In situ molecular profiling of the microenvironment in breast carcinoma

Mustapha Kaira

Degree project in biology, Master of science (2 years), 2015
Examensarbete i biologi 45 hp till masterexamen, 2015
Biology Education Centre, Uppsala University, and Department of Biochemistry, Stockholm University
Supervisor: Jessica Svedlund
External opponent: Marco Mignardi
Abstract

High stromal PDGF receptor B expression was shown to have strong prognostic value in a study involving over 600 breast cancer patients however, the molecular role of the receptor in tumor development remains unclear. In this project we studied the spatial distribution and expression levels of a panel genes and markers associated with PDGF signaling, in breast cancer tumor microenvironment (TME) using a newly developed technique -in situ sequencing. The technique relies on padlock probes which we validated with corresponding RNA sequencing, microarray, and immunohistochemistry data. Our results showed that high PDGF receptor B mRNA co-localized with markers of two pathways, TGFβ and Hedgehog signaling; this suggests that they might contribute to the PDGF-receptor B-driven tumor growth. We also showed that stromal PDGF signaling is stimulated predominantly by tumor cells. Finally, further expression profiling of each individual gene revealed that CXCL14 was mainly expressed in the stroma, ACTA2 expression was enriched in the tumor/stroma boundary while the stem-cell marker, OCT3, was expressed in the interior of the tumor cells.
List of abbreviations

BSA  Bovine serum albumin

cDNA  Complementary deoxyribonucleic acid

DAPI  4',6-diamidino-2-phenylindole

DCIS  Ductal carcinoma in situ

DEPC-PBS  Diethylpyrocarbonate-treated phosphate-buffered saline

DEPC-PBS-T  Diethylpyrocarbonate-treated phosphate-buffered saline with 0.05% Tween-20

dNTPs  Deoxynucleoside triphosphates

ECM  Extracellular matrix

EDTA  Ethylenediaminetetraacetic acid

EGF  Epidermal growth factor

EMT  Epithelial mesenchymal transformation

IDT  Integrated DNA Technologies, Inc

IHC  Immunohistochemistry

MIP  Maximum-intensity projection

mRNA  Messenger ribonucleic acid

PDGF  Platelet-derived growth factor

PFA  Paraformaldehyde

RCA  Rolling circle amplification

RCP  Rolling circle product

RNA  Ribonucleic acid

TGFβ  Transforming growth factor beta

TME  Tumor microenvironment

UNG  Uracil-DNA glycosylase

VEGF  Vascular endothelial growth factor
# Table of Contents

Abstract .......................................................................................................................... 1

List of abbreviations ......................................................................................................... 2

1.0 Introduction .................................................................................................................. 5

1.1 Breast cancer .............................................................................................................. 5

1.2 The tumor microenvironment (TME) ....................................................................... 5

1.21 Fibroblast cells ........................................................................................................ 6

1.22 Immune cells .......................................................................................................... 6

1.23 Tumor vasculature ................................................................................................ 7

1.24 Extra cellular matrix (ECM) ................................................................................ 7

1.3 PDGF signaling in breast cancer ............................................................................. 8

1.4 Targeting the tumor microenvironment (TME) in breast cancer treatment .......... 9

1.5 In situ sequencing ..................................................................................................... 9

2.0 Objectives .................................................................................................................. 11

3.0 Materials and methods ............................................................................................. 11

3.1 Breast cancer tissue ............................................................................................... 11

3.2 Cell culturing .......................................................................................................... 11

3.3 Padlock probes used .............................................................................................. 12

3.4 In situ sequencing ................................................................................................... 13

3.41 In situ reverse transcription .................................................................................. 13

3.42 Rolling Circle Amplification (RCA) .................................................................... 13

3.43 Sequencing by ligation .......................................................................................... 14

3.44 Image acquisition ................................................................................................ 14

3.45 Stripping of sequencing oligos ............................................................................. 15

3.5 Image analysis ......................................................................................................... 15

4.0 Results and Discussion ............................................................................................. 16

4.1 Assay validation ...................................................................................................... 16

4.11 Correlation between in situ sequencing and mRNA levels .................................. 16

4.12 Correlation of in situ sequencing and protein levels ............................................ 18

4.2 Gene expression coupled to morphological features .............................................. 21

4.3 PDGF signaling ....................................................................................................... 22
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.31</td>
<td>PDGF Receptor Alpha</td>
<td>22</td>
</tr>
<tr>
<td>4.32</td>
<td>PDGF Receptor Beta</td>
<td>23</td>
</tr>
<tr>
<td>4.33</td>
<td>PDGF ligand expression</td>
<td>24</td>
</tr>
<tr>
<td>4.34</td>
<td>Expression profiles of tumors and the tumor microenvironment</td>
<td>25</td>
</tr>
<tr>
<td>5.0</td>
<td>Conclusion</td>
<td>27</td>
</tr>
<tr>
<td>6.0</td>
<td>Acknowledgement</td>
<td>27</td>
</tr>
<tr>
<td>7.0</td>
<td>References</td>
<td>28</td>
</tr>
</tbody>
</table>
1.0 Introduction

1.1 Breast cancer

Breast cancer has been reported to be one of the most dangerous forms of cancer, being the second most deadly form of cancer after lung cancer in women. An estimation of 300,000 new cases has been predicted for 2013 and the vast majority (78%) of those will be invasive. Of these cases, about 40,000 are expected to die from the disease for the USA alone (DeSantis, Ma, Bryan, & Jemal, 2014).

The forecast of the outcome (prognosis) depends a great deal on the tumor size, its presence or absence at a secondary sites and its proximity to a lymph node. Tumors that arise from cells that line ducts generally have worse prognosis than tumors that arise elsewhere. This distinction results in two main types of breast cancer: Ductal carcinoma in situ (DCIS) and Invasive ductal carcinoma (Sharma, Dave, Sanadya, Sharma, & Sharma, 2010).

DCIS are tumors that have not yet invaded or will not invade other sites. This form of breast cancer is contained within one site and in most cases does not require chemical treatment once surgically removed. However some invasive breast cancers originate from tumors that later penetrate the walls of the duct. Once this has happened, the cancer cells grow in the fatty tissue and may spread their ‘seeds’ to other parts of the breast (Sharma et al., 2010).

1.2 The tumor microenvironment (TME)

The mammary gland comprises of a portion of epithelial parenchyma emended in mostly stromal cells and extracellular matrix (ECM) (Hens & Wysolmerski, 2005). Its development undergoes a dynamic change through hormonal stimulation conducted through the stroma and the ECM. Like normal cells, breast tumor cells also depend a lot on their environment for nutrition,
immune cells, vasculature, signaling molecules and a matrix to grow. It has been repeatedly demonstrated that breast cancer develops in particularly dense stroma and rarely develops in less dense stroma (Walker, 2001) so the role of tumor microenvironment is very important in breast cancer progression (Fig. 1).

1.21 Fibroblast cells
Most of the cellular components of TME comprise of fibroblasts. Just like in wound healing, tumor cells secrete paracrine signals that cause neighboring fibroblasts to differentiate in pro-tumorogenic fibroblasts (Schäfer & Werner, 2008). These cells in turn synthesize pro-tumorogenic elastic fibres, collagen, cytokines glycoproteins, ECM modulating enzymes and other non-cellular components of the ECM. The arrangement of fibroblasts and ECM form a dense desmoplastic shield around the tumor against drug- and immune attack. A dense desmoplasia is usually seen at the invasive front of tumors and has long been established as a high risk factor in breast cancer patients.

Cancer associated fibroblasts also produce growth factors such as insulin-like growth factors, epidermal growth factors (EGF) and fibroblast growth factors which have all been found to be mitogenic for tumor cells. Cancer associated fibroblasts also synthesize transforming growth factor β (TGFβ) which induce epithelial mesenchymal transformation (EMT) (Dumont et al., 2013) a key step in metastasis.

1.22 Immune cells
The immune system eliminates foreign pathogens from our body but it also armed to eliminate potentially tumorogenic transformed cells. In the TME, however, such immune cells are rendered mute, providing a protective shield of pseudo immune cells around the tumor mass (Motz &
Expression of TGFβ from stromal fibroblasts has been implicated in rendering cells pro-tumorigenic (Bonde, Tischler, Kumar, Soltermann, & Schwendener, 2012). Hypoxic and inflammatory conditions also play a role in the attenuation of immune attack against tumors (Chouaib et al., 2012).

1.23 Tumor vasculature

Blood vessels in the TME are a hallmark of cancer (Hanahan & Weinberg, 2011) because tumors can grow only to a limited size without the need for vascularization (Aguirre-Ghiso, 2007). Blood vessel formation requires multiple TME cell types such as vascular endothelial cells, pericytes and bone marrow-derived precursor cells recruited by signaling molecules released in hypoxic conditions (Weis & Cheresh, 2011). The most prominent signaling molecules for neovascularization include vascular endothelial growth factors (VEGFs) (Carmeliet & Jain, 2011) and PDGFs (Spiller et al., 2014) commonly released by cancer stem cells, tumor associated macrophages and fibroblasts.

Lymphatic vessels contribute a great deal in suppressing host immune response (Swartz & Lund, 2012), about 60-80% of breast cancer have lymphocyte infiltrates (Aaltomaa et al., 1992)(Hartveit, 1998).

1.24 Extra cellular matrix (ECM)

The ECM comprises of the ground substance through which the interplay between tumor cells and stromal cells is conducted. It plays a vital role in breast cancer development (Arendt, Rudnick, Keller, & Kuperwasser, 2010) through its influence on tissue density. High tissue density in the breast also known as desmoplasia, is one of the strongest prognosis of breast
cancer, it confers 3-5 times higher risk factors to breast cancer (Wolfe, 1976)(Boyd et al., 2007). Increased stromal density is caused by increased ECM deposition by cancer associated fibroblasts (Weigelt & Bissell, 2008) driven primarily by PDGFs and TGFβ (Shao, Nguyen, & Barsky, 2000).

The ECM is not static; it can be selectively modulated by tumor cells to make way for tumor growth, angiogenesis and metastasis of the primary tumor. Stromal cells can also express various protein fibers, proteolytic enzymes and cross linkers to remodel it (Levental et al., 2009) (Erez, Truitt, Olson, Arron, & Hanahan, 2010).

1.3 PDGF signaling in breast cancer
An earlier study revealed that high stromal PDGF receptor B expression is strongly correlated to breast cancer development (Paulsson et al., 2009). Moreover, HER2 expression, proliferation rate, tumor size and other features associated with poor prognosis highly correlate to high stromal PDGF receptor B levels. PDGF signaling regulates cell growth and multiplication in many tumor associated cells including fibroblast, pericytes and smooth muscle cells (Betsholtz, 2004). It is involved in many aspects of tumor stroma development (Furuhashi et al., 2004)(Guo et al., 2003) and recruitment of tumor associated factors into the TME (Abramsson, Lindblom, & Betsholtz, 2003).

Anti-angiogenesis targets such as VEGF and PDGF inhibitors have been approved for many different types of cancers (Carmeliet & Jain, 2011). Anti-VEGF drugs are most effective against blood vessels with little or no pericyte because the collapse of endothelial cells leaves the PGDF-driven pericytes more or less unaffected. Combination of the two drugs against VEGF and PDGF results in a greater reduction in tumor vasculature compared with either drug individually
(Bergers, Song, Meyer-Morse, Bergsland, & Hanahan, 2003). In summary, there is a strong prognostic significance of PDGF receptor expression in breast cancer development and it is a potential contribution to cancer treatment (Paulsson, Ehnman, & Östman, 2014), however its molecular role in the transition is not fully understood.

1.4 Targeting the tumor microenvironment (TME) in breast cancer treatment

Our understanding of the supportive and protective role of the TME in tumor growth suggests a new avenue towards cancer treatment. However, to successfully target the tumor microenvironment for cancer diagnosis, prognosis and drug development, we need a better understanding of the inner workings of the TME. Therefore, in this study we will investigate the tumor microenvironment of invasive breast cancer tissues by using a newly developed technique called in situ sequencing (Ke et al., 2013). In the assay we used padlock probes targeting PDGF signaling, Notch signaling, EMT, TGFβ signaling and Hedgehog signaling pathways. In the panel of genes, fibroblast markers, stem cell markers and the housekeeping genes ACTB and GAPDH were included.

1.5 In situ sequencing

The technique is based on the use of padlock probes that when bound to its cognate target, are amplified by rolling circle amplification (RCA) to detect nucleic acids (figure 1). Padlock probes are short oligonucleotides probes that have arms complementary to their target and an embedded detection sequence which features a barcode unique for each target (Nilsson et al., 1994). Once the padlock probe binds to its target and its two arms are ligated, it is replicated multiple times into a long DNA polymer so called rolling circle amplification product (RCP). The target of the padlock probe is a Complementary Deoxyribonucleic Acid (cDNA) synthesized from the target a
messenger ribonucleic acid (mRNA) (Larsson, Grundberg, Söderberg, & Nilsson, 2010). Next is the binding of the anchor primer, right next to which a fluorescently labelled sequencing library is ligated — sequencing by ligation (Drmanac et al., 2010). Multiple cycles of ligation, imaging and stripping of the sequencing library is performed until the entire barcode is sequenced. Specificity of the technique is achieved by the selectivity of Ampligase to ligate only correctly base paired terminal nucleotides of the padlock arms. Finally the various sequences detected during imaging are analyzed with a CellProfiler pipeline and decoded using Matlab. Further details of the sequencing protocol has been published by Ke R. et al (Ke et al., 2013).

Figure 1: The in situ sequencing technology: 1) Binding of the reverse transcription primer. 2) mRNA is reverse transcribed into cDNA and then the mRNA template is RnaseH digested. 3) The padlock probe is hybridized to the cDNA, with its two arms facing each other. 4) Ligation of the two arms to form a closed circle. (5). The padlock probe sequence is then replicated multiple times into a long DNA polymer, the rolling circle amplification product (RCP). 6) Binding of the sequencing library to the barcode, right next to the anchor primer. 7) Image of cells with RCPs. (Image courtesy of Rongqin Ke - unpublished)
2.0 Objectives

In this project we will use the in situ sequencing technique to investigate the tumor microenvironment and the role of PDFR signaling in breast cancer. The specific aims are:

1. Validation of the in situ sequencing assay by comparison with mRNA expression levels (RNA sequencing and microarray data) and protein expression levels (Immunohistochemistry, IHC),

2. Correlate our data with morphological features of the tumor.

3. Investigate the possible role of PDGF signaling in tumor development.

4. Reveal the expression profile of tumors and their immediate microenvironment.

3.0 Materials and methods

3.1 Breast cancer tissue

Two fresh frozen invasive breast cancer tissue sections were used, called sample K and sample I. The samples were sectioned into 5µm sections and put on Superfrost Plus slides. The tissues were retrieved from Uppsala Biobank (Muggerud et al., 2010).

3.2 Cell culturing

Confluent cells were treated with 0.25% (w/v) trypsin-Ethylenediaminetetraacetic acid (EDTA) (Sigma) and then re-suspended in 5ml of culturing media. 3ml of the cells were then seeded on to 5 Superfrost Plus slides (Thermo) in a 150 mm × 25 mm Petri dish (Corning) containing 22ml of culturing media. This was then incubated overnight (12–24 h) at 37 °C, 5% CO2. The cell slides were washed 2 times with diethylpyrocarbonate-treated phosphate-buffered saline (DEPC-PBS) and then fixed in 3.7% (w/v) paraformaldehyde (PFA) (Sigma Aldrich) diluted in DEPC-PBS at room temperature for 30 min. The fixed slides were washed again in DEPC- PBS and then
dehydrated in an ethanol series of 70%, 85% and 100% (BDH Prolabo) for 5 minutes each. Finally the dried slides were stored at -20 °C for short-term storage (one week) or -80 °C for longer storage.

Fresh frozen tissues received from the storage (-80 °C) were fixed in 3.7% PFA for 45 minutes at room temperature and then washed 2 times in DEPC-PBS. Finally they were dehydrated in an ethanol serie of 70%, 85% and 100% for 5 minutes each.

### 3.3 Padlock probes used

Table 1: Summary of padlock probes

<table>
<thead>
<tr>
<th>Padlock probe target</th>
<th>Padlock probe target</th>
<th>Padlock probe target</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOTCH1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>ALDH1A1&lt;sup&gt;e&lt;/sup&gt;</td>
<td>CDH1&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>HEY1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>OCT4&lt;sup&gt;e&lt;/sup&gt;</td>
<td>CXCL14&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>HES1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>SOX2&lt;sup&gt;e&lt;/sup&gt;</td>
<td>GDF15&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>CD36&lt;sup&gt;b&lt;/sup&gt;</td>
<td>GLI&lt;sup&gt;c&lt;/sup&gt;</td>
<td>SNAIL&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Enos&lt;sup&gt;b&lt;/sup&gt;</td>
<td>PDGF receptor&lt;sup&gt;d&lt;/sup&gt;</td>
<td>TWIST&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>FAP&lt;sup&gt;b&lt;/sup&gt;</td>
<td>PDGF receptor&lt;sup&gt;d&lt;/sup&gt;</td>
<td>ZEB1&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>nNOS&lt;sup&gt;b&lt;/sup&gt;</td>
<td>PDGFA&lt;sup&gt;d&lt;/sup&gt;</td>
<td>CUX1&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>ACTA2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>PDGFB&lt;sup&gt;d&lt;/sup&gt;</td>
<td>ID1&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>S100A4&lt;sup&gt;b&lt;/sup&gt;</td>
<td>PDGFC&lt;sup&gt;d&lt;/sup&gt;</td>
<td>LIF&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>VIM&lt;sup&gt;b&lt;/sup&gt;</td>
<td>PDGFD&lt;sup&gt;d&lt;/sup&gt;</td>
<td>ACTB&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Notch signaling  <sup>b</sup>Fibroblast markers  <sup>c</sup>HH signaling  <sup>d</sup>PDG signaling  <sup>e</sup>Stem cell markers  <sup>f</sup>EMT  <sup>g</sup>TGFβ signaling  <sup>h</sup>Housekeeping genes
3.4  *In situ* sequencing

3.41  *In situ reverse transcription*

First we mounted the Secure-Seal hybridization chamber (Invitrogen) over the sample, on the slide. Through a port in this chamber, the reaction mixes were added.

The sample was then permeabilized with 0.1M HCl for 5 minutes at room temperature and then washed it with DEPC- PBS with 0.05% Tween-20 (Sigma) (DEPC-PBS-T), 2 times. The sample was then treated with the reverse transcription mix which contains 250 μM deoxynucleoside triphosphates (dNTPs) (Gdansk), 2 μg/μl bovine serum albumin (BSA) (New England Biolabs), 5 μM of random decamers, 1 U/μl RiboLock RNase Inhibitor (Gdansk), 20 U/μl of reverse transcriptase (Gdansk), in RT buffer (Gdansk), overnight at 45 °C. This was followed by washing with DEPC-PBS-T and a post fixation step with 3.7% PFA for 10 minutes in cells and for 30 minutes for tissue sections. Finally we washed twice again with DEPC-PBS-T.

3.42  *Rolling Circle Amplification (RCA)*

All 5’ phosphorylated padlock probes were hybridised to the cDNA synthesised above with a reaction mix containing 2 μg/μl BSA (New England Biolabs), 100 nM of each padlock probe, 50 mM KCl, 20% formamide, 0.5 U/μl Ampligase (Epicentre Biotechnologies) and 0.4 U/μl RNaseH in 1× Ampligase buffer (Epicentre Biotechnologies) The reaction was carried out at 37 °C for 30 minutes to degrade all ribonucleic acid (RNA) and then at 45 °C for 45 minutes to ligate the now hybridized padlocks. The sample was then washed twice with DEPC-PBS-T.

To roll the padlock, an amplification mix containing 0.25 mM dNTPs, 2 μg/μl BSA, 5% glycerol and 1 U/μl phi29 polymerase (Thermo Fisher Scientific) in 1× phi29 polymerase buffer. The
reaction was carried out overnight at room temperature for tissues and at 37 °C for 3 hours in cell lines. The slides were finally washed three times with DEPC-PBS-T.

3.43 Sequencing by ligation

We sequenced the barcode by first hybridizing the anchor primer for the sequencing library. The reaction mix contained hybridization buffer (2× Saline Sodium Citrate and 20% formamide) (Sigma Aldrich) and 0.5 μM of the corresponding anchor primer. The reaction was incubated at room temperature for 30 minutes and then washed twice with DEPC-PBS-T.

Next we hybridized and ligated the sequencing library with the following reaction mix: each of the three interrogation probes (0.1 μM for cell lines and 0.5 μM for tissue sections), 1× T4 ligase buffer (Gdansk), 1 mM ATP (New England Biolabs) 2 μg/μl BSA, 0.1 U/μl of T4 ligase (Gdansk), The reaction was incubated at room temperature for one hour which was followed by 3 times DEPC-PBS-T washing, each washing step incubated for 1 minute. We then stained the cells for 5 minutes with 1 μg/μl 4',6-diamidino-2-phenylindole (DAPI) (Biotium)

Next we scraped the back of the slides with a circle to mark the position were the experiment was carried out. An ethanol series of 70%, 85% and 100% for 3 minutes each was carried out next and then the hybridization chambers were peeled off. Finally we added 15 μl of mounting media (SlowFade®) over the sample and covered it with a cover slip.

3.44 Image acquisition

All images were acquired with an AxioplanII epifluorescence Zeiss microscope at 20X magnification. To capture all the RCPs, a stack of images were acquired at different focal plains and merged by maximum-intensity projection (MIP) into one image using the Zeiss AxioVision.
software. Multiple images were acquired until the entire field of view of interest was covered. These individual images were then stitched together to form one image for further analysis.

3.45 Stripping of sequencing oligos

After imaging, the anchor primer and the ligated probes were stripped off to prepare for the next sequencing cycle. First we did an ethanol series as before, mounted the hybridization chamber and washed once with DEPC-PBS-T. We then treated the sample with Uracil-DNA glycosylase (UNG) in a reaction mix containing 2 μg/μl BSA and 0.02 U/μl UNG in 1× phi29 polymerase buffer. The reaction was incubated for 10 minutes at room temperature and then washed twice with DEPC-PBS-T. As a final step we washed the slide 3 times for 30 seconds each with 65% formamide. After the stripping, the sample is ready for another round of sequencing by ligation.

All oligonucleotides were ordered from Integrated DNA Technologies (IDT), sequencing library from Life-Technologies.

3.5 Image analysis

Images from different channels were separated into different files, next was the alignment of the Cy7 images (anchor primer) from all the sequenced bases using a CellProfiler pipeline. All other images from other channels were optimally transformed to this alignment. Different features were defined as such: nuclei were defined by the DAPI stain, the cytoplasm was defined by all pixels within fixed distance from the nuclei. RCPs were defined as objects approximately 10 pixels in diameter. Information and coordinates of all identified objects were extracted and saved as a .csv file used in MATLAB to decode and plot the reads on a given background.
4.0 Results and Discussion

4.1 Assay validation

4.11 Correlation between in situ sequencing and mRNA levels

All 31 padlock probes were sequenced in situ in the SUM159PT cell line to validate the padlock probe performance. The in situ data was compared to RNA sequencing data from the same cell line (done on bulk). The correlation between the RNA sequencing and the in situ sequencing data sets was $R = 0.73$ for the raw data values and $R = 0.79$ for log2 converted values (Figure 2).

Most of the data correlated fairly well but the following points were outliers, S100A4, TWIST, ID1, SOX2 and PDGFB. The two data sets did not correlate 100% probably because of biases introduced by the techniques or stochasticity.

![Figure 2: Linear regression analysis between log2 converted reads for each gene for in situ sequencing signal reads (RCPs) and RNA sequencing reads for the breast cancer cell line (SUM159PT).](image-url)
The probes were also tested on a fresh frozen breast cancer tissue section (called sample I). This tissue section had corresponding microarray data available (Muggerud et al., 2010). Linear correlation between microarray data and in situ sequencing data in the breast cancer tissue was quite high (R = 0.93 for raw values and R = 0.64 for its log2) (Figure 3). There were two outliers, ACTB and ZEB1 which could be explained by the heterogeneous nature of the tumor (microarray data and in situ data are from different parts of the tumor) or possible probe unspecificity.

Figure 3: Linear regression analysis between log2 converted reads for each gene for in situ sequencing signal reads (RCPs) and microarray data (fluorescence signal intensity) for breast cancer tissue (Sample I).

As mentioned earlier, preliminary data and data from literature both strongly show that PDGF receptor B is highly expressed in invasive breast cancer tissues as opposed to its counterpart PDGF receptor A, which is suppressed. We observe the same trend in the microarray data but the in situ sequencing data shows a reverse of this trend (figure 4) and therefore we believe that the
PDGF receptor probes might not work properly. When we looked at the secondary structure of the PDGF receptor mRNA we observed that the target sites for the padlocks were not relaxed (i.e. there was a hairpin fold), this is a possible explanation of the failure of the probes. Therefore we relied on immunohistochemistry data for these two targets.

Figure 4: A) Graph showing the comparison between PDGFRA and PDGFRB transcripts detected with in situ sequencing vs Microarray data.

4.12 Correlation of in situ sequencing and protein levels
The same section used for in situ sequencing was later stained with an antibody against α-SMA (red stain) and panCK (epithelium) (brown). There is a co-localisation of ACTA2 transcripts and the α-SMA proteins (ACTA2 mRNA encodes the α-SMA protein) (figure 5). Some of the reads that do not correlate (co-localise) might be false positives but it might also be due to the varying stabilities and translocation between mRNA and protein molecules.
Figure 5: A and B) Fresh frozen breast tissue section stained with an antibody against α-SMA (red stain) and panCK (epithelium) (brown). Hematoxylin-and-eosin was used for staining of nuclei and cytoplasm (blue and pink, respectively). B) Symbols for ACTA2 transcripts (magenta dots) detected with *in situ* sequencing and plotted on top of the IHC image.

PDGF receptor correlation between *in situ* sequenced reads and IHC stains was quite poor. In the IHC stains most of the PDGF receptors (both A and B) were in the stroma, unlike the *in situ* sequencing data, which showed a more even distribution (figure 6). However, it is important to also note that mRNA and protein expression levels and distribution may vary partly due to their intrinsic natures (Human Protein Atlas database).
Reads from VIM showed an unexpected expression pattern. VIM is a known fibroblast marker so we expect to see its signals primarily in the stroma yet we observed a lot of them in the tumor cells (figure 7). This unexpected expression pattern could be explained by the similarity of its barcode (GACG) to the barcode of the highly expressed housekeeping gene, ACTB (GACC).

Figure 7: VIM morphological correlation in stroma (light regions) and tumor cells (dark regions)
4.2 Gene expression coupled to morphological features

*In situ* sequencing using 31 padlock probes was performed on two fresh frozen breast cancer tissues sections (sample I and sample K) (Figure 8). IHC was done on the same section with antibodies against α-SMA protein and panCK (endothelial cells). In sample K a total number of 426988 reads (85.3%) were the expected reads, 71340 (14.62%) unexpected reads and 2347 (0.5%) were homomers (background and autofluorescence). Sample I showed 127263 expected reads (84.4%), 21977 (14.5%) unexpected reads and 1480 (1%) were homomers.

![Image of molecular profiling and IHC stain with symbols](image)

**Figure 8:** Molecular profiling of two breast cancer tissue sections. *In situ* sequencing signal reads (RCPs) were plotted as symbols on top of the IHC stain (right).
4.3 PDGF signaling

4.31 PDGF Receptor Alpha

When we compared regions with high PDGF receptor A protein expression levels to regions with low expression levels, we observed different expression patterns. The regions were identified based on IHC staining of the receptor on a consecutive tissue section (Figure 9). In the PDGF receptor A protein enriched regions, genes ALDH1, CXCL14, TWIST and PDGF receptor A mRNA were highly expressed. On the other hand CDH1 and GDF15 expression was relatively lower.

![Figure 9: Gene expression profiles for PDGFRA enriched regions](image)

*Relative frequency is defined as the number of detected signals per pixel.*
4.32 PDGF Receptor Beta

We also compared regions with high PDGF receptor B protein expression levels with regions with low expression levels. Again, the regions were identified based on IHC staining of PDGF receptor B proteins in a consecutive tissue section (Figure 10). In the PDGF receptor B protein enriched regions, genes ALDH1, FAP, GLI1, ID1, TWIST and CXCL14 expression was relatively higher. On the other hand, CDH1 and HEY1 expression was relatively lower.

Certain genes such as ALDH1, CXCL14 and TWIST were enriched in both PDGFRA and PDGFRB rich region while CDH1 expression was repressed for both receptor types.

Figure 10: Gene expression profiles for PDGFRB enriched regions (Sample K) A) IHC with PanCK antibody (brown) and ASMA antibody (red) was performed on fresh frozen breast tissue section that has undergone in situ sequencing. B) Consecutive section fresh frozen breast tissue section stained with PDGFRB antibody (brown) and PanCK antibody (red). C) Expression profile of PDGFRB protein enriched regions. *Relative frequency is defined as the number of detected signals per pixel.
We observed expression patterns that were unique to each region (summarized in table 2). Of all the regions compared, PDGF receptor A mRNA molecules were enriched only in the PDGF receptor A protein enriched region, while GDF15 was repressed only in this same region. In parallel, FAP, GLI1 and ID1 were enriched only in the PDGFRB protein rich region while HEY1 was only repressed here. ‘Repression’ and ‘enrichment’ here is relative to the regions compared.

Table 2: Summary of expression profiles unique in PDGFRA and PDGFRB enriched regions

<table>
<thead>
<tr>
<th>Enriched</th>
<th>PDGFRA</th>
<th>PDGFRB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PDGFRA</td>
<td>FAP, GLI1, and ID1.</td>
</tr>
<tr>
<td>Repressed</td>
<td>GDF15</td>
<td>HEY1</td>
</tr>
</tbody>
</table>

Enrichment of GLI and ID1 in the PDGF receptor B region is in line with the prognostic relevance of stromal PDGF receptor B. Hedgehog signaling pathway has been implicated in propagating a wide range of cancers including breast cancer (Barcellos-Hoff & Akhurst, 2009) and GLI1 is a commonly used a marker for Hedgehog signaling while ID1 is a marker for the TGFβ signaling. Both pathways have been implicated in promoting breast cancer development (Kasper, Jaks, Fiaschi, & Toftgård, 2009) (Barcellos-Hoff & Akhurst, 2009).

4.33 **PDGF ligand expression**

All PDGFR ligands were enriched in the tumor cells and were less expressed in the stroma (except for PDGFD which showed a low signal density overall). Expression of the ligands in the tumor cells coupled with the low expression of the receptors in these tumor cells (figure 11 and
figure 12) suggest that stimulation of the PDGF pathway in the stromal cells came mostly from tumor cells.

![PDGF expression images](image)

**Figure 11:** Example of the distribution of PDGF ligand and PDGF-receptor-B expression in tumor cells and its microenvironment (Sample K). Images 1-4: *In situ* sequencing signals (red) were plotted on top of the IHC; tumor cells (dark greenish), stromal cell nuclei (blue). Image 5: Tumor cells (purple) PDGF receptor A (brown)

![PDGF expression images](image)

**Figure 12:** Example of the distribution of PDGF ligand and PDGF-receptor-A expression in tumor cells and its microenvironment (Sample K). Images 1-4: *In situ* sequencing signals (magenta dots) were plotted on top of the HE-IHC stain; tumor cells (greenish), stromal cell nuclei (blue). Image 5: Tumor cells (purple) PDGF receptor B (brown)

4.34 **Expression profiles of tumors and the tumor microenvironment**

We observe a localization of ACTA2 transcripts around the tumor/stroma border (figure 13).

Translation of ACTA2 transcripts yields α-SMA proteins, which are generally used as markers for smooth muscle cells. OCT3 is a stem cell marker also used to create induced pluripotent stem
cells (Wernig et al., 2007). Its expression is highly regulated as it is critical in maintaining stem cell properties of the inner cell mass of blastocyst and embryonic stem cells (Rodda et al., 2005). In this study we observe an enrichment of OCT3 mRNA molecules in the inner cell mass of tumor cells (figure 13). GAPDH is a housekeeping gene and it is more strongly expressed in the inner cell mass of the tumor, this could be explained by higher cellular activity in the middle of the tumor. We found the expression of the chemokine CXCL14 to be higher in the stroma, it has an anti-cancer role and its repression indicates a poor prognosis for breast cancer patients (Gu et al., 2012).

![Gene expression patterns in tumor cells and its microenvironment. In situ sequencing signals (magenta dots) were plotted on top of the HE-IHC stain; tumor cells (greenish), stromal cell nuclei (blue).](image-url)

Figure 13: Gene expression patterns in tumor cells and its microenvironment. In situ sequencing signals (magenta dots) were plotted on top of the HE-IHC stain; tumor cells (greenish), stromal cell nuclei (blue).
5.0 Conclusion
All probes were validated by comparison with RNA sequencing, microarray, and immunohistochemistry. There was a significant correlation for most of the probes in all three techniques except for VIM, ACTA2, PDGFRA and PDGFRB for which there was evidence suggesting otherwise. Co-localisation analysis suggest that PDGFR regulation might involve TGFβ signaling and Hedgehog signaling. PDGF ligand and receptor expression patterns suggest that growth of stromal PDGF-receptor-B-positive cells is stimulated predominantly by tumor cells. We also noticed that stem cell marker OCT3 and GAPDH are highly expressed in the interior of tumor cells. ACTA2 is enriched in the tumor/stroma border while CXCL14 is mostly expressed in the stroma.

6.0 Acknowledgement
First of all I thank my very able supervisor Jessica Svedlund for the tireless support she has offered me but I would also like to thank our group leader, Professor Mats Nilsson for offering me the opportunity to work with all the great people in Molecular Diagnostics Lab. The study opportunity was sponsored by the Swedish Institute and I would like to thank them for their generosity. Thanks to all my friends and family.
7.0 References


http://doi.org/10.2214/ajr.126.6.1130