2017

Optimization of a mRNA based *in situ* padlock probe approach

Trabajo de Fin de Grado/Bachelor Thesis

Trabajo de Fin de Grado performed at the Institute of Cell Biology, Histology and Embryology, which belongs to the Medical University of Graz.

Universidad Zaragoza

Javier Alaiza Gil
Grado en Biotecnología. Universidad de Zaragoza
28/06/2017

Facultad de Ciencias
Universidad Zaragoza

Este documento deberá ser entregado por el estudiante en la Secretaría de la Facultad dentro del plazo establecido al efecto

ANEXO IV

AUTORIZACIÓN DEL DEPÓSITO DEL TRABAJO FIN DE GRADO EN BIOTECNOLOGÍA

(VISTO BUENO DIRECTORES/PONENTES)

ESTUDIANTE

Apellidos: Alaiza Gil **Nombre:** Javier

TITULACIÓN: Grado en Biotecnología

TÍTULO DEL TRABAJO: (en castellano y en inglés)

Optimización de una estrategia de sondas candado basadas en mRNA.

Optimization of a mRNA based in situ padlock probe approach.

Dpto. Responsable de la propuesta: Inst. of Cell Biol./Histol.&Embry./Medical University Graz-(AUT)

- Los abajo firmantes, dan su visto bueno y autorizan el depósito de la memoria correspondiente al Trabajo Fin de Grado arriba indicado, realizado bajo nuestra tutela y que se ajusta a los criterios que marcan la normativa y la Guía Docente correspondiente.

- Y autorizan su publicación en el repositorio institucional de la UZ. SI ☒ NO ☐

Zaragoza, 15 de Mayo de 2017

Fdo.: Amin El-Heliebi	Fdo.:	Fdo.: Milagros Medina
Director	Director	Ponente

(Esta autorización deberá ser firmada por el director o directores del Trabajo Fin de Grado, así como el ponente si lo hubiere)

C/ Pedro Cerbuna, 12 • 50009 Zaragoza • Telf.: 976761294 • Fax: 976761159 •
E-mail: admcienz@unizar.es

Optimization of a mRNA based *in situ* padlock probe approach

Autor / Author: Javier Alaiza Gil

Director: PhD Amin El-Heliebi

Ponente / Rapporteur: PhD María Milagros Medina Trullenque

This project has been carried out at the Institute of Cell Biology, Histology and Embryology, belonging to the Medical University of Graz, Austria.

Acknowledgements

I thank Amin El Heliebi for his generosity and help throughout my stay in Graz and afterwards, as well as for all the patience shown when dealing with bureaucracy.

I thank María Milagros Medina Trullenque for accepting to be my rapporteur and for all the help she provided.

I thank María Luisa Peleato Sánchez for accepting to be my tutor in the University of Zaragoza and dealing with bureaucracy related to Erasmus + programme during the summer of 2016.

I especially thank Pablo Zardoya Laguardia for all the immeasurable help he has selflessly provided to me. I cannot express how much your constant support, even when you were busy enough with your own thesis, meant to me. Not many people would go to such lengths in order to help others. I wish you every bit of luck that dwells in placentas after you obtain your PhD.

Index

Abstract

English

In situ padlock probes experiments are a common tool in single cell genomics. When applied to detect mRNA sequences, they allow quantification and localization of targeted transcript in each individual cell. These assays are mainly performed on cells attached to a glass slide or fixed tissue samples. However, if the padlock probes could be applied on cellular suspensions, the technique could be combined with a wide range of follow-up techniques, such as micromanipulation.

In order to test this, in situ padlock probing assays were performed on cellular suspensions and on cell seeded slides. The probes targeted mRNAs for β-actin and androgen receptor full length and were employed in three different cell lines: LNCaP, VCaP and HT-29. The results were observed by fluorescence microscopy and quantified by CellProfiler. In order to test the feasibility of a low coverage genetic sequencing, single cells were selected from the probed cellular suspension via micromanipulation, its DNA was amplified using whole genome amplification and the quality of the amplification product was verified by multiplex PCR.

In situ padlock probes assays could be applied to cellular suspensions and expression of the transcripts could be quantified and located in a similar manner to the methods for cells attached onto slides. The number of signals between both supports showed comparable results. Hence any further modifications of the protocol should be focused on improving its efficiency. DNA from probed cells might be of good quality to undergo sequencing.

Castellano

Los ensayos in situ de sondas candado son una herramienta común en genómica de células individuales. Cuando se utilizan para detectar secuencias de mRNA, permiten cuantificar y localizar la expresión del transcrito diana en cada célula individual. Estos ensayos se realizan principalmente en células adheridas a láminas de cristal y en muestras de tejido fijadas, sin embargo, si pudiesen ser aplicadas en células en suspensión, se podrían combinar con un amplio rango de técnicas, como por ejemplo la micromanipulación.

Con objetivo de probar esto, se realizaron ensayos in situ de sondas candado en suspensiones celulares y en placas sembradas con células. Las sondas detectaban los mRNA de actina β y del receptor de andrógenos de longitud completa y fueron empleadas en tres líneas celulares distintas: LNCaP, VCaP y HT-29. Los resultados fueron observados en microscopio de fluorescencia y cuantificados usando CellProfiler. Para comprobar la posibilidad de usar una secuenciación de baja cobertura en muestras que hayan sufrido el sondeo, se seleccionaron células individuales mediante micromanipulación, su DNA fue amplificado usando secuenciación de genoma completo y la calidad del producto amplificado fue comprobada por PCR multiplex.

El ensayo de sondeo in situ por sondas candado pudo ser aplicado a suspensiones celulares y sus resultados pudieron ser cuantificados y localizados usando métodos similares a aquellos utilizados en células adheridas. El número de señales detectadas mostró resultados comparables entre ambos soportes por lo que cualquier modificación posterior debería centrarse en mejorar su eficiencia. El material genético obtenido de células sondeadas podría tener la calidad suficiente para justificar la secuenciación.

Background

Large scale assays that analyze the content of a cell population tend to miss molecules present in rare cells and can't relate what cell produced each molecule. What is more important, a single cell may exhibit different expression patterns than the cell population mean. Thus, single cell assays are needed to study individual variations in genome and transcriptome (1,2). *In situ* padlock probes and its capabilities when combined with rolling circle amplification have been known for more than 15 years (3). Single nucleotide polymorphism (SNP) detection via padlock probes has been traditionally used to detect and genotype several pathogenic point mutations in mitochondrial DNA (4). *In situ* padlock probes have also been applied to detect repetitive genomic sequences (5), bacterial (6) and viral infections (7) via detection of pathogen-specific DNA sequences; and detection of short single-copy sequences within genomic DNA in human cells (8). Then, the **method was adapted to detect mRNA** sequences and sequence variants (1), allowing a remarkable improvement in the field. The study of gene expression via *in situ* padlock probing techniques provides a **method to localize and quantify the expression of transcripts** (1,2,9) or proteins (2).

When used to quantify mRNA expression, *in situ* padlock probing has been **mainly performed on cells attached to glass slides** (1,2,9) **or on fixed tissue samples** (1,9,10). Although efficiency of *in situ* padlock probing has been estimated to be near 30% of the detected by quantitative polymerase chain reaction (PCR) for β-actin transcripts, **its results correlate with the relative expression levels in different cell lines and the detection efficiency is consistent among different transcripts** (1).

In order to quantify the expression of a transcript from a sample, firstly it is mandatory to isolate its mRNA, then convert it to cDNA by reverse transcription and ribonuclease H (RNase H) digestion (1,2,9). This technology is based on highly selective probes: linear oligonucleotides of approximately 70-100 nucleotides (nt) in length. The 5' and the 3' ends contain the complementary sequence to the target flanking sequence, thus allowing probe-specific union. Hybridization of the padlock probe to the target renders both probe ends next to each other in a head-to-tail orientation (1,2). The ends will then be joined through enzymatical DNA ligation and the probe will be circularized. Ligation between 5' and 3' ends assures high enough selectivity to discriminate single nucleotide variations in the target sequence, as it will only occur if those ends are perfectly complementary to the target (1,2). The padlock probe also counts with a backbone region. The backbone contains a sequence which can be targeted by an oligonucleotide, which is able to conjugate with a fluorophore or a primary antibody, in order to detect and quantify the expression (2). Ligation of padlock probe allows performing a rolling circle amplification which increases the number of backbone sequences and increases the posterior signal to detectable levels (2).

In situ padlock probing assays have shown **great specificity** and its **utility in medical diagnosis**, to the point that it had been possible to simultaneously detect different fusion transcript variants, clinical relevant point mutations, and differentially expressed genes(9). Single cell genomics are useful tools for the understanding of several diseases along with their diagnosis(11). However, effectively detecting SNP and other base alterations in single cells has proven to be **challenging because of the pre-requisite of whole genome amplification**. Despite the advances on the technique, whole genome amplification still creates artefacts that can lead to **false detections** (12).

If *in situ* padlock probing assays could be **combined** with successful whole genome amplification, this **constraint would be eliminated**. The amount of information obtained from a single cell would considerably increase, resulting in a better understanding of genomics and transcriptomics, an easier study of rare cells as well as an improvement in the diagnosis, prognosis and overall knowledge of genetically based diseases, such as cancer.

Regarding cancer, **copy number alterations** are relevant disease markers (13,14). Copy number alterations are mutations that **change the number of copies** of genes, entire chromosomes and parts of chromosomes. Many of them are pathogenic and they can have a wide range of effects, not only in cancer but also in different genetic diseases (14). If the **product of *in situ* padlock probes** assays could be analysed, for example, using a low coverage **genomic sequencing** (15), it would be possible to **study copy number alteration along with gene expression**.

Aims and approach

In situ padlock probes have been used on samples – cells or tissues – attached to a support, for example, glass slides. However, if we could apply the technique to cells in suspension and combine it with a wider range of methods, we would then increase the information that is obtained from a single set of cells.

Objectives

The purpose of this thesis is to determine whether *in situ* padlock probing assays can be adapted to cell suspensions. If so, it aims to: (i) compare the technique performance in both supports, (ii) optimize *in situ* padlock probing protocol for cell suspensions and (iii) test if the resultant material meets the quality requirement to undergo subsequent analyses.

(i) In order to compare new results from cell suspension assays to previous experiments performed on seeded slides, certain points should be addressed. **First** of all, it must be determined whether results obtained from *in situ* padlock probing assays performed on cells in suspension can be compared to those obtained from cells attached to a slide. **Second**, it should be assessed what is the overall efficacy of the technique compared to the one performed on slide and what factors may affect its outcome.

(iii) As it is impossible to test every existent analytical technique after *in situ* padlock probing, the project is focused on elucidating whether a low coverage genetic sequencing can be used after it. This is especially interesting in cancer diagnosis, as it would allow identifying copy number alterations.

Structure of the project

A major part of the project has aimed to compare the performance of *in situ* padlock probes experiments targeted to mRNA. The experiments were performed in parallel on both settings (cellular suspensions and cells attached onto glass slides). In order to include more variables, three different cell lines were used and the technique was directed to detect two targets. The targets are the transcripts for β-actin and AR-FL (androgen receptor full length). β-actin is expected to be present in all three, while the transcript for androgen receptor should not be found in HT-29 cells due to the fact that they do not express it. Only VCaP in suspension were used in micromanipulation and whole genome amplification due to the higher cost of the latter.

The overall structure of the project was the following: The three cell lines were cultured. Half of cultured cells were taken and split; then they were fixed in formaldehyde and stored in a glycerol stock at -20ºC. *In situ* padlock probe protocol was performed on cells seeded on slides (which had been stored at -80ºC until required) and on cell suspension aliquots from fixation. Once *in situ* padlock probing was accomplished, part of the cell suspension was put on a slide. Then, images were taken using a fluorescence microscope and quantified by Cell Profiler (16), an open source software.In parallel to that, one VCaP aliquot followed a different set of experiments: Once the probing protocol was over, a micromanipulator was utilized to select single cells and transfer them into PCR tubes. Whole genome amplification was performed to have enough DNA for sequencing. Instead of sequencing, a Control Multiplex PCR was used to assess its quality.

Materials and Methods

This thesis involves a significant amount of work with cellular suspensions. Because of that and in order to avoid needless repetitions, it must be considered that every centrifugation involving cells in suspension inherently contains the removal of supernatant. Cells would always settle to the bottom of the centrifuged recipient and centrifugation steps were performed with the intent of removing the liquid medium that contains them, unless it is stated otherwise. There is a list of initials used throughout the text available on Annex II.

Cell culture

Cell lines were cultured in an incubator humidified at 37ºC with 5% CO_2 using different media in order to provide a supply for cellular suspensions. The composition of cell culture medium is available on Annex II. The three cell lines employed are LNCaP, VCaP and HT-29. All of them are epithelial adherent cells derived from human tumors.

After incubation, culture media was removed and cell cultures were washed with HBSS solution. A brief (5 min) incubation with Accutase (Invitrogen) at 37ºC loosened cell adherence to the flask. Culture medium was added to deactivate Accutase and the detached cell suspension was split in two tubes; one would be cultured to maintain the cell line, the other would be used in the following assays. The tubes were then centrifuged (300 *g*, 5 min, room temperature (RT)). Those intended for continuation the cell culture were resuspended in culture media and transferred to T25/T75 culture flasks (Sigma), then placed in the incubator. The tubes intended to be used for experimentation were resuspended in DEPC-PBS and employed for the next step: fixation.

Fixation

Cells from culture were centrifuged (300 *g*, 5 min, RT) and resuspended in DEPC-PBS twice before the pellet was mixed with 3.7% formaldehyde (Sigma) buffered in DEPC-PBS. After 15 min, DEPC-PBS was added to dilute formaldehyde and the solution was centrifuged (300 *g*, 5 min, RT). Formaldehyde was removed and the pellet was washed with DEPC-PBS, centrifuged again and then resuspended in 500 µl of PBS. Cells were counted using a Neubauer chamber (Marienfield-Superion, Visionären). Cells were split into aliquots containing at least 10^6 cells, then centrifuged and resuspended in 30 µl of PBS. Finally, 150 µl of glycerol (Sigma) (50%, diluted in DEPC-H_2O) were added to each aliquot, which was then stored at -20ºC. This would provide a stock to perform future experiments and compare its results to those obtained in this thesis, in example, starting from the same cell aliquot.

In situ padlock probe assay

In situ padlock probing was performed in two different settings: cell seeded Superfrost Plus slides (Menzel/Thermo Fisher Scientific, Braunschweig, Germany) and cellular suspension in 1,5 ml eppendorf tubes. Oligonucleotides that served as padlock probes had been already designed using CLC Main Workbench software (CLC Bio Workbench Version 7.6, Qiagen, Venlo, The Netherlands) according to the guidelines published by Weibrecht et al (2). The protocol followed for cell seeded slides was an adaptation of the one described by Weibrecht and her colleagues (2).

In situ padlock probe experiments targeting β-actin and AR-FL mRNA transcripts were performed in parallel on both supports to keep assay conditions as similar as possible. By doing that, the differences between assays on the same cell line would be due to the distinct physical setting. As it was applied on two distinct supports, protocol employed on cell suspensions slightly differs to the one used on cell

seeded slides. Unless it is stated so, the following method has been applied to cell seeded slides and tubes containing cell suspension alike.

Slides: Secure-seals hybridization chambers (Sigma) are used in order to keep the added reagents on specific areas of the slide and to assure that the concentration is the intended. Secure seals have a certain volume, in this case, 35 µl so all the volumes used for slide are 35 µl. Any liquid can be added and removed with a micropipette as if they were usual tubes. Every liquid addition includes removing it after the specified time. As cells are adhered to the slide surface, a centrifugation step is not needed to remove the liquid.

Tubes: As they contain cell suspension, every volume addition inherently includes a centrifugation step (300 *g*, 5 min, RT) and the removal of its supernatant so that only the cell pellet continues to the following step. The addition of master mix is an exception: cell solutions would be incubated or placed in the thermocycler first.

Pre-treatment

Slides containing attached cells were taken from -80ºC refrigerator. 35 µl secure seals were mounted on the slides; then, cells were rehydrated with DEPC-PBS-Tween at room temperature for 6 min, after that, 0.1M HCl-DEPC-H_2O is added for 5 min to permeabilise cells. At last, two washes were made with DEPC-PBS-Tween for 6 min each in order to remove HCl. Then, the slides were ready to initiate reverse transcription.

The tubes containing cell suspensions were taken from -20ºC to room temperature. Once they reached room temperature, 400 µl DEPC-PBS-Tween were added to remove glycerol. Then they were resuspended in 200 µl DEPC-PBS-Tween and transferred to PCR tubes. 50 µl 0.1M HCl-DEPC-H_2O was added to permeabilise cells and is removed after 5 minutes. At last, two washes with 400 µl DEPC-PBS-Tween remove HCl. Each of these washes implied 3 minutes waiting steps before centrifugation. After that, the pellets were ready to initiate reverse transcription.

Reverse Transcription

The composition of reverse transcription master mix is available on Annex II.

Reverse transcriptase copies the mRNA transcripts into cDNA using the locked nucleic acid primers as guides. Therefore, only the targets are reverse-transcripted.

First of all, the reverse transcription master mix was prepared and added to secure seals and PCR tubes containing cells. A secure seal chamber was loaded with master mix that lacks *TranscriptMe RT* to serve as negative control. Then, PCR plate seals were applied to each secure seal to prevent evaporation. Slides were placed in a humid chamber and incubated for 3 hours (h) at 45ºC while PCR tubes used a thermocycler which followed the same temperature pattern, 3 h at 45ºC.

After incubation, PCR tubes were centrifuged and master mix was removed from tubes and slides. Then tubes and slides were incubated for 15 min at RT with 35 µl (tubes) / 90 µl (slides) of 3% formaldehyde (Sigma) buffered in DEPC-PBS. 200 µl of PBS were added to the PCR tubes to dilute the formaldehyde and thus minimize the impact of the additional time the cells are in contact with it during centrifugation. After that, both tubes and slides were washed twice with DEPC-PBS-Tween. Although it is recommended to wait 2 min before removing the media from secure seal chambers so that formaldehyde is completely eliminated, this is not necessary in PCR tubes as centrifugation took 5 min.

In situ ligation

The composition of ligation master mix is available on Annex II.

RNase H is an unspecific endonuclease that cleaves RNA in a RNA/DNA heteroduplex. In this case, RNase H degrades mRNA leaving a single stranded cDNA chain that will join to the newly added padlock probes. Ampligase (Illumina) is a commercial thermostable DNA ligase that catalyses ligation of adjacent 3' and 5' ends in duplex DNA structures that are stable at high temperature. RNase H degrades mRNA during 37ºC incubation and Ampligase is inactive. Then the padlock probes join to their cDNA targets through their 5' and 3' ends. RNase H is inactivated at 45ºC and Ampligase activates. 5' and 3' ends of padlock probes are covalently bonded so the previously linear padlock probes become circular strands.

The padlock targeting AR-FL had a backbone sequence called Lin16, which allows its identification. Because of the same reason, the padlock targeting β-actin had a backbone sequence called Lin33. Formamide was used to increase stringency, *ergo*, the extent to which hybridization between nucleic acids with mismatched sequence can occur and thus lowering melting temperature of hybrids.

The ligation master mix was prepared and added to each secure seal and PCR tube. It is necessary to apply PCR plate seals to each secure seal chamber on the slides to prevent evaporation while being incubated. Slides were then placed in a humid chamber and were incubated for 30 min at 37ºC and 45 min at 45ºC. PCR tubes were placed into a thermocycler which followed the same temperature program.

Afterwards, PCR tubes were centrifuged and master mix was removed from slides and tubes. Then, they were washed once with saline sodium citrate-Tween for 5 min to remove formamide and twice with DEPC-PBS-Tween to eliminate any trace of the reagents. Saline sodium citrate was used to remove formamide because it disrupts hydrogen bonds and may affect amplification.

In situ rolling circle amplification

The composition of rolling circle amplification master mix is available on Annex II.

phi29 polymerase (Thermofisher Scientific) uses the 3' end of cDNA as primer and the circularised padlock probe as template to synthesize a strand complementary to the padlock probe via rolling circle amplification. The new strand contains several complementary copies of the probe, including the probe detection sequence, due to the strong strand displacement activity featured by the commercial polymerase (17). This amplifies the posterior signal, increasing the fluorescence intensity to an easily detectable threshold.

The rolling circle amplification master mix was prepared and added to the secure seal chambers and PCR tubes, which were not centrifuged afterwards. Secure seal chambers were sealed with PCR plate seals. After that, PCR tubes and slides were incubated in a humid chamber at room temperature overnight.

Detection Probe Hybridisation

The composition of detection master mix is available on Annex II.

Detection oligonucleotides join to the complementary detection sequences present in the rolling amplification product. Each of them was conjugated with a different fluorophore so every fluorescence was proportional to the specific original transcript concentration: Cy3 to β-actin and Cy5 to AR-FL.

Nuclei were stained with DAPI (4',6-diamidino-2-phenylindole), a fluorescent dye that binds unspecifically to A-T rich regions of DNA.

After incubation, PCR tubes were centrifuged and master mix was removed from all slides and tubes. Then, they were washed twice with DEPC-PBS-Tween. In the case of the slides, it was necessary to wait 2 min before removing each time the washing buffer.

Detection master mix was added to secure seal chambers and PCR tubes. Secure seal chambers were sealed with PCR plate seals. After that, PCR tubes and slides were incubated in a humid chamber at 37ºC for 30 minutes. Afterwards, PCR tubes were centrifuged and master mix was removed from all slides and tubes. Then, they were washed once with DEPC-PBS-Tween for 5 min and twice with DEPC-PBS-Tween. In the case of the slides, it was necessary to wait 2 min before removing each time the washing buffer.

Preparation for microscopy

Once they were stained with fluorescent dyes, preparations had to be protected from light.

Slides: Secure seal chambers were removed and slides were submerged in ethanol series (70%, 85% and 97%). Once they were dry, 50 µl of SlowFade Gold Antifade Reagent (Thermo Fisher Scientific) were added and a 24x55 mm coverslip was mounted on top.

PCR tubes: 10 µl of the cell solution were taken to a slide, where they were mixed with 30 µl of SlowFade Gold Antifade Reagent (Thermo Fisher Scientific) and covered with a 24x55 mm coverslip.

Micromanipulation

Single cells were selected via micromanipulation from an aliquot containing a suspension of VCaP cells which had undergone *in situ* padlock probing. Before that, the cells were forced to form a single layer on a slide using cyto-centrifugation:

First, using a PAP pen (Sigma-Aldrich) cell suspension was prevented from spilling, then a cytochamber was mounted on the slide and 1 ml of the cell solution was transferred to it. The slides were centrifuged at 150 *g* for 1 min. In this step, there was no removal of the supernatant because the purpose of cyto-centrifugation was not to separate cells from the medium. The protocol used for cyto-centrifugation was adapted from the one described by Kronesis and El-Heliebi (18) in their section "Live Cell Immunostaining".

Using a capillary needle connected to a vacuum system, single cells were taken from the cell suspension to a PBS reserve in order to wash any substance that may interfere with posterior steps. After that, cells were transferred to sterile PCR tubes containing 1 µl of PBS. The protocol used for micromanipulation was adapted from the one described by Kronesis and El-Heliebi (18).

Whole Genome Amplification

Sequencing needs a minimum amount of 5 ng to be effective; however, a diploid human cell is estimated to only provide 7 pg(11). Increasing the DNA quantity of the previously selected VCaP cells was mandatory; consequently whole genome amplification was performed. In order to do so, "Ampli1™ WGA" commercial kit was used. It follows an already published (19,20) adaptor-ligation-mediated whole genome amplification: DNA is digested by the restriction enzyme *Mse*I and amplified using a primer pair that binds to the generated restriction ends in a PCR reaction. The current one-day protocol that is now commercialised was not the one employed due to schedule incompatibilities. Instead, the previous three days kit was preferred. An equivalent protocol is described by Geigl and

Speicher (21) in their section “Whole Genome amplification by linker-adaptor PCR”. Whole genome amplification was performed inside a biological safety cabinet Safe 2020 (Thermofisher Scientific) to avoid contamination.

GoTaq PCR

GoTaq® Green Master Mix (Promega, Vienna, Austria), primer mixture (22), nuclease free water and an aliquot of whole genome amplified material were used in this reaction, which was carried out in a thermocycler. The composition of GoTaq® Green master mix is available on Annex II. The method is based on the one published by van Beers and his colleagues(23) and consists on a multiplex PCR that uses four primer pairs to amplify four specific DNA fragments found in GAPDH gene and downstream of it. The fragments have different molecular weights (100, 200, 300 and 400 bp) so that they are easily spotted on 1% agarose gel electrophoresis. Depending on the quality of amplified material, a different number of bands were seen: two bands denoted the minimum quality appropriated for further sequencing while three or four bands indicated that more accurate results might be expected; note that four bands were a prove of a better quality than three. A complete guide to the method, including the primers’ sequence and the exact concentrations of materials, is published elsewhere (22).

Microscopy

At the end of the *in situ* probing protocol, cells have been stained with three different fluorophores: DAPI, Cy3 and Cy5. Images were captured using the Zeiss Observer.Z1 inverted microscope (Carl Zeiss, Oberkochen, Germany) and the AxioVision software (Carl Zeiss, Version 4.8.2.0). The fluorescence of each compound was captured through a different channel and then Z-stacks were combined in a single layer as a maximum intensity projection with ZEN 2012 black software (Carl Zeiss, Version 8.1). Brightness and contrast of every image were adjusted for better visualization with ZEN 2012 black software (Carl Zeiss). Each dot suspected to be positive was verified by checking its position through all other fluorescent channels, as false positive signals are typically visible in multiple wavelengths.

Quantification

Images were analysed using Cell Profiler Version 2016-05-03T18:31:00 ac0529e (Carpenter Lab, Broad Institute of Harvard and MIT), is an open source software that allows recognising several objects and areas in an image in order to identify and quantify the objects on it. In this case, fluorescent dots were quantified. Following the designed pipeline, Cell Profiler used nuclei, which appeared as large areas stained with DAPI, as starting points from which cytoplasm area is calculated. The fluorescent dots found within that area were then counted and assigned as AR-FL and β-actin transcripts depending on which channel its fluorescence was detected on. After quantification, Cell Profiler generated an image, showing what the calculated cytoplasms were, how many dots were assigned to each cell and how many dots were not assigned to any cell. Results were also transferred to a Microsoft Excel file.

Each fluorescent microscopy image was a composition of three different layers, one for each fluorescent channel used: one corresponds to DAPI channel, other to Cy3 and the last one corresponds to Cy5 channel. These raw information files were what Cell Profiler used as input and therefore Cell Profiler activity was not affected by any of the image editing mentioned on the “Microscopy” section. By analysing raw microscopy data files, Cell Profiler can detect dots that may not be visible on the edited image and differentiate individual dots that may merge and appear as one single dot on an image.

However, Cell Profiler is prone to errors in certain circumstances, such as when cells are clustered and several nuclei are found next to each other. Even though it has tools that help discriminating these scenarios, a manual revision was needed. This consisted on comparing fluorescent microscopy images to their computer-generated homologues, where it was possible to trace every count the software has made. To manually correct the output images and quantification, the following discrimination rules were set:

- First, when several nuclei were considered a single cell, their measurements were excluded. Exceptions were made when the nuclei were separated enough to trace a line between them and dots were located in one of the areas that that line separates. Therefore, the dots belonging to each nucleus would be manually counted and added to quantification as new sets of dots assigned to different cells.
- Second, nuclei that were overlapped with others were excluded from quantification, as it was not possible to elucidate what cell each dot belonged to.

After correcting Cell Profiler quantifications of all the images obtained from the same sample, for example, HT-29 cells attached to slide which underwent *in situ* padlock probing on 09/08/2016; average and standard deviation were calculated from the number of detected dots/cell.

Results

Cell culture and fixation in formaldehyde

The three cell lines were cultured. Cells were split when the culture reached confluence, which was confirmed by microscopy. Splitting was imperative each three to five days. No sudden colour change was observed in culture media, which contains pH indicators, and no existent contamination was revealed by the posterior DAPI staining during fluorescent microscopy. None of the cells lines showed other signs of contamination throughout the duration of the experiments.

In situ padlock probing generally took place the day following fixation (*in situ* probing on 09/08/2016) or several days after it (*in situ* probing on 29/08/2016, suspension cells from 18 and 22/08/2016), storage in glycerol did not affect *in situ* padlock probing efficiency.

Optimization of *in situ* padlock probing assay to cellular suspensions

In situ padlock probing protocol for cell suspension was adapted from the one optimized for cells seeded onto slides. *In situ* probing was performed on cell suspension and cells attached onto slide simultaneously so that their performances could be compared. Adaptation of the protocol included:

Centrifugation steps: The addition of any volume of reagent only involved its addition via micropipette, a waiting time and its removal via micropipette when working with slides and secure chamber seals. However, when working with cell suspensions this step involved centrifugation and removal of the supernatant. Centrifugation ought to be as short as possible in order not to hinder the advance of the protocol on slide. Shorter centrifugation times imply stronger centrifuge force to precipitate all cells in solution, but if an excessive force is applied, cells might break. After several attempts, the centrifuge force used throughout the protocol was fixed at 300 *g* and centrifugation time, to 5 min.

Reduction of waiting steps: Some waiting steps that are necessary for *in situ* assays on slides can be omitted or shortened for cell suspension due to the fact that each centrifugation step adds 5 additional minutes of contact between cells and liquid. However, centrifugation is essentially a separation technique. As time passes; lesser number of cells is found within the intended reagent. In comparison to *in situ* protocol for seeded slides, the protocol for cellular suspension eliminates all waiting steps lesser or equal to 2 min and reduces all 6 min waiting steps to 3 min plus 5 min centrifugation.

Taking into consideration both factors, the total time added by centrifugation throughout a complete *in situ* padlock probing protocol amounts to 61 extra min.

Volumes of reagents and washings: Secure seal chambers only allow a specific volume to be added, in example, 35 µl in 30 µl secure seal chambers prevents formation of bubbles inside the chamber because it surpasses the maximum capacity. When working with cells in suspension, volumes vary during the protocol and tend to be larger than those used in secure seal chambers. For example, one washing step when using 1.5 ml eppendorf tubes equals 400 µl; nevertheless, one washing step when using PCR tubes equals to 200 µl. Significant changes were made to the step in which formaldehyde was added to cellular suspensions: while secure seal chambers allow its quick removal, the centrifugation step needed to remove it from cell suspension increased the time of contact with the reagent. This time should be meticulously followed to avoid damaging the cells. In consequence, it was diluted to a third of its concentration after the stipulated time had passed. Master mix volumes were not changed and remained as 35 µl per sample.

In situ padlock probing assays

Several successful *in situ* padlock probe experiments were performed. The set of VCaP which had undergone *in situ* padlock probing on 29/07/2016 were the source of the cells selected by micromanipulation to proceed to whole genome amplification. Two experiments were quantified after the probing assay. One on 09/08/2016 and the other, on 29/08/2016. All data and images, except for the part involving the above mentioned VCaP aliquot, refer to those

Fluorescence Microscopy

After *in situ* padlock probing on 09/08/2016, 60 µl of Slowfade Gold Antifade Reagent (Thermo Fisher Scientific) were used on slides with attached cells, which was excessive and resulted in the coverslip floating on top of the slides instead of adhering to them. Regarding cell suspensions, 20 µl of cellular solution were transferred to slides, where they were mixed with 60 µl of Slowfade Gold Antifade Reagent, which proved to be an excessive volume, as some of the suspension was spilt. This not only resulted in a dilution larger than intended but also in the loss of cells within the spilt liquid.

After *in situ* padlock probing on 29/08/2016, 50 µl of Slowfade Gold Antifade Reagent (Thermo Fisher Scientific) were used on cells attached onto slides whereas 30 µl were used on cell suspensions.

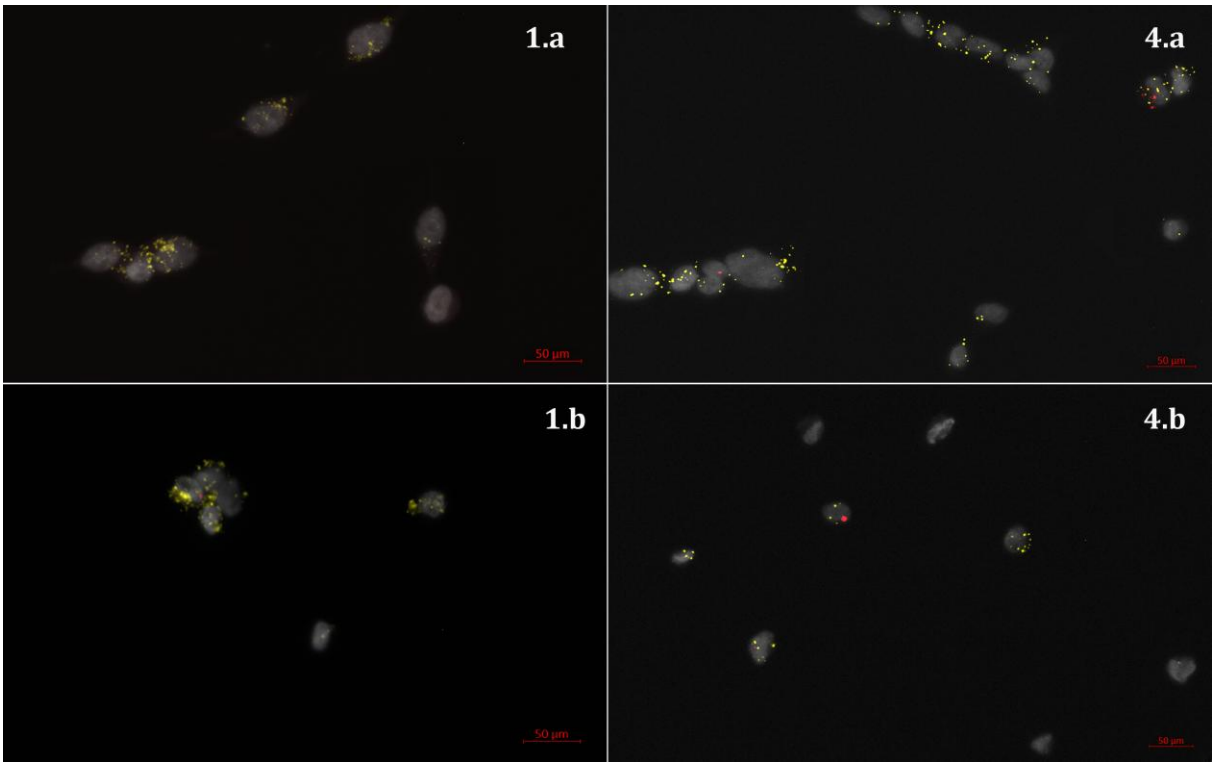

Figure 1: Image 1 (left) and *Image 4* (right). Images in higher magnification as well as images of the other cell lines are available on Annex I, Set 2.

Image 1: LNCaP from the *in situ* padlock probing assay performed on 9th of August 2016. 1.a: LNCaP seeded on slide. 1.b: LNCaP in suspension. Nuclei stained by DAPI are displayed in gray, β-actin expression dots in yellow and AR-FL in red. The variability in dots detected is remarkable in 1.a while clustered cells are more notable in 1.b. Despite that, both occurrences were common in all images.

Image 4: LNCaP from the *in situ* padlock probing assay performed on 29th of August 2016. 4.a: LNCaP seeded on slide. 4.b: LNCaP in suspension. Nuclei stained by DAPI are displayed in gray, β-actin expression dots in yellow and AR-FL in red.

In situ padlock probing on 09/08/2016 was successful on cellular suspensions for all cell lines and its results were quantified, as identifiable expression dots were detected and quantified. However, another probing assay was performed on 29/08/2016 due to: the variability revealed as standard deviation after processing and quantifying the images, problems found in assigning dots to individual cells from clustered groups, the already commented irregularities during preparation of the samples for microscopy and the fact that few AR-FL (androgen receptor full length) signal was detected for LNCaP seeded on slide. Images of samples from cell suspensions after *in situ* probing showed that cells were scarce and often clustered, which caused difficulties to quantification.

LNCaP, VCaP and HT-29 observed by fluorescence microscopy on 10/08/2016 had undergone *in situ* padlock probing on 09/08/2016. LNCaP and VCaP observed by fluorescent microscopy on 30/08/2016 had undergone *in situ* padlock probing on 29/08/2016. Images from those assays are available in Annex I, Set 1.

In situ padlock probing assay on 29/08/2016 provided samples of LNCaP and VCaP where most of the previously stated defects were corrected. HT-29 could not be tested. LNCaP cells showed AR-FL (androgen receptor full length) expression, the dots were more separated and clustered cells were avoided when taking pictures.

Quantification

All fluorescent images were analyzed by CellProfiler and manually corrected, after that, average fluorescent dots per cell and its standard deviation were calculated for each cell line (see Figure 3). As every image followed the same processing and manual correction, only one image analysis will be explained in detail because it shows all the cases that had to be corrected:

The cells displayed are HT-29 attached to slide after they underwent *in situ* padlock probing. It shows a fluorescent microscopy image, which is the composition of three different layers one for each fluorescence channel used. These layers are what Cell Profiler used as input. *Image d* represents the output image and some markers to make the explanation easier.

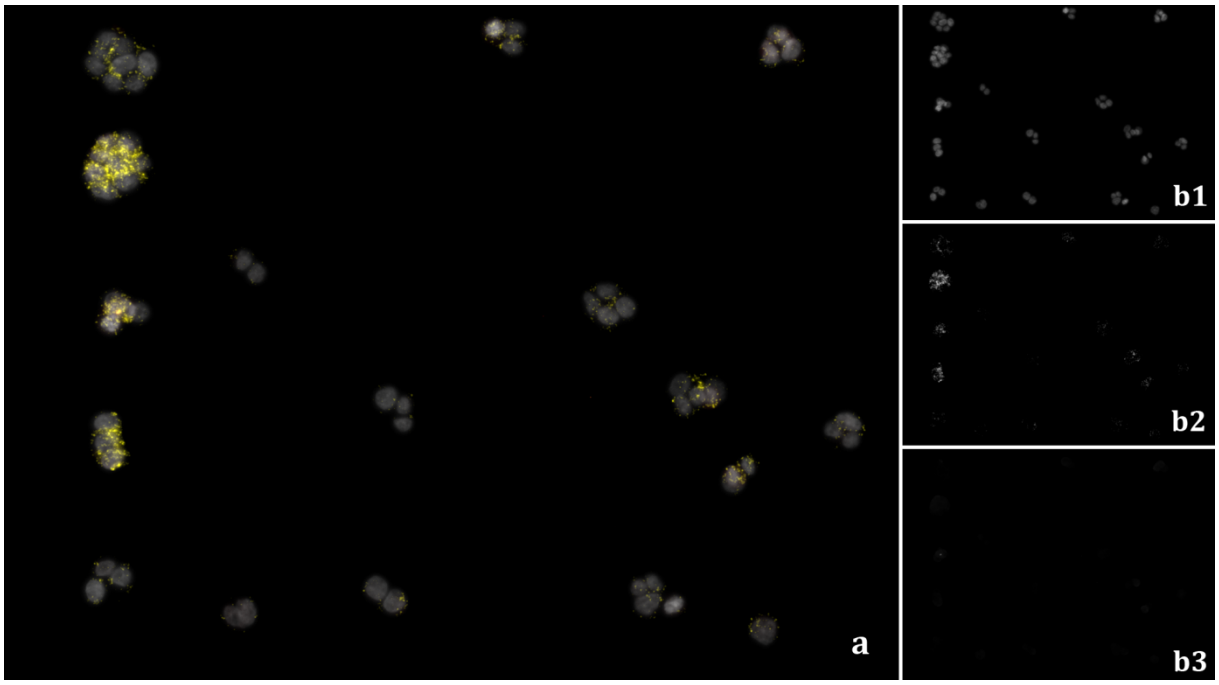

Figure 2: Image a, image b1, image b2 and *image b3*. *Image a*: Fluorescence microscopy image. It contains three different layers, one for each fluorophore. Nuclei (DAPI) in grey, β-actin expression (Cy3) in yellow and androgen receptor full length expression (Cy5) in red. *Image b1*: Fluorescence microscopy image. It contains only fluorescence detected through DAPI channel. *Image b2*: Fluorescence microscopy image. It contains only fluorescence detected through Cy3 channel. *Image b3*: Fluorescence microscopy image. It contains only fluorescence detected through Cy5 channel. Images in higher magnification on annex I, Set 2

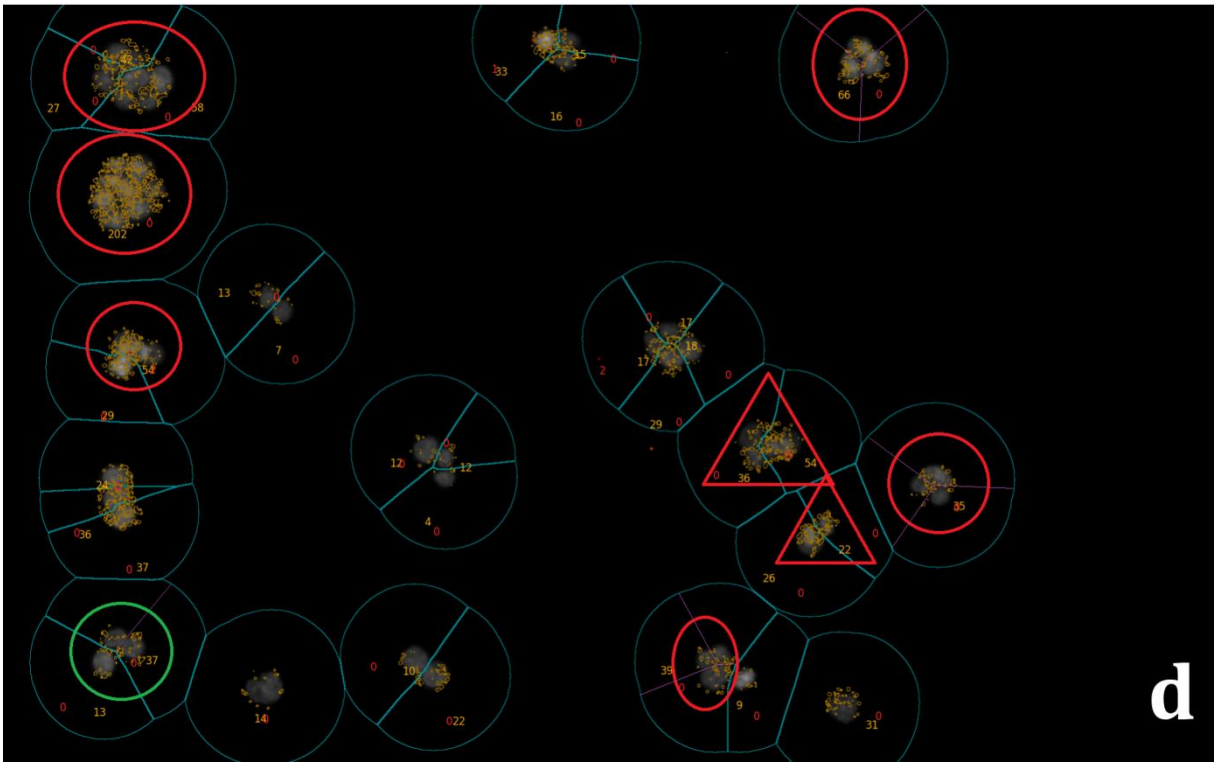

Figure 2: Image d: Output image showing the quantification results on it in addition to explanatory markers. Cytoplasm areas are delimited by aquamarine perimeters and surround what CellProfiler identified as a single nucleus. Yellow areas encircle what cell profiler identified as a single Cy3 fluorescent dot and red areas encircle what CellProfiler identified as a single Cy5 fluorescent dot. The number of dots counted in each cytoplasm is noted in yellow (Cy3) and red (Cy5) figures.

Explanatory markers: Thick circles mark the cases where the first rule for exclusion was applied: red circles surround cases where the measures were excluded from quantification, while green circles surround cases where the measures could not be excluded but were manually revised. Purple lines differentiate manually traced lines from those traced by CellProfiler.

Examples of the application of the first rule are surrounded by a circle. Some (upper left corner) cells are directly excluded as no line can be traced to separate the nuclei. Others are more difficult to exclude, a purple line is manually traced and the nuclei are separated. Most of them are excluded (red circles) because some dots are located in the line traced and not in one of the newly formed areas. One cannot be excluded (green circle). Therefore, the dots belonging to each nucleus are manually counted and added to quantification. Examples of the second rule are surrounded by a red triangle: their nuclei overlap so their measures are excluded from quantification.

β-actin quantification

LNCaP reached higher β-actin values on the first assay (10.08.2016 LNCaP) than on the latter assay (30.08.2016 LNCaP): 16.48 dots/cell for cell suspensions and 22.21 dots/cell for seeded slides compared to 9.80 and 11.11, respectively. VCaP had 14.06 for seeded slide and 6.72 for suspension on the first *in situ* experiment while the images of the latter one accounted only for 9.15 and 4.

Images from the only *in situ* experiment performed with HT-29 show 7.69 dots/cell in average for cell suspension and 23.73 for seeded slides.

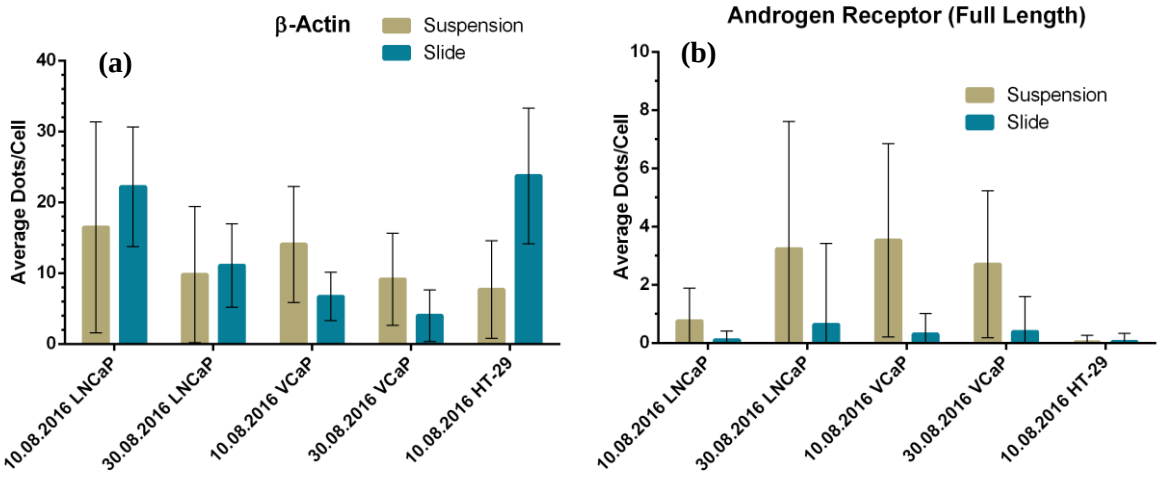

Figure 3. Average expression dots per cell after quantifying all images (a): average β-actin dots/cell of every cell line analysed, along with its standard deviation. (b): average androgen receptor full length dots/cell found in all cell lines analysed, along with its standard deviation.

Androgen Receptor Full Length quantification

LNCaP reached 0.76 dots/cell on average in cell suspension and 0.11 on seeded slide from the first assay performed. The cells from the last one had 3.24 dots/cell in cell suspension and 0.64 dots/cell seeded on slide. VCaP showed higher values: 3.53 dots/cell in suspension and 0.30 dots/cell seeded on slide on the first assay, while cells from the one performed on 30th August yielded 2.70 for suspension and 0.39 for slide. HT-29 had values of 0.03 for cellular suspensions and 0.05 for seeded slides despite the fact that this cell line does not express androgen receptor.

As for the standard deviation, it seemed to be low for seeded slides. Seeded LNCaP from the last experiment (30.08.2016 LNCaP) were an exception, showing higher standard deviation than the other slides.

LNCaP relative expression

For AR-FL, the average dots/cell detected on seeded slides was 13.65% of the average detected in cell suspension on the first *in situ* padlock probe experiment. Images from the last assay exhibited a percentage of 19.67%.

Regarding β-actin, its percentages were based on seeded slides because β-actin expression was larger in that setting. Images from the first *in situ* probing assay revealed that cells in suspension exhibited 74.2% of what seeded cells had, while in the last assay they exhibited 88.19% of the dots/cell its counterparts on slide showed.

VCaP relative expression

For AR-FL, on the first assay, seeded slides revealed 8.59% of what cell suspensions showed. In the case of the last one, this value was 14.38. For β-actin, in the first experiment, seeded slides reached 47.8% and on 30th August, seeded slides exhibited 43.7%.

HT-29 relative expression

A single *in situ* padlock probing assay was performed with HT-29. This is the relative efficacy (dots per cell in suspension/dots per cell on seeded slides):

Cells in suspension revealed 61.4% of the average AR-FL found in seeded slides. The average β-actin dots/cell found in suspension cells was 32.4% of the average from seeded slides.

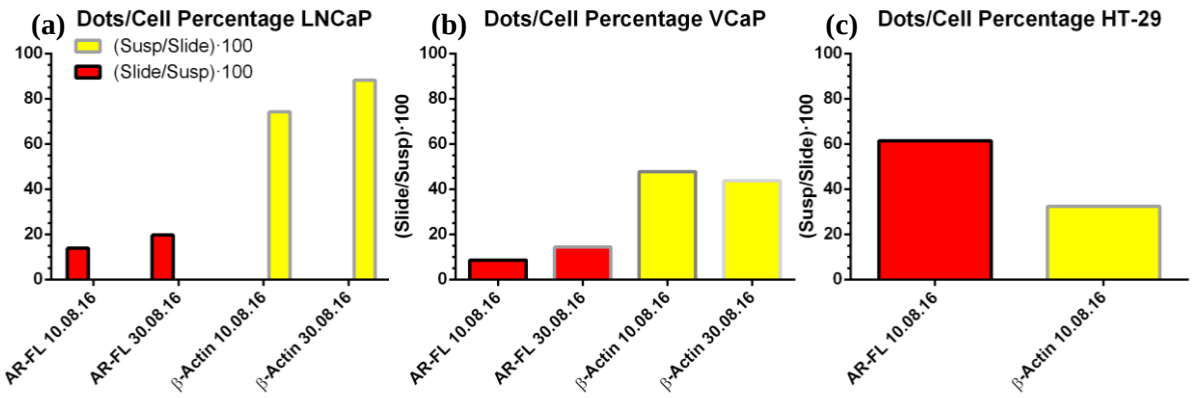

Figure 4. Relative efficacy of assays. (a): relative efficacy of all assays for LNCaP. Percentages are based on the setting that reached higher average dots per cell (b): relative efficacy of all assays for VCaP. Percentages are based on cell suspension average dots per cell. (c): relative efficacy of all assays for HT-29. Percentages are based on the average dots per cell reached by cells attached onto a slide.

Whole Genome Amplification and Quality control PCR

VCaP were selected by micromanipulation and transferred to 9 PCR tubes. The content of each tube was eventually analysed in an electrophoresis gel, so it will be referred directly as Lane X instead of Tube X.

Lane 1 contained PBS and was used as negative control. Lanes 2 to 5 contained a single cell. Lanes 6 to 8 contained pools of 15, 10 and 10 cells respectively. Lane 9 contained Human Genomic DNA: Female (Promega), which served as positive control. Several incidents occurred during whole genome amplification: the lid of the tube corresponding to Lane 1 broke but its content remained intact and the tube corresponding to Lane 2 opened during the last centrifugation step of the protocol, loosing half of its content. Control multiplex PCR and electrophoresis (Figure 1) of its products were conducted on 13 days after whole genome amplification. Until that moment, the product was stored at 4ºC.

If the quality of whole genome amplified DNA had been sufficient for being forwarded to sequencing, the respective QC-PCR product would have yielded three or four positive bands. The quality associated bands had four distinct molecular weights: 400 bp, 300 bp, 200 bp and 100 bp. Lane 9, which contained the positive control exhibited 4 bands with the expected weight so the whole genome amplification and Control PCR performed as intended. There were no bands in lane 1, as it would be expected for the negative control. The bands seen below 100 bp were the employed primers(22). No bands could be seen in either lane 2 or 5.

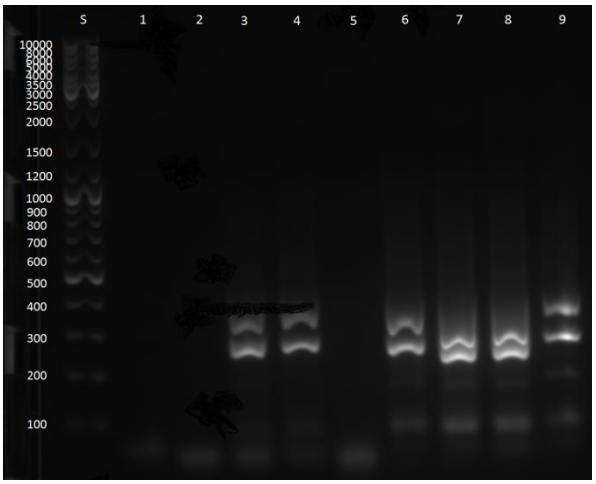

Figure 5: Gel Electrophoresis containing the product of quality control PCR (Image in higher magnification on Annex I, Set 1). The black lines had been probably painted on the outer case that contained the gel with a pen only visible under ultraviolet light. They could not be removed through digital editing without affecting the interpretation of the gel.

Lanes 3 and 4 contained 3 bands. One is slightly heavier than 300 bp and other, lighter. A faint band at 100 bp was also observed. Therefore, the amplified product had good quality.

Lanes 6, 7 and 8, which contained pools of several cells, showed 4 bands: one slightly lighter than 100 bp. A faint band slightly lighter than 200 bp,

another lighter than 300 bp and a last one which weights 300 bp in lane 6 and lighter in lanes 7 and 8.

"GoTaq® PCR" produced 4 bands that weight as specified. The positive and negative controls ensured that the technique worked as intended. There were trails following the bands. The bands nearer to specified weight of products (100 and 200 bp) were followed by faint trails. The lighter band near 300 bp was followed by a more noticeable trail while the heavier one near 300 bp was followed by the most intense trails.

Discussion

The absence of contamination in cell culture means that the detected signals should only belong to the specific cell line of each sample. Storage in glycerol did not affect the technique performance, as some samples were stored up to 11 days before undergoing *in situ* padlock probing.

Optimization of the protocol

In situ padlock probing was successfully performed in cell suspensions, as proven by the microscopy images and quantification. The protocol was developed from one that was optimized for seeded slides. Although the method was modified and it is suitable for cellular suspensions, it is not totally optimized. These are the changes that need to be implemented to the protocol:

First, its duration is excessive compared to the protocol for cell seeded slides. Apart from washing only once with DEPC-PBS-Tween after the washing step with saline sodium citrate-Tween in the ligation part of *in situ* padlock probing, no other steps may be omitted without risking the overall performance of the procedure. Other possible solution would be raising the centrifugal force so that the centrifugation time decreases. Being repeated 17 times during the protocol, any reduction in the centrifugation step duration will greatly affect its length. To discover its effects, I recommend performing a simple assay: counting the initial concentration of cells in suspension, centrifuging and removing the supernatant 17 times and counting cells at the end. A balance would be needed to achieve considering the time saved and the loss of cells that the researcher is willing to accept.

Second, having all the cells attached to a glass slide allows more trustworthy quantification because all cells are arranged in a single layer. When using cell suspension, it is common to find several cells in different layers at the same position, thus being impossible to quantify their spots. This handicap would be easily avoided if the cell solution was cyto-spun before microscopy using the same method that had been employed before micromanipulation. In particular, this additional cyto-centrifugation should be performed after the addition of the antifade reagent and before the coverslip is mounted on top of the slide. By doing so, the cells would be protected against sunlight and the risk of losing fluorescence during the centrifugation would decrease.

Throughout the assays, cell suspensions tended to rapidly deposit. This was especially perceptible when counting cells before storing them in glycerol. A delay of one minute before transferring cells from a newly made dilution to the Neubauer chamber greatly influenced the number of cells counted (data not shown). It would be advisable to perform this step as swift as possible in order to obtain an accurate cell number.

Feasibility of follow-up techniques

A number of accidents happened during whole genome amplification that partially explained the results of the latter gel electrophoresis. Despite its lid broke, tube corresponding to lane 1 was probably not contaminated because no bands are seen. Even if it was contaminated, the foreign organism did not possess sequences complementary to the primers used. Lane 9 contained positive

control and its 4 bands showed that the PCR was successful. The absence of bands in lane 2 was most likely due to the fact that the tube opened itself during centrifugation and most of its content was lost. Lane 5 does not show any band in spite of not being involved in any accident. It is highly probable that no cell was transferred to the PCR tube associated with lane 5 during micromanipulation. A single cell was sucked into the capillary needle and then released inside 1µl PBS droplet contained in the PCR tube. Due to the fact that the micromanipulator was located in the same room the fluorescence microscope, it was necessary to work in poor light conditions.

As stated in Results, lane 9, which contained the positive control, showed 4 bands with the expected weights (400 bp, 300 bp, 200 bp and 100 bp), therefore whole genome amplification and control PCR worked as expected. There were no bands in lane 1, which contained the negative control; therefore, the bands were not caused by an unknown contamination of PBS or of the cell medium. Interpretation of the remainder lanes is more difficult as the results were ambiguous.

The 100 bp bands in lanes 3, 4, 6, 7 and 8 might have the same weight as the one appearing in lane 9, however, the latter was too faint to confirm it. Bands slightly lighter than 200 bp appeared in lanes 6, 7 and 8 but they do not match the one found in positive control. There were bands at 300 bp in lane 6 that seemed to match the one found in lane 9. Lanes 3, 4 and 6 showed heavier bands and lanes 7 and 8 display lighter bands, all of which were considered as non-specific PCR products, along with the previously mentioned bands at 200 bp.

The guidelines for this method (22) state that non-specific PCR products must not be present and that at least two specific bands must be present in order to consider that the whole genome amplification product has enough quality to be sequenced. Therefore, only a small subset of samples would have yielded good quality DNA for follow up analysis, namely those of lanes 3 and 4 although to conclude whether or not the samples had the appropriate quality, another gel electrophoresis should be performed. The cause for the bizarre positions occupied by the bands is the heterogeneous agarose content of the gel. This can be seen in the curved shape of all bands, the trajectory of the DNA staircase and the fact that the positive control in lane 9 did not yield bands at the same heights that those of the DNA staircase. Defects in several components needed for electrophoresis, such as the DNA staircase or agarose caused the failure of two previous electrophoresis so it should not be discarded their involvement on these results. For the other samples, it might be speculated that *in situ* padlock probing protocol was the cause of the low quality due to the fact that the positive control DNA had suffered the same procedures since it was added after the selection of single cells. If the prolonged storage of amplified genome was the cause, lane 9 should have displayed the same problems. Also, "Ampli1™ WGA" has shown to commit errors (24) which might have also contributed to those results.

Comparison between cellular suspension and cells attached onto slides

Figure 3 (a) represents the average dots/cell detected for β-actin and its standard deviation. LNCaP and HT-29 showed more dots/cell in cells fixed on slide, while VCaP showed more dots/cell in cell suspension. Both LNCaP and VCaP cells had more dots/cell on images (captured on 10/08/2016) from the first *in situ* padlock probing assay than on images (captured on 30/08/2016) from the one performed on 29/08/2016. It is also worth mentioning that the difference between slide and suspension observed for LNCaP in 10/08/2016 images was notably larger than the one in 30/08/2016 images.

Figure 3 (b) represents the average detected dots/cell for AR-FL along with its standard deviation. LNCaP reached higher number of dots/cell in images from 30/08/2016 than in images from 10/08/2016. However, the average for VCaP suspension decreased and the one for VCaP attached onto slides remained unchanged. Average dots/cell detected in cells fixed on slides was similar. For cell

suspension, results were similar except in the case of LNCaP on 10/08/2016. HT-29 results were not compared because that cell line does not express androgen receptor. As for the standard deviation, it was higher in cell suspensions. LNCaP attached onto slides from 30/08/2016 also showed high standard deviation.

Figure 4 represents the average dots/cell fraction in order to show the difference between *in situ* padlock probing assays performed on cell suspension and on cells attached onto slides. Its denominator is the higher average reached. The closer to 100%, the smaller the difference between the two settings. Figure 3 (a) shows the percentages for LNCaP, where the decrease of difference for β-actin that was dispayed as standard deviation for Figure 3 (a) is noticed again.

These graphs showed that despite the average number of detected dots per cell changed from one assay to the next, their relative value did not. Although more assays should be conducted to confirm it, this could mean that the efficacy of *in situ* padlock probing is not only influenced by the setting it is performed on (namely, suspensions or attached to a slide), but also by the target and cell line involved in each particular assay. On that note, it should be remarked that the difference between cell seeded slides and cell suspension is not limited to the usage of a different physical setting but also comprehends the exposure of cells to drastically distinct environments, which may affect gene expression: the cells in suspension that were used for *in situ* padlock probing came from aliquots stored in glycerol, which in turn were directly obtained from cell culture; whereas cells that were seeded on slides had been stored at -80ºC before being used for *in situ* padlock probing.

The various differences between the results obtained from the same cell line in the same support can be attributed to the more adequate volumes of SlowFade Gold reagent utilised and the more cautious selection of images, avoiding clustered groups of cells. An adequate volume in the samples from cell suspension would have reduced the number of depth layers in which cells appeared. If two cells appeared on the same spot on different layers, they would have been mistaken for a single cell and the number of dots per cell would have increased. Avoiding clustered groups when taking images would have had the same effect. This can be observed in Figure 3 (a): the average number of dots per cell decreased in the last assay. The average number of AR-FL dots per cell in LNCaP (Figure 3 (b)), seems to oppose this supposition. However, the value from *10.08.2016 LNCaP* (see figure 3 (b)) was considered as abnormal when it was obtained and was one of the factors that signalled the necessity of repeating the assay on 29th August. Thus, it should not be considered when analysing the results.

As it can be seen in the standard deviation showed in graphs 1 and 2, the number of dots per cell varies greatly in cell suspension; in fact, cells without dots and cells with a higher number of dots than average were common occurrences (see images 1 and 4). The reason is most likely the different expression of transcripts in each individual cell, for example, if they were in a different cell cycle state.

Conclusions

English

In situ padlock probing assays can be performed using cell suspension instead of cell seeded slides and yield results that are observable through microscopy and susceptible of quantification. The same method for microscopy and quantification used for seeded slides can be applied to cell suspension.

Protocols for *in situ* padlock probing optimized for seeded slides can be adapted to cell suspension just by adding centrifugation steps. Nevertheless it is not as efficient. The protocol for cell suspensions that has been described in this thesis is still time consuming and needs further optimisation despite representing a significant advance. The main points that need to be changed are the following:

- Centrifugation times should be shortened by increasing centrifuge force. However, excessive force may break the cells so a separate assay must be performed to find the optimal time and force.
- Cyto-centrifugation should be performed before microscopy so that it would prevent cells to be found at different layers, which arises difficulties to quantify fluorescent dots.

After *in situ* padlock probing a sample of VCaP suspension, single cells were selected by micromanipulation and subjected to whole genome amplification. A multiplex PCR measured the quality of amplified material, indicating that some samples would be of good quality for sequencing, although gel electrophoresis of the control multiplex PCR should be repeated to safely conclude it.

The results obtained from probing assays on cells in suspension can be compared to those obtained on cells attached onto slides, although the relationship between the two is not clear. Efficacy may depend not only on which support it is performed on but also on the specific target and the cell line involved in the assay. If the cell line and target remain the same, a similar relation is found. However, even in that case, average number of dots/cell is different. This may be caused by the inherent differences to the use of each setting, cells attached to a slide being frozen at -80 ºC and cellular suspensions being kept in glycerol at 4 ºC. Hence, it would be advisable to check how the cell line, target and setting affect the assay results before comparing them with different experiments.

Castellano

Es posible realizar ensayos de expresión con sondas candado en células en suspensión y no solo en placas sembradas con células. Los resultados obtenidos pueden ser observados al microscopio y cuantificados usando los mismos métodos que los empleados con placas sembradas de células.

Los protocolos para el sondeo mediante sondas candado que ya están optimizados para placas sembradas con células pueden adaptarse para trabajar con células en suspensión simplemente añadiendo pasos de centrifugación. Sin embargo, esto no resulta tan eficiente. El protocolo de sondeo aquí descrito, a pesar de ser un avance significativo, es todavía costoso y debe ser mejorado. Los principales puntos que se deben modificar son los siguientes:

- Los tiempos de centrifugación deben ser acortados incrementando la fuerza centrífuga utilizada. No obstante, una fuerza excesiva podría romper células de modo que se debería realizar un ensayo aparte para averiguar la fuerza y el tiempo óptimos para el proceso.
- Se debería llevar a cabo una cito-centrifugación antes de observar las muestras al microscopio, ya que evitaría encontrar células a distintas profundidades, hecho que dificulta la correcta cuantificación de los resultados.

Las células individuales obtenidas tras el sondeo con sondas candado parece que podrían tener suficiente calidad para justificar una secuenciación posterior pero la electroforesis del producto de la PCR debería ser repetido para poder afirmarlo.

Los resultados obtenidos en el sondeo de suspensiones celulares pueden ser comparados a los obtenidos a partir de células adheridas a placas, aunque la forma en que se relacionan no está clara. La eficacia podría depender de otros factores además del soporte, como la diana y la línea celular utilizada en el ensayo. Para la misma diana y línea celular se encuentra una relación similar. Sin embargo, incluso en ese caso el número de puntos detectados por célula es diferente. Esto puede deberse a las características inherentes al uso de cada soporte, ya que las células adheridas a placas habían estado congeladas a -80ºC y las células en suspensión habían sido mantenidas en gliceron al 4ºC. Por ello, sería recomendable comprobar cómo afectan la línea celular, la diana y el soporte a los resultados de un ensayo antes de compararlo con otros experimentos.

Bibliography

1. Larsson C, Grundberg I, Söderberg O, Nilsson M. In situ detection and genotyping of individual mRNA molecules. Nat Methods. 2010 May;7(5):395–7.

2. Weibrecht I, Lundin E, Kiflemariam S, Mignardi M, Grundberg I, Larsson C, et al. In situ detection of individual mRNA molecules and protein complexes or post-translational modifications using padlock probes combined with the in situ proximity ligation assay. Nat Protoc. 2013 Jan 24;8(2):355–72.

3. Christian AT, Pattee MS, Attix CM, Reed BE, Sorensen KJ, Tucker JD. Detection of DNA point mutations and mRNA expression levels by rolling circle amplification in individual cells. Proc Natl Acad Sci. 2001;98(25):14238–14243.

4. Larsson C, Koch J, Nygren A, Janssen G, Raap AK, Landegren U, et al. In situ genotyping individual DNA molecules by target-primed rolling-circle amplification of padlock probes. Nat Methods. 2004 Dec;1(3):227–32.

5. Shaposhnikov S. Detection of Alu sequences and mtDNA in comets using padlock probes. Mutagenesis. 2006 May 30;21(4):243–7.

6. Wamsley HL, Barbet AF. In Situ Detection of Anaplasma spp. by DNA Target-Primed Rolling-Circle Amplification of a Padlock Probe and Intracellular Colocalization with Immunofluorescently Labeled Host Cell von Willebrand Factor. J Clin Microbiol. 2008 Jul 1;46(7):2314–9.

7. Andersson C, Henriksson S, Magnusson K-E, Nilsson M, Mirazimi A. In situ rolling circle amplification detection of Crimean Congo hemorrhagic fever virus (CCHFV) complementary and viral RNA. Virology. 2012 May;426(2):87–92.

8. Yaroslavsky AI, Smolina IV. Fluorescence Imaging of Single-Copy DNA Sequences within the Human Genome Using PNA-Directed Padlock Probe Assembly. Chem Biol. 2013 Mar;20(3):445–53.

9. Kiflemariam S, Mignardi M, Ali MA, Bergh A, Nilsson M, Sjöblom T. *In situ* sequencing identifies *TMPRSS2-ERG* fusion transcripts, somatic point mutations and gene expression levels in prostate cancers: In situ detection and identification of TMPRSS2-ERG fusion. J Pathol. 2014 Aug;n/a-n/a.

10. Siwetz M, Blaschitz A, El-Heliebi A, Hiden U, Desoye G, Huppertz B, et al. TNF-α alters the inflammatory secretion profile of human first trimester placenta. Lab Invest. 2016 Apr;96(4):428–38.

11. Macaulay IC, Voet T. Single Cell Genomics: Advances and Future Perspectives. Maizels N, editor. PLoS Genet. 2014 Jan 30;10(1):e1004126.

12. Voet T, Kumar P, Van Loo P, Cooke SL, Marshall J, Lin M-L, et al. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. Nucleic Acids Res. 2013 Jul 1;41(12):6119–38.

13. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013 Sep 26;45(10):1127–33.

14. Tang Y-C, Amon A. Gene Copy-Number Alterations: A Cost-Benefit Analysis. Cell. 2013 Jan;152(3):394–405.

15. Kader T, Goode DL, Wong SQ, Connaughton J, Rowley SM, Devereux L, et al. Copy number analysis by low coverage whole genome sequencing using ultra low-input DNA from formalin-fixed paraffin embedded tumor tissue. Genome Med [Internet]. 2016 Dec [cited 2017 Jun 27];8(1). Available from: http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-016-0375-z

16. Kamentsky L, Jones TR, Fraser A, Bray M-A, Logan DJ, Madden KL, et al. Improved structure, function and compatibility for CellProfiler: modular high-throughput image analysis software. Bioinformatics. 2011 Apr 15;27(8):1179–80.

17. phi29 DNA Polymerase (10 U/L) - Thermo Fisher Scientific [Internet]. [cited 2017 Jun 27]. Available from: https://www.thermofisher.com/order/catalog/product/EP0091

18. Kroneis T, El-Heliebi A. Whole Genome Amplification of Labeled Viable Single Cells Suited for Array-Comparative Genomic Hybridization. In: Kroneis T, editor. Whole Genome Amplification [Internet]. New York, NY: Springer New York; 2015 [cited 2017 Jun 22]. p. 233–43. Available from: http://link.springer.com/10.1007/978-1-4939-2990-0_16

19. Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Riethmüller G. Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. Proc Natl Acad Sci. 1999;96(8):4494–4499.

20. Polzer B, Medoro G, Pasch S, Fontana F, Zorzino L, Pestka A, et al. Molecular profiling of single circulating tumor cells with diagnostic intention. EMBO Mol Med. 2014 Nov 1;6(11):1371–86.

21. Geigl JB, Speicher MR. Single-cell isolation from cell suspensions and whole genome amplification from single cells to provide templates for CGH analysis. Nat Protoc. 2007 Dec;2(12):3173–84.

22. El-Heliebi A, Chen S, Kroneis T. Using Multiplex PCR for Assessing the Quality of Whole Genome Amplified DNA. In: Kroneis T, editor. Whole Genome Amplification [Internet]. New York, NY: Springer New York; 2015 [cited 2017 Mar 5]. p. 119–28. Available from: http://link.springer.com/10.1007/978-1-4939-2990-0_9

23. van Beers EH, Joosse SA, Ligtenberg MJ, Fles R, Hogervorst FBL, Verhoef S, et al. A multiplex PCR predictor for aCGH success of FFPE samples. Br J Cancer. 2006 Jan 30;94(2):333–7.

24. Deleye L, Tilleman L, Vander Plaetsen A-S, Cornelis S, Deforce D, Van Nieuwerburgh F. Performance of four modern whole genome amplification methods for copy number variant detection in single cells. Sci Rep [Internet]. 2017 Dec [cited 2017 Jun 27];7(1). Available from: http://www.nature.com/articles/s41598-017-03711-y

Annex I - Images

Set 1:

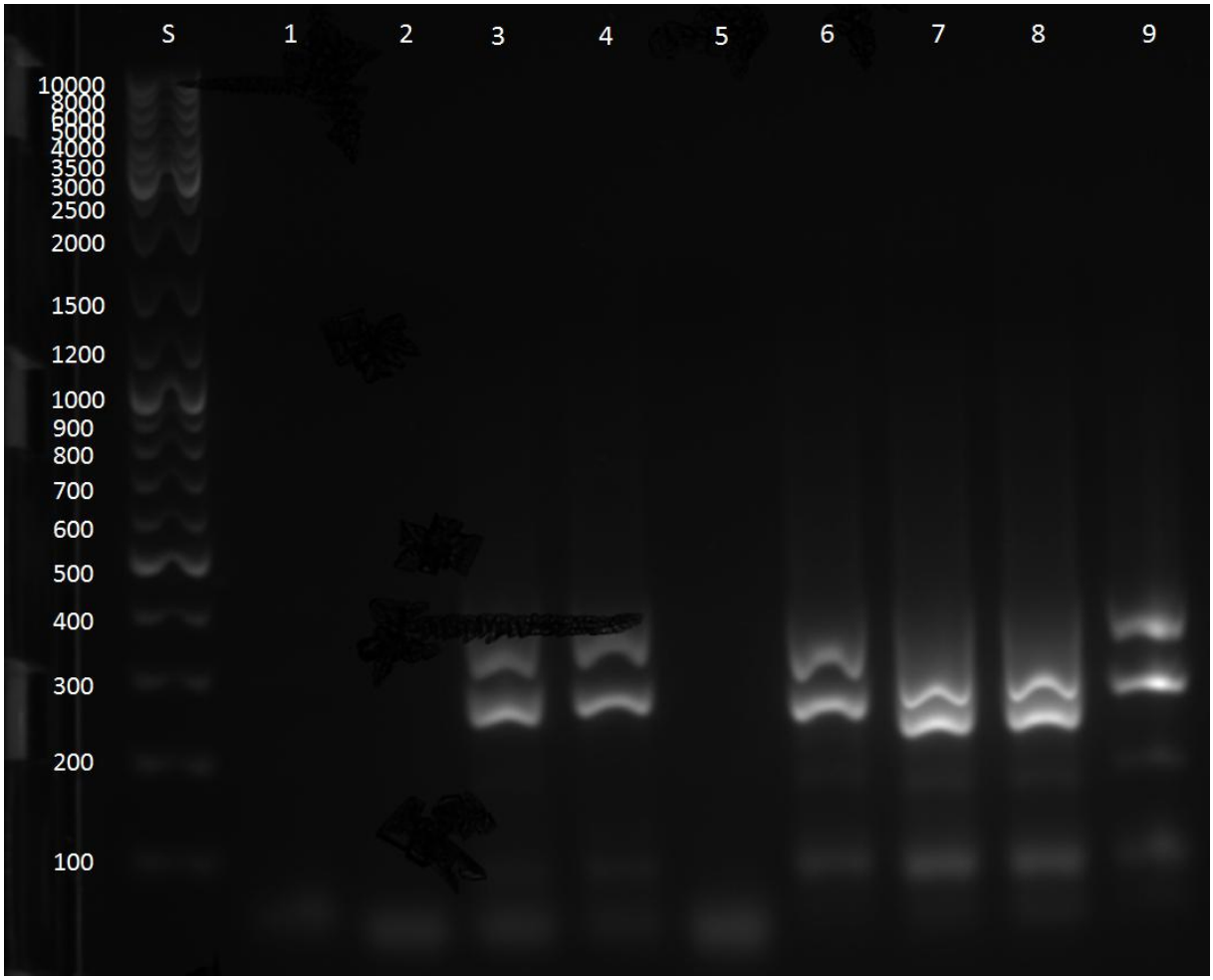

Image 1: Quality Control PCR of Whole Genome Amplification. Agarose gel electrophoresis image showing the product of Quality Control PCR of whole genome amplification products. Lane S shows a DNA staircase ranging from 10000 bp to 100 bp. Lanes 1 to 9 correspond with the content of PCR tubes from whole genome amplification, after undergoing GoTaq® Control PCR. Lane 1 contains negative control (solution without cells), lanes 2, 3, 4 and 5 contain single cells, lane 6 contains a pool of 15 cells, lane 7 contains a pool of 10 cells and lane 8 contains a pool of 10 cells. Lane 9 contains the positive control (Human Genomic DNA: Female (Promega)). The black lines had been painted with a pen only visible under ultraviolet light and could not be removed through digital editing without affecting the interpretation of the gel.

Set 2: Images from fluorescent microscopy. 09/08/2016 and 29/08/2016. Colours are digitally assigned depending of the channel the fluorescence was detected on: white to DAPI channel, yellow to Cy3 and red to Cy5. DAPI stains DNA, thus marking nuclei. Cy3 was conjugated to detection probes which should recognize β-actin transcript presence via circularised padlock probes. Cy5 was conjugated to detection probes which should do the same for Androgen Receptor Full Length (AR-201) transcript.

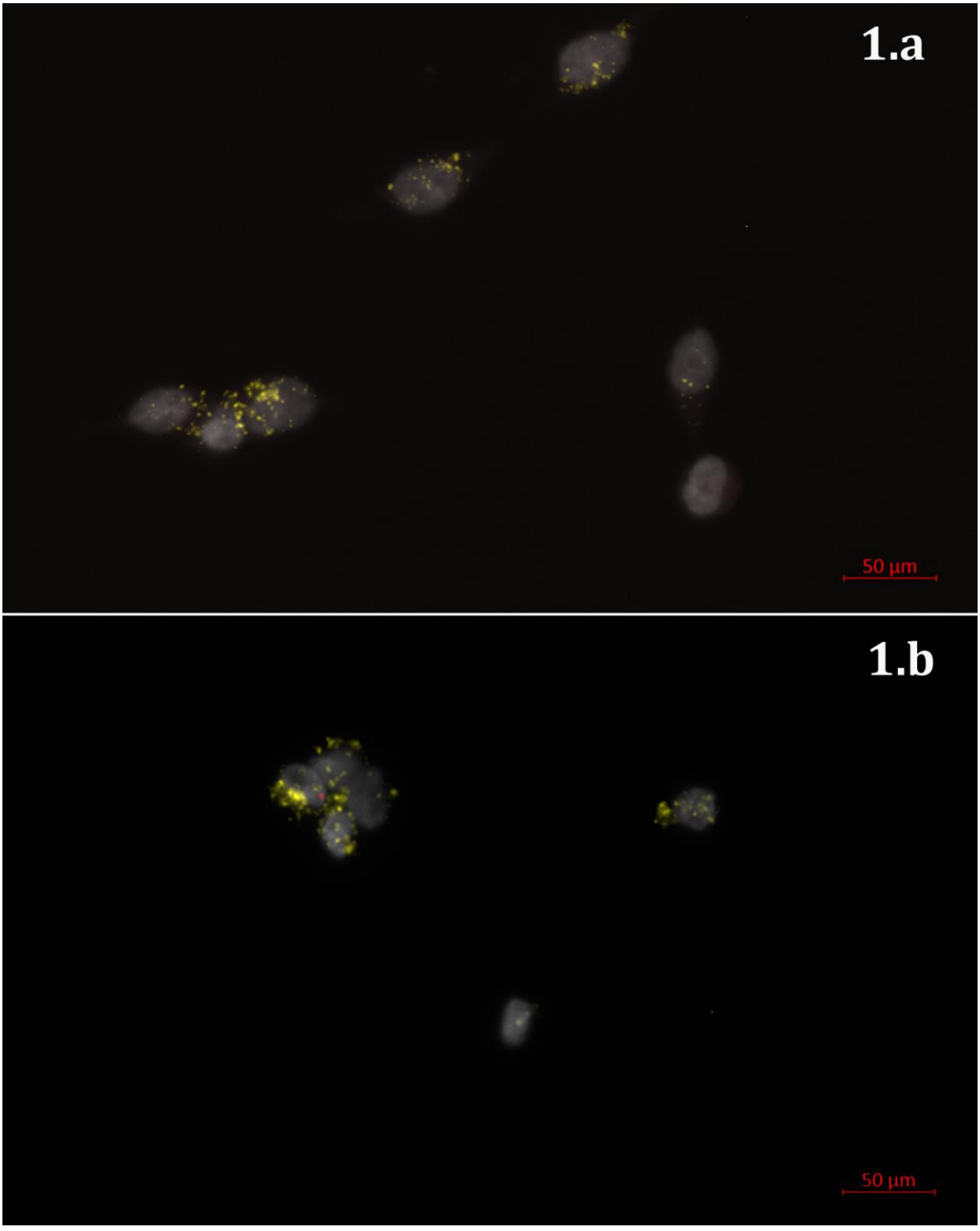

Image 1: LNCaP from in situ padlock probing assay 9th of August 2016. 1.a: LNCaP seeded on slide. 1.b: LNCaP in suspension.

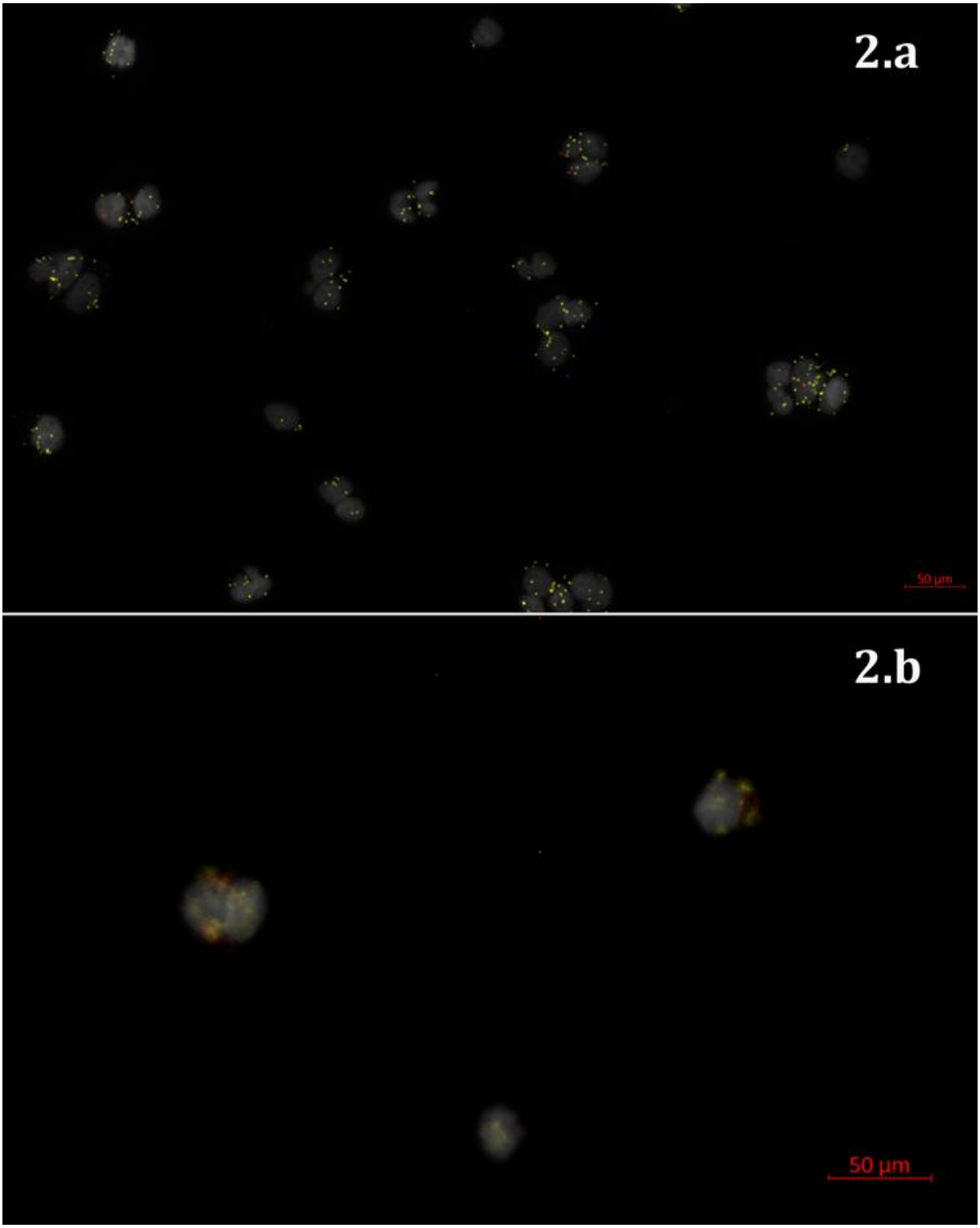

Image 2: VCaP from *in situ* padlock probing assay 9th of August 2016. 2.a: VCaP seeded on slide. 2.b: VCaP in suspension.

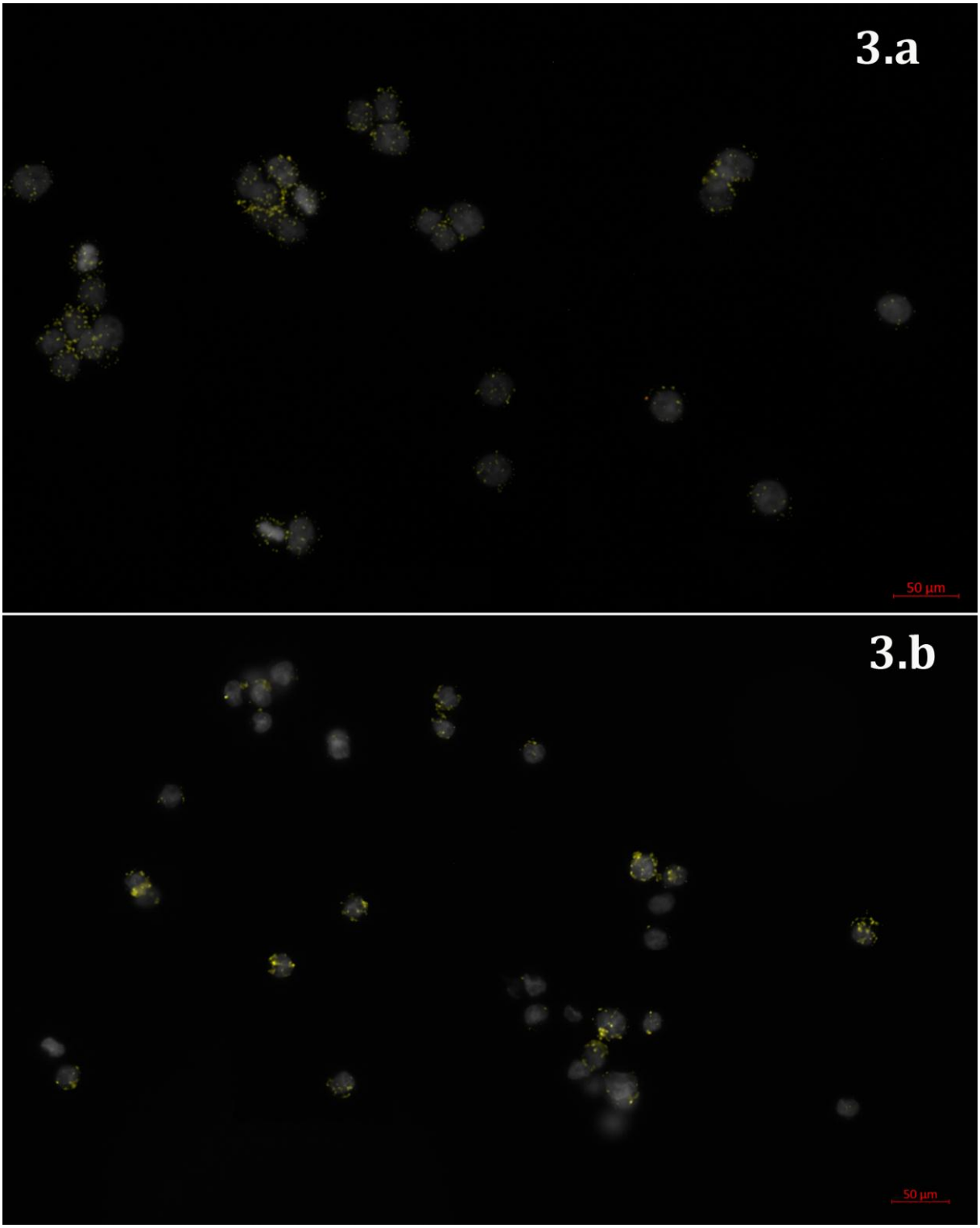

Image 3: HT-29 from *in situ* padlock probing assay 9th of August 2016. 2.a: HT-29 seeded on slide. 2.b: HT-29 in suspension.

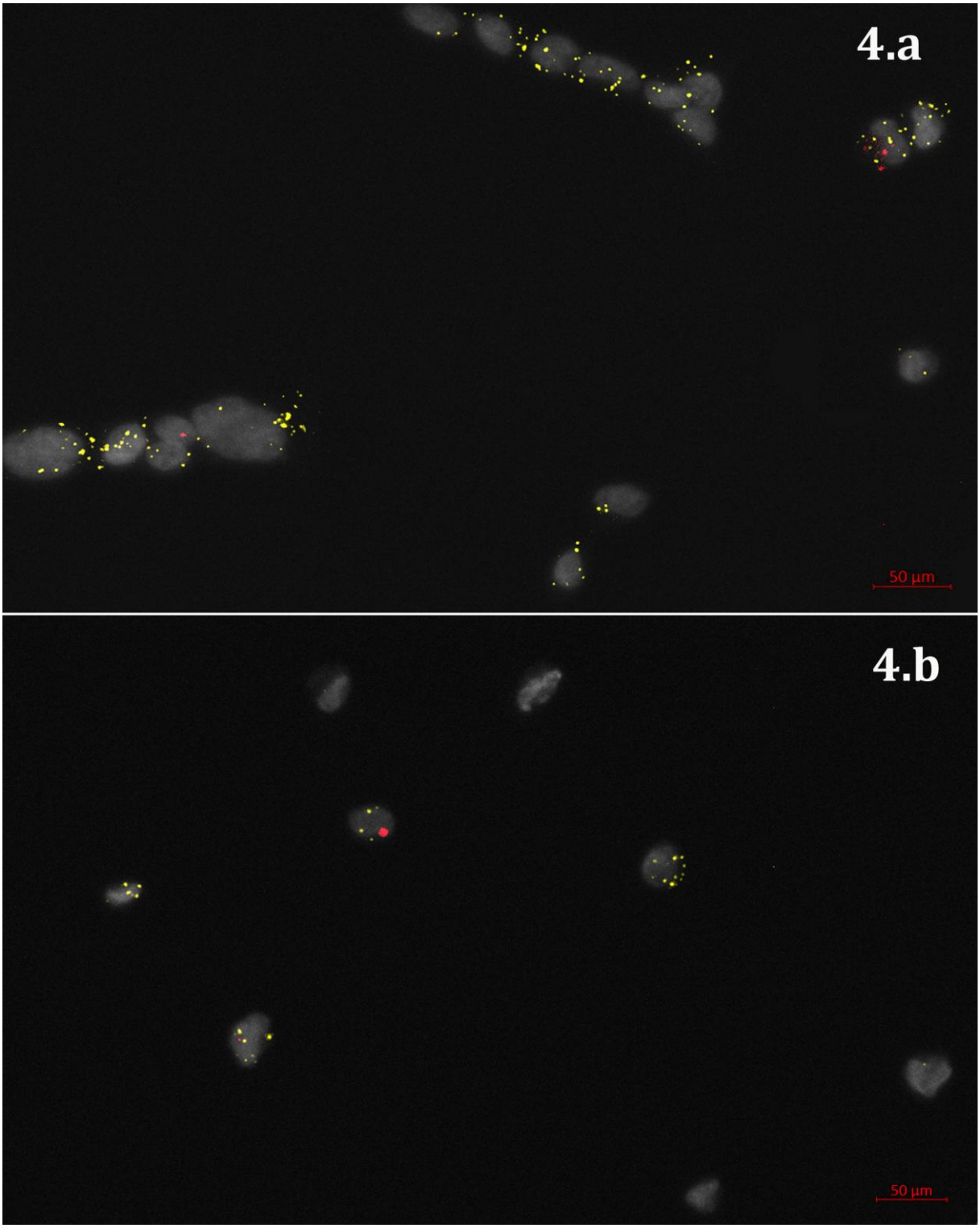

Image 4: LNCaP from *in situ* padlock probing assay 29th of August 2016. 4.a: LNCaP seeded on slide. 4.b: LNCaP in suspension.

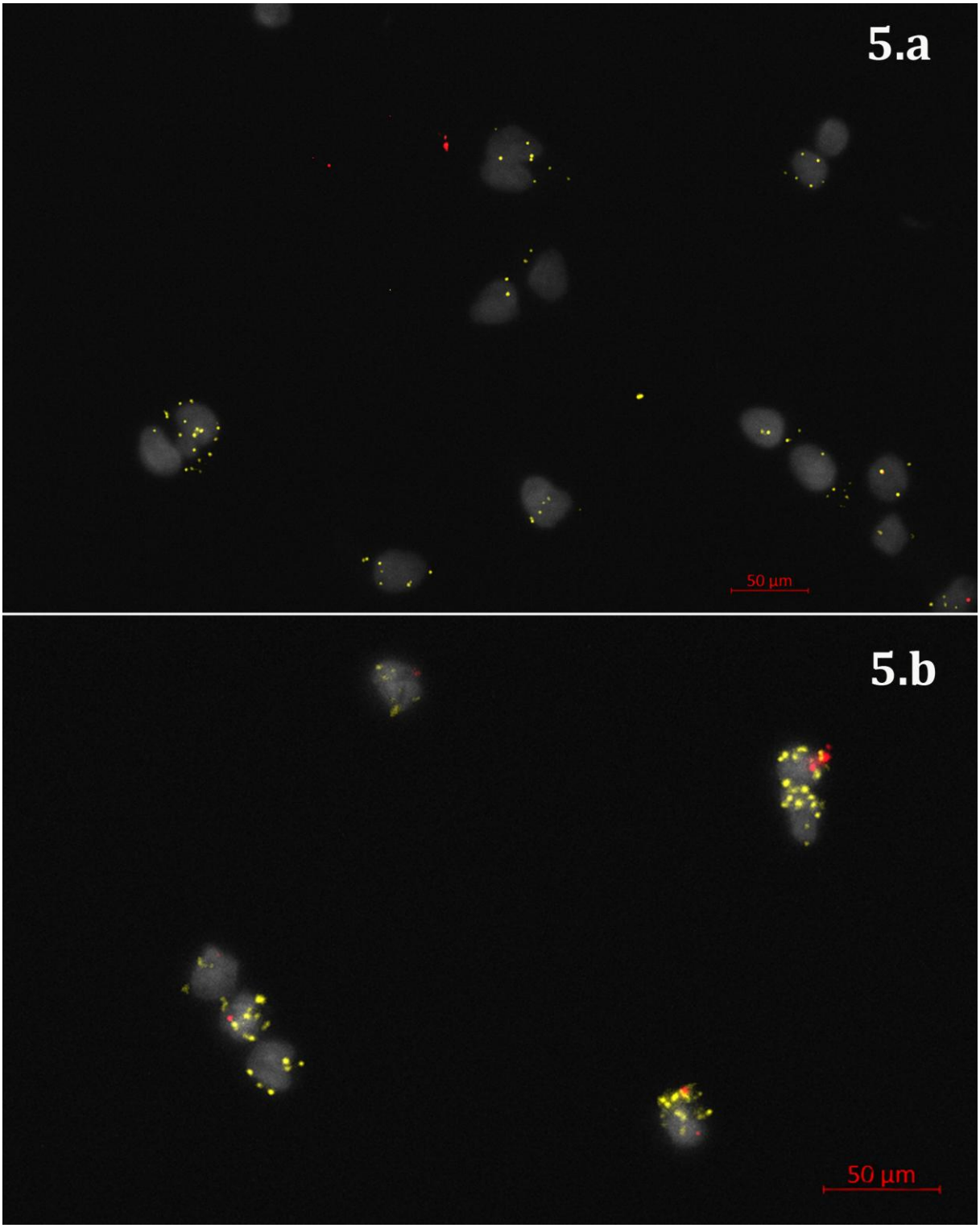

Image 5: VCaP from *in situ* padlock probing assay 29th of August 2016. 5.a: VCaP seeded on slide. 5.b: VCaP in suspension.

Set 3: Explanatory images for Cell Profiler Quantification and Exclusion rules, bigger size. HT-29 cell line which underwent *in situ* padlock probing seeded on a slide. The image was taken on 10th of August 2016.

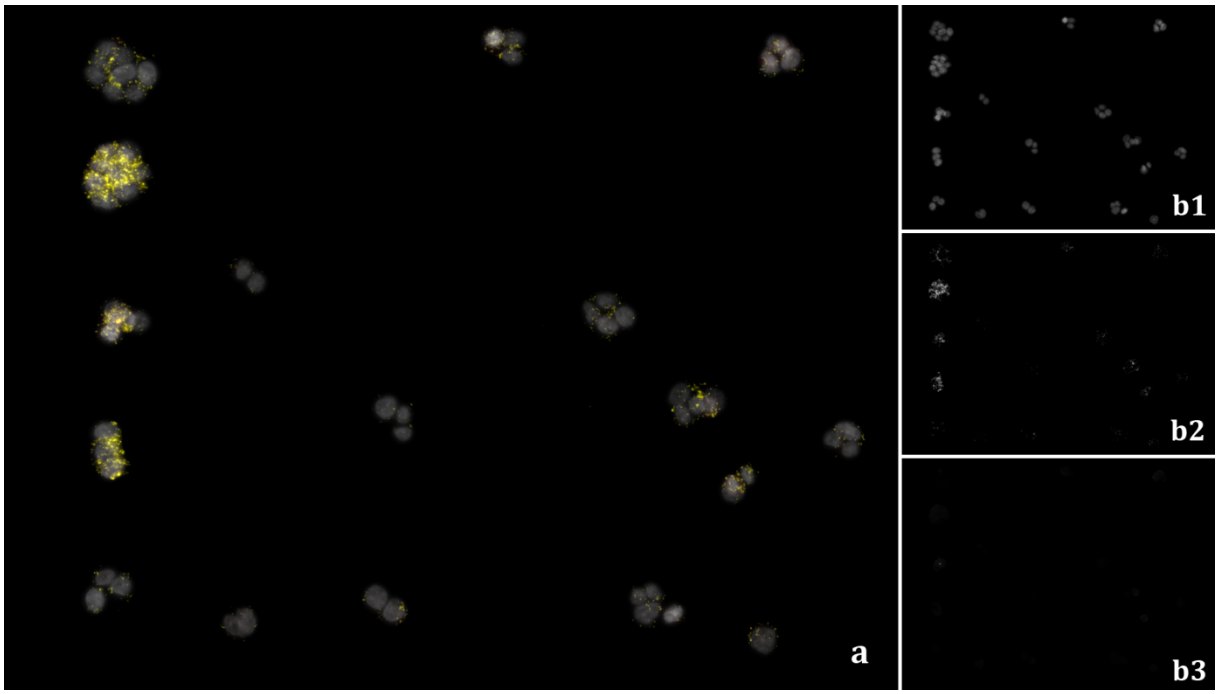

Image a: Fluorescence microscopy image. It contains three different layers, one for each fluorophore. Nuclei (DAPI) in grey, β-actin expression (Cy3) in yellow and androgen receptor full length expression (Cy5) in red. Image b1: Fluorescence microscopy image. It contains only fluorescence detected through DAPI channel. Image b2: Fluorescence microscopy image. It contains only fluorescence detected through Cy3 channel. Image b3: Fluorescence microscopy image. It contains only fluorescence detected through Cy5 channel.

Image c: Output image showing the quantification results on it. Cytoplasm areas are delimited by aquamarine perimeters and surround what CellProfiler identified as a single nucleus. Yellow areas encircle what cell profiler identified as a single Cy3 fluorescent dot and red areas encircle what CellProfiler identified as a single Cy5 fluorescent dot. The number of dots counted in each cytoplasm is noted in yellow (Cy3) and red (Cy5) figures. Image d: Image c with explanatory markers added. Thick circles mark the cases where the first rule for exclusion was applied: Red circles surround cases where the measures were excluded from quantification, while green circles surround cases where the measures could not be excluded but were manually revised. Purple lines differentiate manually traced lines from those traced by CellProfiler.

Annex II – Additional information

List of initials used throughout the thesis:

- DEPC: Diethyl pyrocarbonate. RNases are ubiquitous and must be inactivated to protect the mRNA transcripts that the assay is targeting. Diethyl pyrocarbonate inactives RNases.
- DEPC-PBS: Diethyl pyrocarbonate combined with phosphate-buffered saline, which maintains an isotonic osmolarity.
- DEPC-H_2O: Diethyl pyrocarbonate diluted in water.
- DEPC-PBS-Tween: DEPC-PBS with 0.05% Tween-20 (Sigma). Tween-20 is a detergent that solubilises hydrophobic substances.
- 0.1 M HCl-DEPC-H_2O: HCl (Sigma) diluted in diethyl pyrocarbonate and water to 0.1 mol of HCl per solution litre.
- SNP: Single Nucleotide Polymorphism
- PCR: Polymerase Chain Reaction
- WGA: Whole Genome Amplification
- DAPI: 4',6-diamidino-2-phenylindole

Cell culture medium components:

VCaP media consisted of DMEM high glucose (Gibco, Invitrogen, Paisley, UK) supplemented with 10% FCS (Fetal Calf Serum) and 1% P/S (Penicillin/Streptomycin). LNCaP media consisted of RPMI1640 (Gibco, Invitrogen, Paisley, UK) supplemented with 10% FCS (Fetal Calf Serum) and 1% P/S (Penicillin/Streptomycin). HT-29 media consisted of McCoy's Medium supplemented with FBS (Fetal Bovine Serum), HEPES (hydroxyethyl piperazineethanesulfonic acid), L-Glutamine and P/S (Penicillin/Streptomycin).

In situ reverse transcription master mix composition:

Reverse transcription master mix composition: 20 U/ µl TranscriptMe Retrotranscriptase (DNA-Gdansk, Gdansk, Poland), RT buffer (Gdansk), 1U/µl RiboLock RNase Inhibitor (Thermo Fisher Scientific, Waltham, MA, USA), 1 µM locked nucleic acid primer targeting AR-201 (Exiqon, Vedbaek, Denmark), 1 µM locked nucleic acid primer targeting β-actin (Exiqon), 0.5 mM dNTPs (Thermofisher Scientific), 0.4 µg/µl BSA and DEPC-H_2O.

In situ ligation master mix composition:

Ligation master mix: 0.5U/µl Ampligase (Epicentre, Illumina, Madison, WI, USA), Ampligase buffer (Epicentre, Illumina), 0.4 U/µl RNase H (Thermofisher Scientific), 0.4 µg/µl, BSA, 0,05 M KCl, 20% formamide (Sigma), 0.1 µM padlock probe targeting AR-201 cDNA (Integrated DNA Technologies, Coralville, IA, USA), 0,1 µM padlock probe targeting β-actin cDNA (Integrated DNA Technologies) and DEPC-H_2O.

In situ rolling circle amplification master mix composition:

Rolling Circle Amplification Master Mix: 1U/µl Φ29 polymerase (Thermofisher Scientific), Φ29 polymerase buffer (Thermofisher Scientific), 0.25 mM dNTP (Thermofisher Scientific), 0.4 µg/µl BSA, 5% glycerol (Sigma) and DEPC-H_2O.

Detection probe hybridization master mix composition:

Detection Master Mix: Hybridisation buffer, 0.1 µM detection oligonucleotide complementary to Lin16 conjugated with fluorophore Cy3 (Biomers, Ulm, Germany), 0.1 µM detection oligonucleotide complementary to Lin33 conjugated with fluorophore Cy5 (Biomers), 2 µg/ml DAPI (Thermofisher Scientific) and DEPC-H_2O.

GoTaq® Green master mix composition:

GoTaq® Green Master Mix is composed of: GoTaq® DNA Polymerase, Green GoTaq® Reaction Buffer, dNTPs and $MgCl_2$.