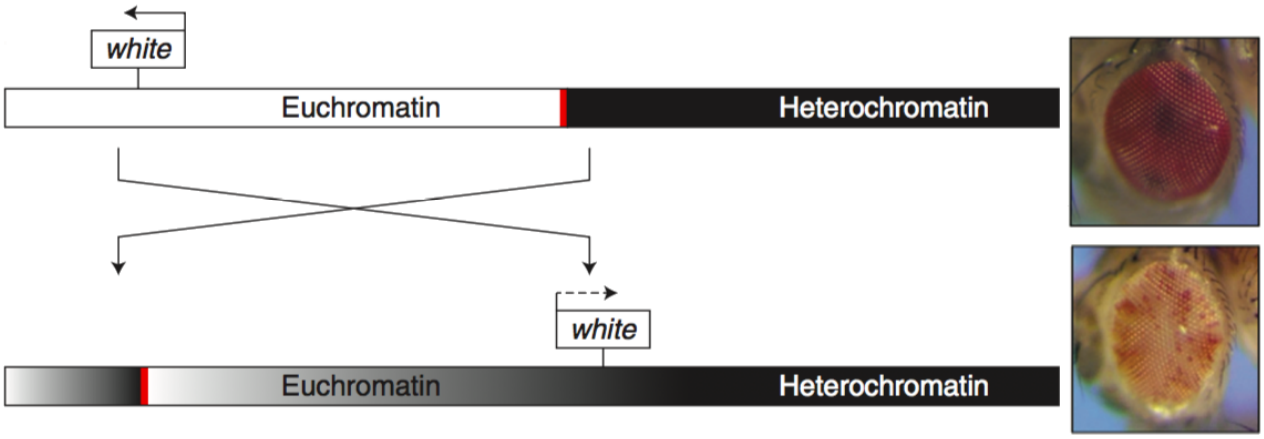


Figure 3: Schema of the X-ray induced rearrangement places the white locus, normally located in the distal euchromatin (white bar) of the X chromosome (see the top line), 25 kb from a breakpoint in the pericentric heterochromatin (black bar; bottom rearranged line). Spreading of heterochromatin packaging into the euchromatic domain results in silencing (causing a white eye in this case); loss of silencing in some cells during differentiation results in a variegating phenotype (bottom line, right). Adapted from Elgin & Reuter, 2013

and transformed to electrocompetent *E. coli* cells in order to amplify the library. An important constraint here is to check the complexity of the library, in other words, to estimate the number of unique molecules you get (plasmids with a different barcode). It is considered a good library these ones with a complexity higher than 1,000,000. Then, the bacterial cultures are grown overnight and plasmid DNA is isolated using a commercial kit.

Once the library is fully constructed, the library is co-transfected with a plasmid carrying the Sleeping Beauty transposase, which will allow the random integration of the barcoded reporters. Then, the cells are put in 2 heat-shocks at 37ºC in order to allow the transposase be expressed (it was cloned under the control of a thermo-inducible promoter) and because it is the optimal temperature for the transposase.

3 days after transfection, GFP expression is checked by microscopy and FACS.

2.2 Experimental procedures

2.2.1 Barcoding PCR

The first step is to do the barcoding PCR according to this protocol:

Reagent
10 ng plasmid
2.5 μL primer GAT155 10μM
2.5 μL primer GAT385 10μM
10 μL GC buffer 10x
1 μL dNTPs 10 mM
0.5 μL Phusion Polymerase
32.5 μL water

with this programme:

T	Time
98°C	2 min
98°C	15 sec
60°C	20 sec
72°C	5 min
x 28 cycles	
72°C	2 min

I checked the outcome of PCR in a 1% agarose gel.

2.2.2 DpnI digestion

The recipe is:

Reagent
45 μL barcoding PCR product
5 μL CutSmart Buffer 10x
1 μL DpnI enzyme

I incubated the mix 1h at 37°C and 20 min at 80°C to deactivate the enzyme.

2.2.3 MinElute kit (Qiagen) purification

This step is necessary to clean up PCR products of primers, nucleotides and polymerase.

1. Add 5 volumes of PB buffer for 1 volume of PCR product and 10 μL of NaAc 3M.
2. Put the column in a 2 mL collection tube.
3. Spin the column 1 min at 13,000 rpm and discard the flow-through.
4. Add 750 μL of PE buffer and spin at 13,000 rpm for 1 min. Discard the flow-through.
5. Spin for 2 min and replace the collection tube (1 mL tube).
6. Elute in 10 μL EB buffer. Repeat the action and wait 5 min and spin for 1 min.

2.2.4 Gibson Cloning

Reagent
5 μL of template DNA
5 μL of water
10 μL of Gibson mix

I incubated this mix 1h at 50°C.

2.2.5 Ethanol precipitation

1. Add 2.5 volumes of pure ethanol.
2. Add 1/10 of sodium acetate 3M and 2 μL of glycogen.
3. Immerse in a mix of ethanol and carbonic snow or store it in a -80ºC freezer 1 hour to freeze and precipitate DNA.
4. Centrifuge for 45 min at 16,000 rpm at 4ºC.
5. Discard the supernatant.
6. Wash the pellet with 75% ethanol and centrifuge for 15 min at 16,000 rpm at 4C.
7. Resuspend the pellet in 10 μL of water.

After that, I checked the concentration of DNA:

conc ng/μL	260/280	260/230
508	1,62	1,92

2.2.6 Transformation of high electrocompetent *E. coli* cells

I diluted the plasmid from the previous experiment to 125 ng/μL so it is the optimal concentration for electroporation.

1. Thaw an aliquot of electrocompetent *E. coli* in ice.
2. In 0.1 cm electroporation cuvettes (Bio-Rad), pipette 25 μL of *E. coli* cell suspension and 2.5 μL of the purified plasmid.
3. Electroporate with the corresponding programme (1 pulse of 1,80 kV) using the MicroPulser Electroporator from Bio-Rad.
4. Add 1 mL of recovery medium (SOC media).
5. Pool the transformed cells in a 15 mL falcon.
6. Prepare 3 LB + Ampicillin agar plates and a dilution bank of the culture (1/1000,1/2000,1/4000):

Dilution	Culture volume	Final volume
1/1000	4 μL	200 μL
1/2000	2 μL	200 μL
1/4000	1 μL	200 μL

7. Plate this 200 μL in this plates and culture overnight at 37ºC.
8. Transfer the rest of the culture into 500 mL of liquid LB + ampicillin 1X (500 μL of 1,000X stock) and culture overnight at 37ºC.

2.2.7 MaxiPrep NucleoBond Xtra Midi kit

1. Harvest cells by centrifuging at 4,500g at 4ºC for 15 min.
2. Resuspend the pellet in 12 mL of RES buffer, add 12 mL LYS buffer and wait 5 min at RT.
3. Equilibrate the column with 25 mL of EQU buffer.
4. Add 12 mL of NEU buffer to the lysate.
5. Mix the tube and load the lysate into the column.
6. First wash. Add 15 mL of buffer EQU.
7. Discard the filter.
8. Second wash. Add 25 mL of buffer WASH.
9. Add 25 mL of ELU buffer.
10. Add 10.5 mL of isopropanol, vortex and centrifuge at 4,000 rpm for 1h at 4ºC.
11. Discard the supernatant and add 4 mL of 70% ethanol and centrifuge at 4,000 rpm at 4ºC for 15 min.
12. Discard the supernatant and resupend the pellet in 100 μL of nuclease-free water.
13. Quantify the library in a Nanodrop and leave the maxiprep at 1μg/mL.

2.2.8 Cell culture and transfection

Kc167 cells were grown at 25ºC with Schneider's *Drosophila* medium supplemented with 5% FBS and antibiotics. Cells were passed twice a week with a split ratio 1/4.

They are needed 20 million cells per transfection.

1. Prepare as many plates as you need (1 plate for transfection) in order to get 20 million cells on the day of transfection.
2. Centrifuge the culture at 800g for 3 min.
3. Remove the supernatant and resuspend the cells in 800 μL x number of transfections.
4. Pipette 20 μg of DNA in an electroporation cuvette (Biorad, 0.4 cm gap). This 20 μg were distributed: 10 μg of barcoded plasmid and 10 μg of Sleeping Beauty plasmid.
5. Add 800 μL of cell suspension and pipette up & down 2-3 times.
6. Electroporate at 1000 μF and 250 V.
7. Wait 5 min.
8. Transfer the cells into a 10 cm cell culture dish containing 10 mL of Schneider medium + 5% FBS previously warmed at RT.
9. Grow cells at 25ºC.

The next day I did 2 heat shocks at 37ºC of two hours each, with at least four hours recovery between them and the day after I did another heat shock in order to induce the expression of the transposase.

Then, I checked the expression of GFP by microscopy and flow cytometry.

3. Results and Conclusion

3.1 Barcoding PCR

I checked the outcome of this PCR running a 1% agarose gel:

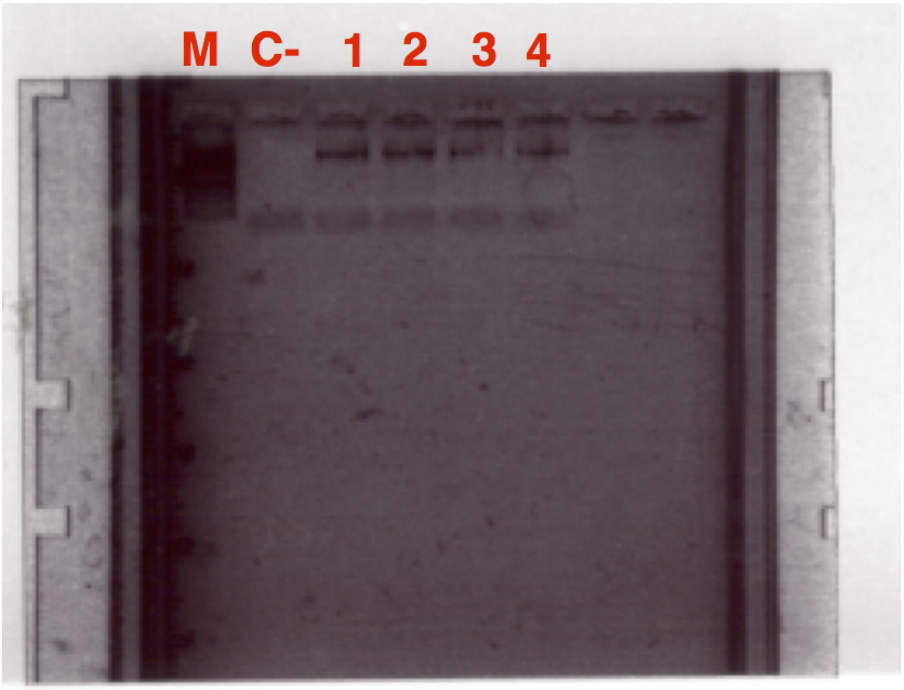


Figure 4: 1% agarose gel showing a 5000 bp band, according to the plasmid length. M is the molecular weight marker, C- is the negative control and 1,2,3,4 are the technical replicates.

The negative control is clean, so apparently, there is not unspecific amplification. The band around 5000 bp indicate the amplification runs well and the absence of other bands warranties the integrity of the vector.

3.2 Complexity of the library and plasmif yield

As I stated before, it is necessary to estimate the complexity of the barcodes; in other words, to estimate the number of different barcodes I got. For that purpose, I prepared 3 agar plates, in which I seeded a known culture volume to estimate the CFU (Colony Forming Units). I used CFU as an estimation if the complexity of the library.

I got the following results:

Dilution	CFU/mL
1/1000	1.000.000
1/2000	1.160.000
1/4000	1.020.000

The complexity is about 1.000.000 CFU/mL (500.000.000 of different barcodes), which is satisfactory so I continued with the plasmid purification and isolation.

After purifying the plasmid, I measured the concentration in a Nanodrop device (special spectrophotometer to measure the concentration of nucleic acids):

conc ng/μL	260/280	260/230
2583	1.94	2.35

With this concentration, I reached a yield of 25 μg of the plasmid. I diluted the plasmid until got 1 ug/μL. I transfected Kc167 cells with this library.

3.3 GFP expression

3 days after the transfection, I checked the expression of GFP by microscopy and flow cytometry:

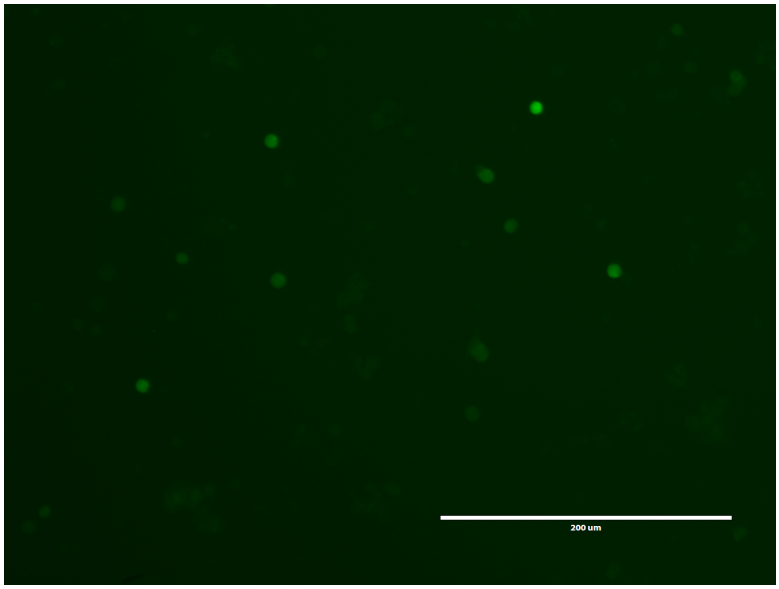


Figure 5: Photograph showing a representative field of cell culture 3 days after transfection. There were enough cells expressing GFP.

The flow cytometry showed that there were 10% of cells expressing GFP with a geometric mean of 2.68.

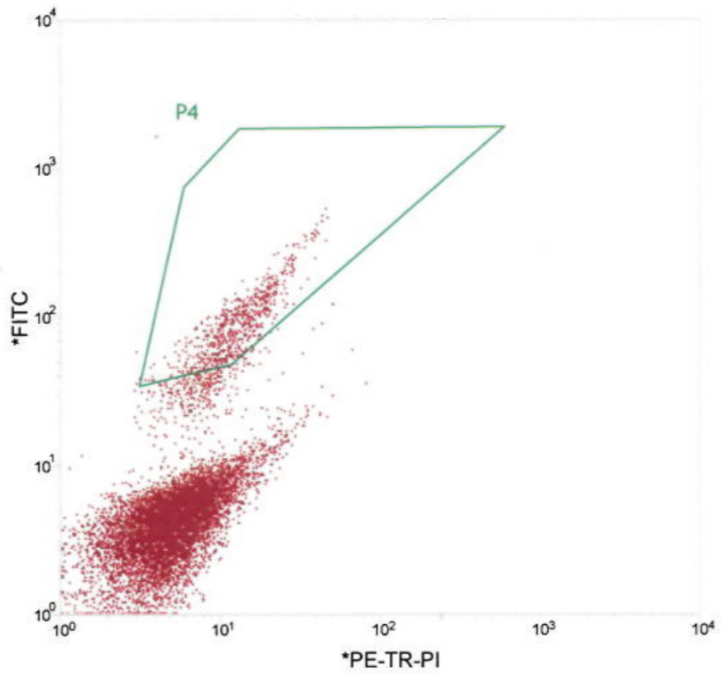


Figure 6: Scatter plot of the events passed for the flow cytometer. Each dot is an event (actually a cell). P4 highlighted the population expressing GFP.

The fraction of the population expressing GFP seems good and I continued with the experiments; which will be part of the Practicum II.

4. Self-assessment

During the development of this practicum, I practised what I learnt about molecular biology, genetic engineering, cell culture and gene regulation.

I accomplished the objectives set at the beginning of the practicum. However, I realized that research timings are very difficult to predict and they can change as the project goes on.

Moreover, I also acquired new skills in molecular biology such as primer design, PCR setup, Sanger sequencing, plasmid manipulation and sample preparation for next-generation sequencing. I also applied cell biology-related techniques like FACS and fluorescence microscopy. As the range of techniques I used is so wide, I consider this practicum has been useful and productive.

References

Akhtar, Waseem, De Jong, Johann, Pindyurin, Alexey V., Pagie, Ludo, Meuleman, Wouter, De Ridder, Jeroen, Berns, Anton, Wessels, Lodewyk F A, Van Lohuizen, Maarten, & Van Steensel, Bas. 2013. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell*, **154**(4), 914–927.

Akhtar, Waseem, Pindyurin V, Alexey, de Jong, Johann, Pagie, Ludo, ten Hoeve, Jelle, Berns, Anton, Wessels FA, Lodewyk, van Steensel, Bas, & van Lohuizen, Maarten. 2014. Using TRIP for genome-wide position effect analysis in cultured cells. *Nature Protocols*, **9**(6), 1255–1281.

Corrales, Marc, Rosado, Aránzazu, Cortini, Ruggero, van Arensbergen, Joris, van Steensel, Bas, & Filion, Guillaume J. 2017. Clustering of *Drosophila* housekeeping promoters facilitates their expression. *Genome research*, **27**(7), 1153–1161.

Elgin, Sarah C R, & Reuter, Gunter. 2013. Position-Effect Variegation, Heterochromatin Formation , and Gene Silencing in Drosophila. *Cold Spring Harbor Perspectives in Biology*, **5**, 1–26.

Filion, Guillaume J., van Bemmel, Joke G., Braunschweig, Ulrich, Talhout, Wendy, Kind, Jop, Ward, Lucas D., Brugman, Wim, de Castro, In??s J., Kerkhoven, Ron M., Bussemaker, Harmen J., & van Steensel, Bas. 2010. Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell*, **143**, 212–224.

Gibson, Daniel G, Young, Lei, Chuang, Ray-Yuan, Venter, J Craig, Hutchison, Clyde a, Smith, Hamilton O, Iii, Clyde A Hutchison, & America, Nature. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, **6**(5), 343–5.

Mátés, Lajos, Chuah, Marinee K L, Belay, Eyayu, Jerchow, Boris, Manoj, Namitha, Acosta-Sanchez, Abel, Grzela, Dawid P., Schmitt, Andrea, Becker, Katja, Matrai, Janka, Ma, Ling, Samara-Kuko, Ermira, Gysemans, Conny, Pryputniewicz, Diana, Miskey, Csaba, Fletcher, Bradley, Vandendriessche, Thierry, Ivics, Zoltán, & Izsvák, Zsuzsanna. 2009. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nature Genetics*, **41**(6), 753–761.

Rego, A., Sinclair, P. B., Tao, W., Kireev, I., & Belmont, A. S. 2008. The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure. *Journal of Cell Science*, **121**(7), 1119–1127.

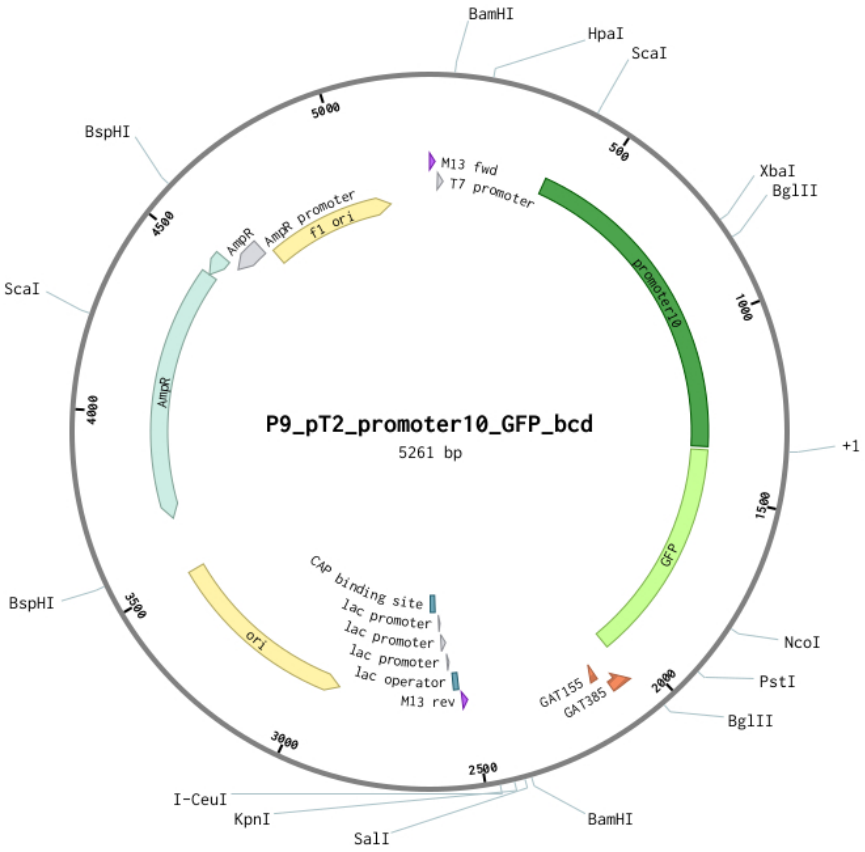
5. Addendum

5.1 Primer sequences

- GAT155: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAAAGGCAATGCTAC-CAAATAC
- GAT385: ACGCTCTTCCGATCTNNNNNNNNNNNNNNNNNNNNCATGCTAGTTGTG-GTTTGTC

5.2 Plasmids

5.2.1 GFP barcoded plasmid



5.2.2 Sleeping Beauty plasmid