Supported by EC Grant Agreement n°259881

Start Date: 01/05/2011- Duration: 48 months

Coordinator: Pr. Kerry CHESTER, University College London

PROJECT FINAL REPORT

Grant Agreement number: 259881

Project acronym: IMAGINT

Project title: HER Imaging and Molecular Interaction Mapping in Breast Cancer

Funding Scheme: Collaborative project

Period covered: from 01/05/2011 to 30/04/2015

Name of the scientific representative of the project's co-ordinator, Title and Organisation:

Kerry CHESTER, Professor, UCL Cancer Institute

Tel: +44 20 76790700

Fax: +44 79 85192051

E-mail: k.chester@ucl.ac.uk

Project website address: http://www.imagint.eu/

TABLE OF CONTENTS

I. Executive summary

IMAGINT is a 48-month collaborative project addressing the objectives of the HEALTH priority of the FP7 Cooperation Programme, and in particular the area “detection, diagnosis and monitoring” applied to cancer. IMAGINT focused on developing tools to investigate factors associated with the human epidermal growth factor receptor 2 (HER2) pathway. HER2, a member family of the tyrosine kinase cell surface receptors (HER1/EGFR, HER2, HER3 and HER4), is a potent oncoprotein that has become an established target for breast cancer treatment. There is urgent need for development of safe, efficient non-invasive HER2-related biomarkers to guide treatment regimes and for prediction, diagnosis, monitoring and prognosis of disease. IMAGINT aimed to develop such tools by combining bio-imaging and molecular testing biomarkers from RNA, DNA and/or protein and their complexes.

IMAGINT brought together an interdisciplinary team to achieve its goals and utilised the potential of Designed Ankyrin Repeat Proteins (DARPins). These are small, antibody-like proteins based on human protein scaffolds. DARPins bind specific targets with high affinity in monovalent form and are readily engineered for site-specific chemical modification to create specific marker-tags.

The consortium developed: new ways of creating and manipulating DARPins; a range of new tools to detect HER2, its interacting partners and associated pathways; and an innovative data bioinformatics system for data handling and analysis. The tools were validated on cells and/or human tissues/xenografts, some showing exciting potential to provide greater understanding of HER2 driven disease and its treatment: (i) Super-resolution microscopy revealed new details in subcellular HER2 distribution, opening the possibility of eventually screening individual patient tissues resected during breast surgery to support personalized treatment schemes. (ii) Imaging Cycler Microscopy (ICM) discovered differences between breast cancer and corresponding healthy tissue, showing potential to provide valuable information about protein clusters and their associated biology. (iii) RNA/protein interaction studies of human cancer cells revealed differences in associating miRNAs/proteins in response to treatment with anti-HER2 therapy. Bioinformatic analysis of gene expression for the differentially detected proteins revealed a possible association with time to progression in breast cancer; these genes warrant further investigation as prognostic biomarkers in patients.

Some IMAGINT tools were developed far enough to demonstrate utility in clinical application. The FRET/FLIM (Förster Resonance Energy Transfer Fluorescence Lifetime Imaging Microscopy) assay to detect EGFR heterodimers (HER2-HER3 and HER1-HER3) in archived human tissue from clinical trials was successfully applied to a large number of patient samples and the results analysed with novel Bayesian statistical methods. As a result The HER1-HER3 assay has validated a new mechanism of resistance to anti-EGFR therapy in patients with triple-negative breast. In another breast cancer study, the HER2-HER3 assay has shown superior prognostic value to the standard HER2 staining assay.

Finally, a new agent for whole body imaging of HER2+ve tumours has been fully developed and generated by the consortium for clinical use. The new imaging agent will be trialled in HER2+ve patients with advanced breast cancer and gastric/gastro-oesophageal adenocarcinoma. The test is non-invasive and has potential to provide hitherto unknown information about the extent of HER2+ve disease and enable clinicians to make informed decisions regarding treatments.

IMAGINT has succeeded in its aims to develop tools that can eventually be used by clinicians to guide treatment regimens and for prediction, diagnosis, monitoring and prognosis of disease.

II. Project context and objectives

IMAGINT is an interdisciplinary project that aims to develop a range of new **tools** for **imag**ing the human epidermal growth factor receptor (HER) family of tyrosine kinase cell surface receptors and their **int**eractions in breast cancer.

Context of the project

The incidence of breast cancer in Europe is > 430,000 per annum. In this chronic disease, individuals whose tumours show increased levels of the HER family member, HER2, are known to have more aggressive cancers and high mortality. HER2 is a potent oncoprotein with a central role in the development and maintenance of breast cancer and, consistent with this pivotal function, HER2 has become an established target for breast cancer treatment, both with antibodies and small molecules. For instance, trastuzumab (Herceptin), a humanized monoclonal antibody against the extracellular domain of HER2 is standard of care in adjuvant treatment and first line treatment of metastatic disease. Current antiHER2 treatments are effective but tumours still progress due to innate or acquired resistance and there is need for new HER2-related biomarkers to guide treatment regimes and for prediction, diagnosis, monitoring and prognosis of disease. Furthermore, innovative new targeted agents for HER2 positive (HER2+ve) disease are in development. Here informative biomarkers are required to help determine optimal duration and sequencing of treatment to maintain HER2 suppression.

IMAGINT aimed to address the urgent need for new biomarkers by developing a new range of robust imaging tools. Because of its central role, IMAGINT focused on tools to dissect HER2-related events. Once developed for breast cancer, it was predicted that the tools would find wide application for other cancers in which HER2 is clinically relevant, for example, gastric cancer, another chronic disease and the world's second leading cause of cancer death. To achieve its goal, IMAGINT employed the properties of Designed Ankyrin Repeat Proteins (DARPins), which are antibody-like proteins based on human protein scaffolds but 10 times smaller than antibodies. DARPins are highly stable and bind specific targets with high affinity in monovalent form; they are also readily engineered for site-specific chemical modification to create specific marker-tags and thus allow for the creation of tools that would be difficult or impossible to create from antibodies.

Concepts of the project

Despite the pivotal role of HER2 in breast and other cancers, the current cutting-edge tools and technologies used in biomarker studies of HER2 related events are confronted by many limitations for early diagnostic, prognostic and monitoring of disease. There are no established tools for whole body quantitative clinical imaging to assess the extent and location of HER2+ve metastases, or for reliably measuring the wide range of HER dimers that occur in human tissues. Furthermore, the RNA/protein interactions that control behaviour of HER2+ve cancer in response to treatment and the topographical protein clusters linked to HER2 signalling remain unexplored. This information cannot be obtained by gene expression microarray studies. Moreover, there is the need for advanced mathematical and computational tools to enable integration of the different types of biomarker information in a holistic manner. As a result of the current tools' limitations, the segregation of breast cancer into more specific subsets is incomplete. Nonetheless, a cost effective approach to the disease in which cancer detection, diagnosis and treatment are tailored to each individual molecular profile is essential. IMAGINT aimed to address these issues and develop innovative DARPin-based tools that will enable researchers to discover new biomarkers – such as HER/HER protein interactions, RNA/protein interactions and cellular clusters, metastatic signatures, imaging agents and software tools - that will enable stratification and monitoring of breast cancer. It was envisaged that the information obtained with these biomarkers would lead to better treatment decisions for existing therapies and assist in the development of new drugs.

Concepts behind characterizing and quantifying the new biomarkers:

- **HER/HER protein interactions:** The HER family of proteins are cell surface tyrosine kinase receptors that are involved in regulation of cellular proliferation and survival. HER2 - the focus of IMAGINT - is activated by a variety of mechanisms, including overexpression. However, HER2 does not work in isolation but functions as homodimer, or a shared co-receptor with other members of the family. There are 4 members of the family; HER1 (EGFR), HER2, HER3 and HER4 and their dysregulation is related to a large fraction of human cancers. HER2 is the preferred dimerisation partner of all family members but the family can undergo a wide variety of homo- and heterodimeric associations, depending on their expression levels, stimulation by external or autocrine ligands, receptor mutations and changes in the molecules interacting on the cytoplasmic side. Therefore, the measurement of homo- and heterodimeric association, as well as potential clustering of receptors or receptor pairs could provide the scientific basis for stratifying tumours and for developing next-generation tools. HER2 plays a central role and research indicates that specific HER dimers will be more important and informative biomarkers than HER2 alone. For example, HER3 has been found to preferentially interact with HER2 and the HER2/HER3 heterodimer is purported to be a particularly potent mitogenic and oncogenic unit for a variety of cancers including breast cancer. Additionally, there is accumulating evidence that two different complexes of HER2 and HER3 exist, one with the liganded form of HER3, the other with the non-liganded form of HER3, and these are formed as a function of HER2 level. While signaling in both cases appears to occur mainly through HER3 phosphorylation, the two different complexes have an important therapeutic consequence: only the second one reacts to trastuzumab. HER2 can also be partnered with HER1, depending on the expression level and on particular mutations in the HER1 gene, which can have important consequences for signaling and response to anti-HER2 therapeutics. In summary, there is increasing evidence that measuring and understanding the real HER2 complexes present in particular tumours will be of benefit for the development of rational therapies. There is urgent need for robust tools to measure these different HER dimers in human tissues. As biomarkers, the specific dimers could provide prognostic information on disease outcome and diagnostic information on disease stage and response to therapy. This will allow the segregation of disease into more specific sub-sets and a personalised therapy could be designed for specific dimer types.
- **Molecular networks:** Unravelling the complexity of the interacting molecules can reveal undiscovered biomarkers or diagnostic signatures. This is important because the therapeutic responsiveness of cancers to HER-targeted therapies does not necessarily correlate simply with the receptor levels. For instance, when trastuzumab is given as a single agent for first-line treatment of HER2-overexpressing metastatic breast cancer, it is associated with only a 40% objective response rate. In addition to primary resistance, HER2+ve breast cancer patients who initially respond can subsequently progress/relapse on trastuzumab treatment (acquired resistance). Also, the incidence of mutations (such as AKT1 and PTEN) in breast tumours is typically low and there may not be complete penetrance. Studying protein function (in terms of activity and spatiotemporal location) has potential to give a broader patient base to integrate into a diagnostic signature, regardless of whether patients have germline or somatic mutations. The protein clusters linked to HER2 signalling cannot detected by gene expression microarray studies but hundreds of molecules can be visualised simultaneously on cancer cells by high throughput robotic handing of fluorescent probes to provide detailed cellular images of co-localised molecules. When this information is combined with disease outcome it could provide new clinically-relevant biomarkers. There is a need to develop imaging and mathematical tools to identify these clusters and link them to disease status.
- **RNA/protein interactions:** Although the signalling pathways following HER activation are well defined and their potential as targets investigated, RNA/protein interactions controlling these pathways are not yet elucidated. These complexes of interacting molecules could be an untapped source of new biomarkers for diagnostic purposes and tools are required for their detection, isolation and functional characterisation.
- **Addressing metastases**: Most deaths from cancer are not directly due to the primary tumour but as a consequence of disseminated metastases. Many patients diagnosed with early breast cancer can expect to survive, due in part to the widespread adoption of adjuvant systemic therapy to combat micrometastatic disease. Some patients derive little benefit from systemic

treatment, however; either because their tumours metastasize despite receiving chemotherapy, or because their tumours would never have metastasized. There remains a critical need for diagnostic tools to identify those individuals who are at high risk of metastatic recurrence and rationally select therapy to which they will respond. IMAGINT aims to develop practical and informatics tools to achieve this.

- **Imaging HER2+ve disease:** Despite the pivotal role of HER2 in diagnosis and treatment of cancer, there are still no established tools for quantitative clinical imaging of the extent and location of metastatic disease. There is a need to develop these tools because non-invasive whole body quantum imaging of HER2+ve metastatic cancer in patients could provide important clinical diagnostic information by early detection of sub-clinical HER2+ve disease, optimal management of current anti-HER2 therapies and response assessment of novel therapeutics.
- **Systems biology:** The new types of biomarker described above will each provide distinct information and understanding about HER2 and its role in driving malignancy. However, increased knowledge and greater understanding can be achieved by integrating the different levels of information in a holistic manner, with each other, with public databases and with information on clinical outcome. Achieving this requires building an efficient integrated database for the accumulating biomarker and clinical data, and the development of new mathematical and computational tools with which to bring together the various sources and quantify their (combined) predictive potential.

III. Main S&T results/foregrounds

III.1. WP1 – Generation and development of DARPins as tools for quantitative imaging and detection of interacting partners

UZH have obtained binders against all members of the EGFR family using both phage display and ribosome display. To extend the generation of DARPins to particular epitopes on all members of the EGFR family, further panning was carried out and additional binders were obtained. To complement this set, a new strategy was also developed, especially to "transplant" binders to valuable epitopes to other family members. For this purpose, use was made of the structures of the anti-HER2 DARPins (Jost et al 2013). The approach consisted of analyzing the structures by ROSETTA, and then replacing computationally the HER2 domain by the corresponding HER3 domain, and calculating the required sequence for the DARPin to bind specifically. This approach was applied to both domain I and domain IV. Furthermore, this approach was used to adapt the DARPins to recognize the murine version of ErbB2.

UZH has also developed robust labelling strategies for DARPins (Tamaskovic et al 2012). The most useful method for this approach has been selective labelling with cysteine-maleimide chemistry. DARPins with multiple cysteines have been engineered, to permit a greater resistance against photobleaching in STED (super-resolution super resolution microscopy) experiments. The cys-maleimide labelling has now been done for over 100 pairs of DARPins and dyes and generally works with >95% yield. Additionally, selective labelling at the N-terminus has been established, and selective labelling with click chemistry, either at the N-terminal Met or at an artificially introduced single Met. Using fluorescently labelled DARPins or fluorescently labelled HER2 in native cells, very encouraging first results from single-molecule detection on cells could be obtained. The obtained diffusion values are within expectations on cells, and suggest that this method can be used in the future to determine the diffusion characteristics of HER2 on cells.

UZH has also developed proximity ligation assays (PLA) with DARPins (Gu et al 2012) for the detection of HER2, UZH developed a straightforward and convenient approach to functionalize recombinant affinity reagents for PLA by expressing the reagents as fusion partners with SNAP protein tags. The crucial improvement is the rigorous purification of the DARPin-oligonucleotide conjugate away from both unconjugated oligonucleotide and from unconjugated protein. After

having successfully established a proof-of-principle using HER2, the method has been significantly improved, by use of a new oligo-capturing strategy (Yan et al 2014), This new strategy leads to a significantly lower background in whole cells. A quantitative analysis shows this very clearly.

Most recently novel chemistries have been develop which no longer require to express the DARPins as SNAP tag fusions, which are (unlike the word "tag" suggests) large enzyme domains. For this purpose, UZH has concentrated on two strategies, both using click chemistry. On the one hand, methionine analogs were used with "clickable" amino acids, on the other hand modified suppressor tRNAs were used for *in vitro* synthesis. Both methods were shown to function, but at the moment the tRNA method requires more expensive reagents.

UZH has also worked closely with UCL to produce G3 DARPins and UZH performed analysis of DOTA-maleimide conjugated HE3-G3 DARPin samples by mass spectrometry (MS) and surface plasmon resonance (SPR). Both MS and SPR showed good quality and reproducibility of the material (WP5).

III.2.WP2 – Tools for a multivariate tumour invasion signature

KCL developed a number of fluorescence based protein detection and FRET-FLIM assays for quantitative measurement of protein levels and their interaction in archived clinical samples (formalin fixed paraffin embedded, FFPE, tissue). Those assays were successfully applied on to samples from three independent cohorts of patients (one breast cancer and two colon cancer patients). Obtained data was analysed as described below (WP6). In order to achieve this goal KCL needed to overcome three challenges.

1: Currently companies producing antibodies adopt rigorous testing to ensure specificity and reliability of antibodies for different applications (western blot, ELISA, immunofluorescence and immunohistochemistry). FRET/FLIM assay development requires additional modification of primary antibodies by attaching fluorophore molecules to the antibody. KCL screened more than 20 antibodies for their ability to recognise target protein in FFPE tissue after labelling with Alexa546 (donor fluorophore, directly exited by light source) or Cy5 (acceptor fluorophore, in close proximity to donor fluorophore can absorb energy of excited donor fluorophore leading to decreased time donor fluorophore spent in excited state, fluorophore lifetime (this is used to detect distance between two fluorophore and hence between proteins labelled with antibodies). The majority of antibodies lost this ability after labelling. To test specificity, KCL utilised breast cancer cells with over-expressed target protein. Those cells were processed to obtain paraffin blocks, cut at 3µm slices and subsequently underwent similar staining procedure as tissue samples. Since KCL had ability to control the expression level of the proteins of interest they also created conditions when interaction between two proteins was forced, enabling them to detect it consistently with labelled antibodies. This test was used every time KCL performed the FRET/FLIM assay on a new cohort of patients' samples and also for quality control for new batches of labelled antibodies.

2. The second challenge was to develop a staining protocol, which would maximise epitope availability for antibody, thus increasing signal intensity consequently leading to increase in signal to noise ratio and minimise background autofluorescence signal. For FRET/FLIM assay background fluorescence cannot be eliminated by simple thresholding of signal intensity as can be done for any confocal/epifluorescence imaging. Hence its contribution needs to be diminished by chemical treatment of tissue before staining. Thus KCL screened multiple protocols for antigen retrieval, including wash buffers, blocking agents and staining regimes to find the best option to produce the strongest signal from antibody specific staining whilst reducing the background signal from autofluorescence.

3. The third challenge was to develop an image analysis algorithm, which can account for any tissue contaminations leading to spurious measurements. This is critical because connective tissue, infiltrating blood cells and multiple products of cell metabolism can contribute to fluorescence signal and can be stronger than signal from the fluorophore. However this fluorescence has distinct low lifetime compare to fluorophore lifetime and thus can be removed from the image prior to extracting true lifetime of the donor fluorophore. After thorough testing KCL succeeding in developing an analysis method, which can eliminate the influence of fluorescence signal that is not related to donor fluorophore.

Newly developed FRET/FLIM assays and image analysis were applied on large number of patient samples and results of those studies are currently in preparation for publication.

TNL used their proprietary Imaging Cycler Microscopy (ICM) technology (previously termed toponome imaging system / TIS, or Multi epitope ligand cartography / MELC) (Schubert et al 2006; Friedenberger et al 2007; Schubert et al 2014) to investigate the spatial protein network code (the toponome) of breast cancer tissue. ICM is a cycling procedure that can co-map thousands of protein complexes or assemblies at a time and at spatial dimensions. By overcoming the spectral fluorescence resolution limit, this functional super-resolution technique has potential to identify HER associated breast cancer specific protein complexes *in situ* and provide novel insights into mechanisms of breast cancer as a basis for diagnostic and/or therapeutic options. Informative tag libraries for toponome studies were available in the TNL laboratory; most of the tags are directed against cell surface molecules (Friedenberger et al 2007; Schubert et al 2006). Thus, TNL were able to start the IMAGINT project with a tag library already validated on prostate cancer (Schubert et al 2009) and subsequently extended the library stepwise with markers against tyrosine kinase receptors, signal transduction proteins as well as transport and cytoskeleton proteins.

As appropriate breast cancer was not available at the start of IMAGINT, TNL used obtainable epithelial tissue to calibrate and classify a corresponding large tag toponome library with an improved toponome imaging approach, later transferable to breast cancer tissue. The rational was as follows: breast secretory tissue (glands) and the epidermis of the skin are both epithelial tissues and have many aspects of this tissue type in common. Classification and calibration of a large toponome library to skin tissue of human biopsies (stored in TNL lab biobank) was taken to fulfil this task. A 100- component molecular library was calibrated to epidermal tissue of human skin (Schubert et al 2012). In these experiments TNL focussed on the epidermis cells and basal lamina, a limiting membrane like structure, which shows rather constant structural features across all tissues (including normal breast tissue), thereby separating epithelial cells from juxtaposed connective tissue, by forming 3 distinct, electron microscopically distinguishable layers. The main findings were: (a) Surprisingly by using similarity mapping (Dress et al 2005; Dress et al 2008) in combination with ICM, TNL were able to identify substructures of the basal lamina (Schubert et al 2012) and could distinguish these layers by visualising a 100-dimensional biomolecular profile specifically associated with the lamina densa. The fact that these insights can be obtained for the first time with a normal epifluorescence microscope by co-mapping a very large number of proteins without having to change hardware is a new dimension of protein systems microscopy. With these advantages and based on the feasibility to visualize large cell surface toponomes, as evidenced in blood T cells (Friedenberger et al 2007), TNL were able to jump start the decoding of the protein systems in breast cancer tissues of different stages.

The breast cancer study started with cryo-sectioning of both normal breast tissue and cancerous breast tissue simultaneously. The first step, after having collected high quality tissue sections, was to perform large series of tests for tissue quality related to the toponome measuring procedures. This was followed by tests with a toponome library that had been pre-calibrated in the first part of IMAGINT (skin, see above). Additionally TNL included fluorescent dye (FITC)-labelled antibodies recognizing different epitopes of HER2- and HER3-receptors. This calibration procedure extended the library progressively in order to find HER2-associated protein clusters. For this procedure TNL applied a similarity approach to interrogate the known threshold based approach to toponome

quantitation (Schubert et al 2006) and align these data with non-threshold based *in situ* protein profiling (Schubert et al 2012). From these comparative approaches TNL was able to decode and understand stepwise the large spatial multi-protein complexes embedding HER receptors. This part of TNL IMAGINT showed that HER2 and HER3 can be dominated spatially by a multi-protein complex that TNL called the CAP complex which appears to be directly associated with HER2 and HER3 internalization.

For each type of breast cancer (2 x HER+ve, 1 x Luminal B) and their corresponding control tissues TNL found a typical CMP (combinatorial molecular phenotype) motif using a tag library to co-map up to 72 different bio-molecules at a time. Except in one case of breast cancer, HER2 was not found as a lead protein. Instead, a signal transduction pathway protein was found as a lead protein in all cancer types. It plays a role as a mediator in the TGF-β pathway and was found in colocation together with a vesicle trafficking protein and surface molecules. Hence, in these samples, there appears to be a specific supramolecular order of the breast cancer toponome that dominates HER2 co-assemblies *in situ*. Together, these quantitative toponome data substantiate the spatial data of the CAP complex; composed of many signal transduction proteins embedding HER receptors.

The toponome analysis of the control tissue showed that these also have a CMP motif in common. This motif shows hierarchical control by three different lead proteins: a cell surface trans-membrane protein, a structural intracellular protein, and a signal transduction protein. Surprisingly, the latter is the same transduction protein we found as a lead protein in the cancerous tissues, but here in a cancer specific new assembly with other proteins. Due to the lack of real (healthy) control tissue (lacking foci of acini pathology or damage) TNL used perifocal normal tissue as “control tissue”. A specific motif discriminating early cancer stages (“controls”) from advanced cancer stage was found. The corresponding toponome, among other things, substantiates that myoepithelial cells are involved in early stages of breast cancer. Following toponome hierarchies and implications of lead proteins controlling disease specific toponomes, TNL suggested that a signal transduction protein as lead protein and part of the TGF-β pathway is a critical component in all analysed breast cancer cases. This is a clinically important notion since similar detection of a critical lead protein has led to efficient therapy of ALS in a clinical trial (Schubert, 2015).

The detailed data of the breast cancer study are now assembled in a toponome data base to be processed for exploitation.

III.3. WP3 – Developing methods for isolation characterization of protein/RNA complexes from clinical tissues

The goal of the WP3 was to develop methods to identify proteins and RNAs involved in HER2+ve breast cancer. In order to accomplish the objectives, several imaging methodologies were exploited:

Initially the consortium tried to develop DARPins that would recognize Argonaute AGO2 protein, as a tool for isolating RNA-induced silencing complex (RISC) complexes for further protein and RNA content analysis. However, obtaining the required amount of AGO2 turned out to be troublesome. INO explored several expression systems (mammalian, bacterial and insect cells) but none of them yielded sufficient protein for DARPin production (mainly due to protein aggregation). However, after substantial process development INO was able to meet the challenge by successfully using endogenous AGO2 to purify the RISC complexes from cancer cells. Encouragingly, using this approach, INO obtained sufficient amounts of protein complexes to use in proteomic approaches. Thus, INO were able to compare RISC from a HER2+ve breast cancer cell line (BT474) in the

absence and presence of the pharmaceutical anti-HER2 antibody, trastuzumab. The two isolated protein complexes were characterized by mass spectrometry with exciting results: in total, around 1365 high confidence proteins (less than 1% mismatch error) were identified with at least two peptides matched. This number was later narrowed down to 209 proteins that have been selected for further study.

In order to increase the yield of proteins and miRNAs detected in the immunoprecipitation procedure, INO included a crosslinking step prior to the immunoprecipitation. This broadly used technique (so called CLIP: Crosslinking immunoprecipitation) stabilizes the interactions between the molecules, facilitating their co-isolation and downstream analysis. AGO2 RISC complexes were isolated by crosslinking immunoprecipitation (CLIP) using UV irradiation (physical method) or formaldehyde (FA) (chemical method) and compared with isolations performed without the crosslinking step. The isolated miRNAs from these studies were prepared for further sequencing analysis. However, as both crosslinking methods dramatically reduced the amount of RISC complexes obtained INO decided to use the material immunoprecipitated from un-crosslinked samples for further studies.

The functionality of the RNA/RISC complex depends on its location in the cell; the complex must be part of the P-bodies or Processing-bodies, the place in the cytoplasm where processing and degradation of the RNA takes place and where the RISC machinery works. In order to use the location of the RNA/RISC complexes in the cell as a complementary marker for studying the effect of anti-cancer compounds, AGO2 and GW182 proteins were detected by immuno-fluorescence. AGO2 is one of the main components of the RNA/RISC complex and GW182 is a protein present in the Processing-bodies and one of those responsible for their stability. To achieve this GW182 and AGO2 protein were stained using specific antibodies, labelled with different fluorochromes and their co-localization determined in different cell lines (CHOk1, HEK293 and SH-SY5Y). The results showed that the approach was feasible: AGO2 localized in CHOk1 nuclei but appeared diffused in the cytoplasm and also restricted to P-bodies in HEK293 and SH-SY5Y cells. Thus, in this task INO established the methodology to study RISC/P-bodies co-localization for subsequent application to human HER2+ve treated breast cancer cells.

INO then used the methodologies developed in previous tasks in order to isolate AGO2-containing RISC complexes from BT474 cells. After isolation, the proteins were identified by mass spectrometry and miRNAs present in those complexes were identified by RNA-sequencing at UCL Genomics and compared with samples treated with trastuzumab to find molecules that could be involved in the trastuzumab response. The comparative study raised 60 proteins uniquely identified in samples treated with trastuzumab whereas the untreated sample showed 34 proteins not present in the former. Network analysis of these protein lists at UCL revealed strong enrichment for proteins involved in RNA processing and RNA silencing. In addition, following analysis at UCL, 13 miRNAs were preferentially found in the trastuzumab treated samples treated vs 6 miRNAs that were mainly present in the untreated samples. Predicted mRNA targets of these miRNAs were determined by KCL and these were further analysed in WP6.

INO then investigated human breast cancer HER2+ve cell lines (BT474 and SKBR3) untreated and treated with trastuzumab in order to investigate whether the P-bodies containing AGO2 could be used as biomarker for response. With the BD Pathway 855 image platform the co-localization of both proteins was studied by immunofluorescence and number and size of AGO containing P-bodies after the treatment with trastuzumab was analysed. This study was done according to different parameters; i.e. the granularity and average area of AGO2 fluorescence and the co-localization with GW182. The results showed that trastuzumab reduced RISC/P-body co-localization after oxidative stress, suggesting that the treatment could affect the location and consequently the activity of AGO2 and its function as miRNA regulator.

Finally, INO investigated the transfer of their newly established IMAGINT technology to human samples. HER2+ve and HER2-ve tumour samples were tested, including HER2+ve tissues from

patients treated with trastuzumab. The most interesting proteins and miRNAs from cell studies were evaluated. First trials showed that, whilst further technical optimization is needed, INO were confident that it would be possible to next explore this direction and determine miRNA and protein co-localization with AGO2 and GW182 (markers for RISC and P-bodies).

III.4.WP4 – Development of radiolabelling and software for quantum imaging

WP4 has been responsible for actions designed to develop optimal radiolabelling for the clinical imaging trial. The development of safe and efficient radiolabelling protocols are necessary because the amount to be injected has to be the same each time and for each patient, and the purity of the injected radiotracer has to be under rigorous control in order to avoid risk for unexpected by-product toxicity. Reproducibility is a key issue.

The development of radiolabelling protocols included the purity analysis of the expressed proteins delivered from UZH and from UCL as well as the amount of radioisotope incorporated as a measure of radiolabelling efficiency. Detected presence of impurities and aggregation of DARPins due to oxidation of Cys residues was reported back to UZ and UCL. It was also evaluated if batches could be handled under conditions to allow a scale of production large enough for clinical imaging. The estimated number of patients and the appropriate doses for each patient were used to determine how large a batch that had to be produced in order for a single batch to be sufficient for the entire trial. This batch was considerably larger than the small batches required for animal studies, and therefore required conditions for large scale production. While batches of less than a mg were typically used for animal studies it was concluded that batches of 30 mg had to be produced for the human trial. Oxygen-free atmosphere surrounding FPLC columns and chromatography elutes were required to prevent oxidation of the Cys side chains and ensure quantitative conjugation of the chelator to the DARPins.

A key part of the project was to identify the best radiotracer for the medical imaging trial, and for this reason the chelator was changed several times and evaluated for imaging quality. The two types of chelators considered were the well-established DOTA and NOTA groups that were incorporated into the DARPin by conjugation to a cysteine residue introduced into the DARPin scaffold by mutagenesis. The free thiol group of the Cys residue was used to anchor the chelator using the well-established Michael addition reaction of the maleimide group. Radiolabelling efficiency at temperatures that did not denature the DARPins was an important parameter. Also the protein expression systems were changed as was the “tag” introduced for protein purification purposes. All in all four different protein batches were evaluated, one of them at two different batch sizes, in combination with two different chelators. Efficiency and yield of conjugation were carefully evaluated for close to ten different constructs. As a consequence the radiolabelling protocols, too, had to be evaluated numerous times in order to obtain the best combination of expression system, tag and chelator. The tasks took advantage of a set of well-established laboratory procedures and were crucial for the development of the optimal radiolabelling protocols for the clinical trial.

Examples of analytical tools for evaluation included SDS-PAGE, mass spectrometry and radiodetector based HPLC. The levels of conjugation were rigorously evaluated with established bioanalytical tests of the availability of free cysteine after conjugation and DARPin concentrations were determined in order to determine conjugation kinetics. The SDS-PAGE analysis provides extremely sensitive evaluations of the presence of protein by-products and was used to secure that the protein did not contain detectable amounts of unlabelled or dimerized DARPin. Mass spectrometry was used to identify that only one chelator had been introduced, and HPLC with radiochemical detection was used to show the amount of radioisotope that had been incorporated.

A protocol for the conjugation of maleimide-linked DOTA to DARPins at the scale suitable for 30 mg batches was delivered to UCL in preparation for GMP production.

When robust conditions for the conjugation of chelator to protein had been established, the radiolabelling procedures were evaluated and optimized. At UU a microfluidic device that allowed miniaturized radiolabelling was used, in order to demonstrate minimal consumption of radioisotope as well as protein, ultimately to be used bedside. The lowest amount of DARPin that could be radiolabelled was 5 nmol. The concentration was optimized, aiming at the lowest concentration possible without experiencing loss of efficiency. The volume that can be used for radiolabelling is 20 uL, and consequently a concentration of around 70 uM.

The microfluidic device had been developed to a level where purification of extremely small amounts of radioisotope by an ion exchange column could be achieved on-line whereas the actual incorporation into the protein remained to be demonstrated and was carried out ex column. Steps were taken during the project to include protein radiolabelling but were not yet completed at the conclusion of the IMAGINT project. The required comparison between PET and SPECT imaging, to be described below, provided the possibility to use the PET isotope ^{68}Ga, with a half-life of 68 min, a radioisotope that does not require access to a cyclotron, and can therefore be used for PET imaging at hospitals where such expensive equipment is not available. ^{68}Ga is obtained from a column where ^{68}Ge is immobilized, by eluting the $^{68}Ga^{3+}$ ions that arise as a result of radioactive decay of the Ge radionuclide. Impurities from the column include other metal ions which needed to be removed to achieve high specific radioactivity. Using the microfluidic device it was possible to take advantage of the coordination chemistry of the Ga^{3+} ion and separate it from other cations by the addition of high concentrations of NaCl. The radiolabelling of DARPin conjugates with ^{111}In was carried out without microfluidic purification as the delivered radioisotope was pure enough.

The purified ^{68}Ga as well as the SPECT radioisotope ^{111}In, obtained from commercial sources were incorporated into the DARPin-chelator conjugate under various experimental conditions where pH, reaction time, temperature and concentration were varied. Radiolabelling using Ga^{3+} was found to be optimal with regards to pH in the range 3 – 5.5. The formation of $Ga(OH)_3$ becomes detrimental above that pH. For animal studies of biodistribution where excised organs were measured, the specific radioactivity was less critical, less than 0.1 MBq/ug was used, but for imaging experiments using ^{68}Ga and ^{111}In where sensitivity was critical the SRA was maximized and 1.5 MBq/ug was achieved with 2 ug to be the injected dose. The level of incorporation was determined by HPLC equipped with a radiodetector.

WP4 carried out a total of four *in vivo* biodistribution studies using between 15 to 65 mice for each study. The studies addressed healthy mice and xenograft mice (BALB c) inoculated with HER2 positive (BT474) as well as HER2 negative (MDA MB468) cell lines. The evaluations included both ex-vivo organ biodistribution studies and animal imaging experiments using both PET and SPECT to evaluate the biodistribution of the various protein constructs, and investigated also the relative merits of PET vs SPECT imaging.

PET is intrinsically a more sensitive technique for medical imaging than SPECT, because the radioactivity of the PET nuclides decay within hours, giving rise to a stronger signal than do the radionuclides used for SPECT. The half-life of ^{68}Ga is 68 min, whereas the half-life of ^{111}In is 2.8 days. However, other parameters become important as well, e.g. the time it takes for the injected radiotracer to be cleared and background irradiation to be reduced relative to how much radioactivity remains in the rapidly decaying nuclides. Therefore it cannot be assumed a priori that PET will give better information than SPECT, it has to be measured. A comparison between ^{68}Ga and ^{111}In labelled DARPins was undertaken and the results are summarized as follows. Based on biodistribution data, both ^{68}Ga and ^{111}In-labelled DARPins were able to differentiate between HER2+ve and HER2-ve tumours. The ratio between the uptake in HER2+ve and HER2-ve tumours was higher for the ^{111}In-labelled DARPin (17 times) than for the ^{68}Ga-labelled DARPin (4 times). The uptake in the kidneys was high for both ligands; ~200 times higher than in the HER2-ve tumour and ~20 times higher than in the HER2+ve tumours, for both ligands. This is not atypical for

small recombinant proteins. Based on these results it was decided to pursue SPECT rather than PET for the clinical trial.

In a similar fashion, two differently expressed DARPins were compared with regards to biodistribution using healthy mice as well as xenograft models. The two DARPins differed by a two amino acid mutation, and it was important to assess whether any advantages could be achieved with a GP mutation. The results were similar for the DARPins and it was concluded that the new DARPin without the GP insertion would be used for the clinical trial.

III.5. WP5 – Phase I trial to assess safety and efficacy of quantitative imaging biomarkers in patients

In WP5 the anti-HER2 G3 DARPin was developed as a whole body imaging agent to visualise the extent of HER2+ve cancers in patients. G3 was chosen as it is biologically inert and binds with high affinity at an epitope that does not overlap with trastuzumab.

The G3 HER2 DARPin was first tested in paraffin-embedded tissue sections and compared to an FDA-approved rabbit monoclonal antibody (clone 4B5; Ventana Medical Systems) in tissue microarrays. The data were correlated with HER2 amplification status measured by FISH. Immunohistochemistry demonstrated that G3 reacts with HER2 in breast cancer that was qualified as 3+ on the HercepTest. Importantly, G3 shows no cross-reactivity, whereas 4B5 does, with the panel of normal tissues studied. Thus, the DARPin was able to detect a positive HER2 amplification status with similar sensitivity to FISH, yet with higher specificity than the FDA approved 4B5 antibody.

UCL than set about: (a) finding a suitable format of G3 for clinical application, and (b) developing a robust manufacturing process for clinical application to standards of GMP (good manufacturing practice). Bacterial (*Escherichia coli*) and yeast (*Pichia pastoris*) expression systems for G3 were evaluated in collaboration with UZH. Production of fully functional G3 was achieved in both cases but the yeast-expressed protein was most readily purified. Therefore the GMP compliant bioprocess for the production of G3 was developed using the yeast *P. pastoris* platform. The process involved interfacing fermentation with immobilised metal affinity chromatography (IMAC), using an engineered N-terminal tag on G3 for primary IMAC capture. Initially a hexahistidine (His_6) tag was employed; this is the most commonly used tag for IMAC in laboratory and clinical applications. It has been reported that the His_6 tag can enhance liver uptake in comparison with the alternative, a HEHEHE (HE_3) tag (Hofstrom et al 2011). Therefore, UCL investigated whether removal of, or changing the His_6 on the G3 DARPin, would reduce liver uptake and result in a more favourable imaging agent. For this, three formats of G3 with different N-terminal tags were prepared: (i) His_6 tag (ii) HE_3 tag, and (iii) an untagged control. These were subsequently manufactured using the yeast *P. pastoris* platform. For (iii), the His_6 tag was followed by the human rhinovirus (HRV 3C) cleavage site, LEVLFQ*GP to allow subsequent tag removal. Since this system resulted in a remaining N-terminal GP (glycine proline) on the untagged control, a GP was added to the His_6 and HE_3 tagged variants for comparative purposes. In all cases, the G3 was engineered with a unique C-terminal cysteine for subsequent site-specific attachment of a chelator to allow radiolabelling. The final proteins had the following N-terminal tags: (i) His_6 tag - HHHHHHGP, (ii) HE_3 tag - HEHEHEGP, (iii) Untagged - GP.

Next, a process was developed and optimised (UCL, KCL and UU) for attachment of the chelator maleimide-mono-amide DOTA to the C-terminal Cys on G3. The final DOTA-maleimide conjugated G3 DARPin was stable and attachment of the chelator did not reduce the affinity of G3 for HER2 (tested by UZH).

The three DARPins were evaluated for ability to detect HER2+ve tumours in experimental *in vivo* models. Results showed that ^{111}In-HE_3-G3 had significantly lower liver uptake than its His_6-G3 and untagged G3 counterparts. Interestingly, there was no significant difference in liver uptake between His_6-G3 and untagged G3, suggesting that it was the presence HE_3 rather than the absence of His_6 that led to the favourable result. ^{111}In-HE_3-G3 was also the lead product for HER2-positive tumour xenografts localisation; achieving superior tumour-to-blood ratios of > 150:1 at 4 hours after administration and showing clear tumour images with microSPECT/CT (Goldstein et al 2015).

Since the 'GP' was added solely for comparative purposes in earlier experiments, the DNA sequence encoding these amino acids was omitted in the final genetic construct that was taken forward for generation of the lead G3 protein for clinical evaluation.

To test the robustness of conjugation, HE_3-G3 was conjugated to DOTA-maleimide by UCL and UU. It was also evaluated by UZH to confirm HER2 binding affinity and by UU (WP4) to ensure appropriate biodistribution. All tests were satisfactory and UCL proceeded to prepare a GMP-compliant cell bank to generate the clinical product.

The full GMP compliant bioprocess for production of the HE_3-G3 anti-HER2 DARPin has now been established in the yeast *P. pastoris* on a 10 L scale in a bioreactor and implementation of a short production run, harvest 5hr post induction of protein expression, was founded. Using a multifaceted approach of affinity capture, directly from the crude culture broth, size separation and charge dependent purification steps to yield a fully functional homogenous product of regulatory compliant levels of host cell protein and immunogens. The process was then adapted for the incorporation of maleimide-DOTA addition via the free cysteine on the C-terminal end of the DARPin. This yielded a final product of approximately 25mg of fully labelled DOTA-maleimide-HE_3-G3 with an endotoxin level of <0.5EU/mg & HCP <3.5ng/mg.

UCL are now preparing for the forthcoming first in human clinical trial of HE_3-G3 DARPin. Based on pre-clinical experiments (Goldstein et al 2015) and those performed by UU (WP4), the radiotracer has been defined as ^{111}Indium and the imaging modality as SPECT (Single Photon Emission Computed Tomography). The clinical trial design has been broadened to include HER2 +ve patients with gastric/gastro-oesophageal adenocarcinoma; because, similarly to breast carcinoma, >20% of gastric cancers show HER2 over expression and/or amplification, and this percentage increases to 33% in gastroesophageal junction tumours. Thus, as in breast carcinoma, pathologists are now asked to evaluate HER2 status in gastric carcinoma samples and there is great need for whole body imaging to establish the HER2 status of tumour burden. UCL have performed pilot experiments using ^{111}In-HE_3-G3 with the OE-19 HER2 positive gastro-oesophageal junction cancer tumour xenografts. Excellent tumour uptake and favourable tumour:healthy tissue ratios were obtained in this model.

The cGMP G3 DARPin imaging agent developed and manufactured within the IMAGINT project will be taken forward for the first-in-human whole body HER2+ve imaging trial after the termination of IMAGINT and will be appropriately linked and attributed to the IMAGINT project regarding any forthcoming acknowledgements, dissemination and IP.

III.6.WP6 – Data management, integrative Bayesian analysis of data derived from preclinical and clinical studies

Data from WPs 1-5 was processed and analysed in WP6 using innovative statistical methods established for the IMAGINT project

KCL developed statistical methodology and software to address three significant challenges posed by data generated from imaging HER2-HER3 protein interactions, together with integrative analysis of patients' long-term follow-up and clinico-pathological information and high dimensional data such as protein expression ('toponomic') (Schubert W 2014) or gene expression ('transcriptomic').

The first challenge is the high dimensionality which is a characteristic of genome-scale data sets such as protein expression ('toponomic') and gene expression ('transcriptomic') where there exists a mismatch between the number of measurements and the number of patients. This is known as the “problem of dimensionality” since the dimension of the data (that is, the number of measurements, such as the number of proteins or genes which are assayed) greatly exceeds the number of patients. For a single patient tens of thousands of measurements are acquired, whereas the total number of patients is in the hundreds or fewer. High dimensional data can make it difficult to detect and extract meaningful patterns or relationships in the experimental data. Due to the large number of measurements it becomes easy to find what appear to be promising patterns, but the enormous number of possible patterns means that it is inevitable that some associations will occur by chance. KCL have developed a novel statistical model that attempts to mitigate this problem by “compressing” the observed data (Barrett & Coolen 2014a). The high dimensional data are represented by a small number of “latent variables” which attempt to capture the information in a more parsimonious form. The information contained in ten thousand measurements can be represented using just ten or one hundred latent variables, for instance, and unlike methods such as PCA there is no assumption of linearity of the underlying low dimensional structure. This dimension reduction from thousands of original variables to tens of latent variables diminishes the imbalance between measurements and patients and leads to more a robust and accurate analysis of the data (Barrett & Coolen 2014a).

The second challenge is posed by integrative analysis of different types of measurements with different dimensionality. For example, measurements of gene expression or 'toponomic' protein expression data typically number up to tens of thousands per sample. On the other hand, molecular imaging data typically contains up to tens of measurements per patient sample, such as the data sets generated by tissue microarray-based imaging of HER2 and HER3 protein levels and HER2-HER3 dimerisation. The integration of heterogeneous data sources is challenging since the small number of imaging-based measurements could be “drowned out” by the huge number of genome-scale gene or protein expression measurements. It is desirable to combine these data simultaneously since there may be information that is common (or unique) to both datasets, and to do so in a manner than takes into account the different sizes of the datasets. The model that KCL have developed is capable of simultaneously representing multiple sources of data in terms of the same latent variables (Barrett & Coolen 2014a). This integrates information from several datasets while at the same time reducing the dimension of the datasets. Another feature of this model is that it can automatically detect the most appropriate number of latent variables that are required to represent the information in one or more datasets. The model has been implemented in the statistical software platform Matlab (Barrett & Coolen 2014b). Results indicate that the model is capable of detecting and extracting intrinsic low dimensional structure from test datasets. The benefit of a more parsimonious representation of the data is that the accuracy of subsequent statistical analyses is increased (Barrett & Coolen 2014a).

The third challenge is the analysis of time-to-event survival data where the patient cohort may be comprised of two or more unknown ('latent') groups with different associations for each group. In particular, we investigated whether there was any evidence of multiple latent groups in response to HER-targeted treatments in breast and colorectal cancers. The presence of two or more groups then guides the search for a biomarker for treatment response. In order to model and probe for putative latent groups, survival analysis algorithms have been developed for the determination of the optimal characterisation of a cohort from time-to-event survival data. The algorithms are able to identify and characterise cohort heterogeneity. As the analysis has been designed with the assumption that there may be correlations between different risks at the cohort level, the algorithms operate effectively even in the presence of heterogeneity-induced informative censoring

and is therefore also suitable for modelling patient follow-up with additional risk types (for example, co-morbidity and associated interventions). These novel algorithms are able to determine whether a cohort is comprised of a number of sub-groups, or latent classes, and if so, how the classes differ in terms of their frailty and their associations to different measured quantities. Having determined the optimal characterisation of a cohort, crude and decontaminated (for informative censoring) survival curves are generated and patients can be retrospectively assigned to their most probable latent class. The corresponding hazard rates, 95% confidence intervals, and p-values are also generated.

The Bayesian latent class proportional hazards model was applied to HER2/HER3 imaging data from two patients cohorts (Ng et al 2015, Barber et al 2015a). HER2 and HER3 protein levels and HER2-HER3 interactions were imaged (Work Package 2) on tissue microarrays of primary colorectal tumour samples, and the evidence for the presence of latent groups for HER-targeted treatment response was evaluated (Ng et al 2015). The detection of putative latent groups and the presence of a biomarker for HER-targeted treatment response related to HER2-HER3 signalling is being studied in a second, validation set of tissue microarrays.

FLS has implemented the following data management tools: (1) A multivariate Biomarker prediction suite consisting of a panel of diverse machine learning algorithms including partial least squares discriminant analysis, support vector machine, k-nearest neighbours, naïve bayes, logistic regression, random forest, CN2, classification trees, supervised principal component analysis and other non-linear projection methods. (2) A set of different model validation strategies including n-fold cross-validation, leave-one-out validation, and different re-sampling schemes for calibration and test samples.

Each of these statistical learning approaches has its strengths and weaknesses, and the objective of the FLS Biomarker prediction suite is to apply all methods in parallel and use a supervisor algorithm to integrate and benchmark the predictions of the individual prediction models. This approach is well known in the machine learning community and is called “Panel of Experts” approach. The prediction suite has been implemented and is functional.

The Bayesian Generalized Latent Variable Model (BGLVM) developed at KCL by the lab of Ton Coolen has been integrated into the analysis pipeline as an optional preliminary step. This program allows reducing the number of variables in the input dataset. These latent variables can then be analysed using the analysis pipeline. This allows a faster analysis of high dimensional datasets generated by imaging or NGS methods for example.

In addition, FLS has compiled a set of test cases of diverse numbers of samples/patients and variables (n x p) including synthetic data, transcriptomics, proteomics and metabolomics data, clinical data like demographics, medication, morbidity, standard haematology, biochemistry and urinalysis parameters. Features extracted from imaging procedures will also be included. The objective of this benchmarking approach is to obtain an as unbiased as possible estimate of the predictive performance of the newly developed “Bayesian Latent Variable Model”.

III.7.References

Barber PR, Tullis ID, Pierce GP, Newman RG, Prentice J, Rowley MI, et al (2013) The Gray Institute 'open' high-content, fluorescence lifetime microscopes. J Microsc 251:154-67.

Barber PR, Rowley MI, Lawler K, Weitsman G, Ng T, Vojnovic B (2015a) A Processing Pipeline for Detecting Protein Interactions in Histological Tissue. MIUA (in press).

Barber PR, Tullis IDC, Rowley MI, Lawler K, Weitsman G, Ng T and Vojnovic B (2015b). FLIM Histology: Detecting protein interactions in human tissue by FRET. MMC (in press).

Barber PR, Tullis IDC, Rowley MI, Martins CD, Weitsman G, Lawler K, et al (2014) The Gray Institute open microscopes applied to radiobiology and protein interaction studies. Proceedings of the SPIE. *In press*.

Barrett, JE, Coolen ACC. (2014a) 'Dimensionality Detection and Integration of Multiple Data Sources via the GP-LVM'. arXiv:1307.0323.

Barrett, JE, Coolen ACC. (2014b). `Gaussian process Regression for Survival Data with Competing Risks'. arXiv:1312.1591.

Barrett, JE, Coolen ACC (2014c). `Dimensionality Reduction for Survival Data via the Gaussian process Latent Variable Model'. arXiv:1406.0812.

Barrett JE. (2014). GPLVM-WPHM. Matlab Exchange. http://www.mathworks.com/matlabcentral/fileexchange/48565-gplvm-wphm.

Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A (2013) Liquid biopsy: monitoring cancer-genetics in the blood. Nature reviews Clinical oncology 10:472-84.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486:346-52.

Dress AWM, Lokot T, Pustylnikov LD, Schubert W (2005) Poisson numbers and poisson distributions in subset surprisology. Ann Comb 8: 473-85.

Dress AWM, Lokot T, Schubert W, Serocka P: Two theorems about similarity maps (2008) Ann Comb 12: 279-90.

Friedenberger M, Bode M, Krusche A, Schubert W (2007) Fluorescence detection of protein clusters in individual cells and tissue sections by using toponome imaging system: sample preparation and measuring procedures. Nat Protoc 2:2285-94.

Goff LA, Davila J, Swerdel MR, Moore JC, Cohen RI, Wu H, et al (2009) Ago2 Immunoprecipitation Identifies Predicted MicroRNAs in Human Embryonic Stem Cells and Neural Precursors. PLoS One. 4:e7192.

Goldstein R, Sosabowski J, Livanos M, Leyton J, Vigor K, Bhavsar G, Nagy-Davidescu G, Rashid M, Miranda E, Yeung J, Tolner B, Pluckthun A, Mather S, Meyer T, Chester K (2015) Development of the designed ankyrin repeat protein (DARPin) G3 for HER2 molecular imaging. Eur J Nucl Med Mol Imaging 42: 288-301.

Gu GJ, Friedman M, Jost C, Johnsson K, Kamali-Moghaddam M, Plückthun A, Landegren U, and Söderberg, O (2012) Protein tag-mediated conjugation of oligonucleotides to recombinant affinity binders for proximity ligation N. Biotechnol 30:144-152.

Hillert R, Gieseler A, Krusche A, Humme D, Röwert-Huber HJ, Sterry W, Walden P, Schubert W. Large molecular systems landscape uncovers T cell trapping in human skin cancer. *In review.*

Hofstrom C, Orlova A, Altai M, Wangsell F, Graslund T, Tolmachev V (2011) Use of a HEHEHE purification tag instead of a hexahistidine tag improves biodistribution of affibody molecules site-specifically labeled with (99m)Tc, (111)In, and (125)I. J Med Chem 54:3817-26.

Ilgen et al (2014) STED Super-Resolution Microscopy of Clinical Paraffin-Embedded Human Rectal Cancer Tissue PloS One.

Jiao Y, Lawler K, Patel GS, Purushotham A, et al (2011) DART: Denoising Algorithm based on Relevance network Topology improves molecular pathway activity inference. BMC Bioinformatics, 12(1), 403. doi:10.1186/1471-2105-12-403.

Jost C, Schilling J, Tamaskovic R, Schwill M, Honegger A, and Plückthun A (2013) Structural Basis for Eliciting a Cytotoxic Effect in HER2-Overexpressing Cancer Cells via Binding to the Extracellular Domain of HER2. Structure 21:1979-91.

Leary A, Evans A, Johnston SR, A'Hern R, Bliss JM, Sahoo R, Detre S, Haynes BP, Hills M, Harper-Wynne C, Bundred N, Coombes G, Smith I, Dowsett M (2014). Antiproliferative Effect of Lapatinib in HER2-Positive and HER2-Negative/HER3-High Breast Cancer: Results of the Presurgical Randomized MAPLE Trial (CRUK E/06/039).

Maciej P and Simpson TI (2014) miRNAtap: miRNAtap: microRNA Targets - Aggregated Predictions. R package version 1.0.0.

Maughan TS, Adams RA, Smith CG, Meade AM, Seymour MT, Wilson RH, et al (2011) Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. Lancet 377:2103-14.

Ng T, Weitsman G, Barber P, Lawler K, Rowley M, Elshtein G, et al (2015) Use of FLIM histology-based HER2-HER3 heterodimer quantification and a Bayesian latent class proportional hazards model to predict cetuximab response in the COIN and new EPOC trials. J Clin Oncol 33 (suppl; abstr e14535).

Persson H, Kvist A, Rego N, Staaf J, Vallon-Christersson J, Luts L, et al (2011). Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. Cancer Res 71:78–86.

Primrose J, Falk S, Finch-Jones M, Valle J, O'Reilly D, Siriwardena A, et al (2014) Systemic chemotherapy with or without cetuximab in patients with resectable colorectal liver metastasis: the New EPOC randomised controlled trial. The Lancet Oncology 15:601-11.

Schubert W (2014) Systematic, spatial imaging of large multimolecular assemblies and the emerging principles of supramolecular order in biological systems. J Molecular Recognition 27:3–18.

Schubert W (2015) Toponomics in Bioanalytics. Wiley. *In press.*

Schubert W (2015) Advances in Toponomics Drug Discovery: Imaging Cycler Microscopy Correctly Predicts a Therapy Method of Amyotrophic Lateral Sclerosis. Cytometry Part A. Doi: 10.1002/cyto.a.22671.

Schubert W, Gieseler A, Krusche A, Hillert R (2009) Toponome mapping in prostate cancer: detection of 2000 cell surface protein clusters in a single tissue section and cell type specific annotation by using a three symbol code. J Proteome Res 8:2696-707.

Schubert W, Gieseler A, Krusche A, Serocka P, Hillert R (2012) Next-generation biomarkers based on 100-parameter functional super-resolution microscopy TIS. N Biotechnol 29:599-610.

Schubert W, Bonnekoh B, Pommer AJ, Philipsen L, Böckelmann R, Malykh Y, Gollnick H, Friedenberger M, Bode M, Dress AW (2006) Analyzing proteome topology and function by automated multidimensional fluorescence microscopy Nat Biotechnol 24:1270-8.

Tamaskovic R, Simon M, Stefan N, Schwill M, Plückthun A (2012) Designed Ankyrin Repeat Proteins (DARPins): from research to therapy. Methods Enzymol 503:101-134.

Tao JJ, Castel P, Radosevic-Robin N, Elkabets M, Auricchio N, Aceto N, et al (2014) Antagonism of EGFR and HER3 enhances the response to inhibitors of the PI3K-Akt pathway in triple-negative breast cancer. Sci Signal 7:ra29.

Thery C (2011) Exosomes: secreted vesicles and intercellular communications. F1000 biology reports 3:15.

Weitsman, Paul R Barber, Lan K Nguyen, Katherine Lawler, Natalie Woodman, Sarah Pinder, et al. HER2-HER3 dimer quantification by FRET-FLIM predicts breast cancer metastatic relapse independently of HER2 expression. Clinical Cancer Research *submitted*.

Wyman SK, Parkin RK, Mitchell PS, Fritz BR, O'Briant K, Godwin AK, et al (2009) Repertoire of microRNAs in Epithelial Ovarian Cancer as Determined by Next Generation Sequencing of Small RNA cDNA Libraries. PLoS ONE 4(4): e5311.

Yan J, Gu GJ, Jost C, Hammond M, Plückthun A, Landegren U, and Kamali-Moghaddam M (2014) A Universal Approach to Prepare Reagents for DNA-Assisted Protein Analysis. PLoS One 9, e108061.

III.8.WP7 – Dissemination, IPR and ethical issues

The IMAGINT webpage (www.imagint.eu) was established at the beginning of the project and has since then been continuously updated with information about results and dissemination from the IMAGINT project.

The webpage has evolved with new sections about such as Public/Patient Engagement and information of the 'Independent Cancer Patients' Voice (ICPV).In addition, the internal collaborative portal to exchange confidential documents and is now in use by all partners with 'viewing' permission. Editing is controlled by the IMAGINT webmaster.

The IMAGINT consortium has been very active in disseminating the results and more than 82 oral or poster presentations have been performed at international conferences and meetings. Also, 28 papers have so far been published in international peer-review journals. IMAGINT has also established a significant number of collaborations with other important international initiatives and projects for disseminating and exploiting results.

At the beginning of the project a Project Ethic Committee (PEC) was set up composed of a representative from each participating country. The PEC played an essential role during the entire course of the IMAGINT project. The PEC provided ethical guidance to the consortium as well as ensuring that all experimental work is carried out within national and European regulations. The PEC ensured that no experimental studies foreseen by any partner will start before prior approval of all relevant local ethics committees. The partners are all continuing to ensure that the experimental work, animal welfare, human tissue handling, GMP protein production and clinical trial progression is carried out in compliance with national and European regulations and that any ethical issues are being addressed, specifically:

●All animal experiments at UU were adhered to the following laws and regulations: Animal Welfare Act (SFS 1988:534), Swedish Animal Welfare Ordinance(SFS 1988:539), Code of regulations of the Swedish Board of Agriculture SJVFS 2012:26 (L150) SJVFS 2010:2, Directive 2010/63/EU.

●All animal experiments at UCL were performed in accordance with the UK Home Office Animals Scientific Procedures Act 1986, and United Kingdom Co-ordinating Committee on Cancer Research Guidelines for the Welfare and Use of Animals in Cancer Research (Workman et al 2010) and approved by the UCL Animal Welfare and Ethical Review Body (AWERB) under project license PPL 70/7309

●Human tissues were provided by the King's Health Partners Cancer Biobank via Dr Cheryl Gillett, IMAGINT PEC member. Dr Gillett is also the HTA Designated Individual responsible for ensuring that biobank-associated personnel are aware of and comply with the Quality Management System.

●The KHP Cancer Biobank meets the terms of all current laws and ethical guidelines associated with tissue collection and use for research. This includes having a Human Tissue Authority licence for research purposes (licence number 12121) and approval from the National Research Ethics Service (NRES) to collect, store and issue biological samples and associated clinic-pathological data (REC number: 12-EE-0493), subject to specified conditions.

- The KHP Cancer Biobank Management Committee oversees all aspects of governance and management. This Committee includes senior doctors, scientists, independent and patient representatives. The Management Committee reports to Guy's and St Thomas' NHS Foundation Trust Chief Executive via the Trust Risk and Quality Committee (TRaQ).
- The KHP Cancer Biobank Quality Management System sets out the general principles on quality of work that all biobank- associated personnel are expected to uphold. This includes a framework that enables personnel to work in an environment that is compliant with current legislation, regulations, guidelines and standards.

●Data Management and statistics are compliant with the Good Clinical Practice (GCP) –ICH E2A & ICH E6. The identity of the patients is kept confidential at the investigator center or hospital and patient are anonymized with a code list and ID.

After IMAGINT became an operative running project, an Intellectual Property Committee (IPC) was set up in order to monitor and approve all dissemination activities and also to deal with all Industrial and Intellectual Property Rights issues according to the Consortium Agreement. An IPR manager or Chairman of the IPC committee was appointed in the first period of the project.

III.9.WP8 – Management

The role of the management team, composed by the project coordinator (UCL) and the project management partner (NOVAMEN) was to ensure that consortium activities are performed in accordance to the EC expectations in terms of content and timescale.

The Management Manual of IMAGINT was produced at Month 3 of the project and gave the main decision-making rules and the main procedures regarding the management of the project (reports, deliverables, communication, etc. More particularly, this deliverable D8.18 included: i) a description of the project management structure, organization and procedures; ii) the composition, role and responsibilities of the different Project Committees as well as the responsibilities and duties of the Project Coordinator, Work Package Leaders and partners; iii) the general rules for meetings and communication; iv) the internal procedures for reporting and deliverables; v) the major milestones and project deliverables to be produced during the project.

The operational management organization was installed since the beginning of the project. Strategic, management and Steering Committee meetings were taking place every year to monitor the execution of the project, the work progress and to define the project strategies. They also hosted members of the Project Ethics Committee (PEC) and the Intellectual Property Committee (IPC). In addition to these Steering Committee meetings, regularly meetings and conference calls were organized on the WP level to ensure effective communication and coordination of the scientific work. Work progress and results were reported to the European Commission by the consortium in the Periodic Activity Reports that were submitted at Months 18, 36 and 48. In addition to these reports IMAGINT consortium submitted 17 deliverables.

IV. Potential impact and main dissemination activities and exploitation of results

As described in the DOW, IMAGINT project was aiming to tackle the following challenges:

- Supporting novel scientific and technological breakthroughs
- Improving health and quality of life
- Enhancing the research capacities at the European level and strengthening the competitiveness of the European industry; structuring the ERA

Regarding **novel scientific and technological breakthroughs**, IMAGINT has completely fulfilled the planned expected outputs:

The consortium has been able to develop robust tools to measure different HER dimers in human tissues and even shown the clinical utility of this tool by the HER2-HER3 heterodimer assay, used in a cohort of patients with breast cancer. This tool showed superior prognostic value to standard HER2 IHC assay and enabled greater insight in the understanding of the real complexes of HER2 present in specific tumours. The consortium has also developed a range of practical and innovative mathematical and informatics tools, including artificial neural network methods, to analyse multivariate data and help to determine novel biomarkers to classify disease. Indeed, these informatics tools have already shown clinical use in analysing HER heterodimers in patients with triple-negative breast cancer treated with anti-EGFR antibodies, validating a new mechanism of resistance to anti-EGFR therapy supporting rational selection of therapy. New imaging and mathematical tools to identify the protein clusters have also been developed, to study protein function in terms of activity and spatiotemporal location leading to discovery of specific breast cancer protein clusters with potential to be informative for future diagnosis and therapy. Development of cutting edge tools and technology for the detection, isolation and functional characterisation of complexes of interacting molecules (RNA/protein) has led to discovery of proteins and miRNAs that are differentially associated with drug-treated breast cancer cells. Bioinformatics analysis of gene expression for the differentially detected proteins has revealed an association with time to progression in breast cancer. Thus the combined RNA/protein and informatics tools have provided an untapped source of new biomarkers and shown future potential to uncover new biomarkers in patients. The consortium has also succeeded in development of robust tools for quantitative clinical imaging of the extent and location of metastatic disease. These tools have been fully evaluated and are poised to enter the clinic to be trialled as an Investigative Medicinal Product for whole body imaging of HER2+ve disease in patients with advanced HER2+ve breast and gastric or gastro-oesophageal junction cancer. Imaging HER2+ve disease in these patients is predicted to provide important and hitherto unprecedented clinical diagnostic information to support optimal management of current anti-HER2 therapies.

In addition, innovative statistical methodology and software have been developed as follows: (i) for survival analysis for characterisation of cohorts with latent heterogeneity and in the presence of informative censoring. This software performs generic latent class hazard rate modelling and optimal cohort characterisation using Bayesian model selection formalism by detection and optimal characterisation of latent heterogeneity in time-to-event type survival data such as clinical trial data. This provides effective survival analysis in the presence of informative censoring, and retrospective latent class allocation using Bayesian arguments aiding identification of informative covariates. (ii) Methodology and software for Gaussian process latent variable Weibull proportional hazards model (GPLVM-WPHM) were developed providing improved accuracy of statistical analysis of high dimensional biomedical data. (iii) TMA image, a package for the statistical software R, was developed to facilitate the curation and analysis of tissue microarray (TMA) imaging data after image processing. This facilitates the systematic inspection and manipulation of imaging parameters and image-derived values, in particular from multiple tissue microarrays and large numbers of samples. Continued improvements to implementation and features are on-going alongside tissue microarray-based imaging research.

It is clear that many of the new tools developed by IMAGINT will have substantial impact in terms of **health improvement and quality of life** of patients,. For example, the HER heterodimer assays have already shown efficacy in supporting patient stratification. This is predicted to eventually lead to more personalised therapies with related improvements wellbeing and health costs. The whole body imaging of HER2+ve tumours in patients with breast and gastric or gastro-oesophageal junction cancer is expected to support new paradigms in the management of these diseases; increasing treatment efficacy and decreasing screening costs. In addition, the potentially exciting new markers discovered by the innovative tools for (i) identification of complexes of interacting molecules (RNA/protein), and (ii) identification of breast cancer-specific protein clusters enabling identification of related protein networks, have provided a platform to increase understanding of cancer, provide new markers for diagnosis/prognosis and patient stratification; leading potentially to more personalised treatment and supporting more rational selection of therapy.

The 5.68M€ funding provided by the European Commission to the IMAGINT consortium has been able **to strengthen the European Competitiveness at different scale**: In particular IMAGINT has achieved this by developing new methods – including cGMP protocols - and designing new techniques for the fast growing world market of biotechnology. Moreover, due to the multidisciplinary and transversal approach of IMAGINT, consortium partners have gained visibility regarding their cutting approach in the cancer fields. Different exchanges regarding the use of DARPIN have been conducted during the many international conferences where IMAGINT partners were present as invited or high profile speakers. This is evidenced by the list of dissemination activities in the report on societal implications.

We describe below the variety of outcomes that could be exploitable at different levels, including potential licensing opportunities and knowledge transfer of the methods developed under IMAGINT.

UCL/UZH/UU

A major potential exploitable outcome of the IMAGINT consortium is the novel DARPIN mediated HER2 imaging modality being developed by UCL, UZH and UU. This could benefit patients with advanced HER2 positive breast cancer, by staging their disease, assessing the indication for anti-HER2 therapy and characterising the response to anti-HER2 therapy. It could also have benefits for early breast cancer through improved monitoring and detection of early relapse. The product will also have applications in diagnosing and staging of HER2 positive gastric cancer. The market for HER2 therapy is considerable with annual sales of Herceptin (marketed by Roche/Genentech) of over $ 6.4 billion (Genentech Annual Report 2014). The HER2 market it set to continue to grow as Kadcyla, the HER2 targeted antibody drug conjugate, is now approved in the major pharmaceutical markets with sales expected to peak in the $ billion range. A reagent that helps facilitate and expand this market through detection of relapse, and earlier detection of breast and gastric cancer will be of significant commercial value.

UCL, UZH and UU have developed the following exploitable package of results:

1) GMP-compliant manufacturing processes and associated know-how. One of the main exploitable outputs of the IMAGINT project has been to develop labelling procedures that are fully GMP-compliant and can routinely be used in a hospital setting; having the possibility to make it truly accessible to hospitals all across Europe.
2) A Clinical data package to be developed in partnership with the CRUK Centre for Drug Development and UCL. CRUK have agreed to sponsor and support the follow-on clinical study of the DARPIN at UCL.

With the support of Cancer Research Technology (the tech transfer office of CRUK) and UCL Business we are currently in the early stages of discussions with a potential partner for this program. These have been very encouraging. We see this as an out licensing opportunity and would out license a package of GMP know-how and clinical data to an interested licensor. As positive clinical data arises from the follow on clinical study our intention would be to step up conversations with this partner and also engage with other parties who are working with DARPin

technology. Our aim would be to partner this with a company who has the capacity and experience in bringing complex imaging biologics to the market. It would also be a benefit if they had experience and expertise of working with DARPin technologies. Currently Molecular Partners, Allergan and Johnson and Johnson are active in this technology area. **The manufacturing know-how, coupled with the clinical validation and accompanying data package make this an attractive and high-value licensing opportunity**.

KCL

The mathematical methods for biomarker dimension reduction and regression in the presence of cohort and/or disease heterogeneity developed within IMAGINT are now being applied more widely to medical data (e.g. breast cancer data from the Swedish AMORIS study, paper just accepted by BMC Cancer; lung cancer data from the TOPICAL trial, in collaboration with UCL), and we are in discussion with the statistics team of Glaxo-Smith-Kline UK who have expressed their interest in working with KCL to **develop the methodology further** and apply it to clinical data sets at GSK.

TNL

The direct individual finding of breast cancer driving protein complexes *in situ* will – if systematically applied – improve the chances to treat breast cancer more efficiently, since the ICM based detection of lead proteins controlling these complexes provide highly specific novel target proteins. On the long term, progressive implementation of the ICM technology in breast cancer specific diagnostics and therapy planning will help to improve the healthcare system in the field of breast cancer by reducing the number of patients not responding to current therapy strategies.

On the basis of the European and international penetration of the ICM technology, TNL is presently actively establishing the concept/platform for company growth to enhance the market penetration. TNL has active relationships to big Pharma and our technology brings their pharma partners is an extreme degree of error elimination (Schubert Cytometry 2015, an IMAGINT paper), thereby overcoming current drawbacks of the so called precision medicine (Wang A.Z. Precision cancer medicine: hype or hope ? Science translational medicine 7 (306) pp306, 2015). The dissemination of the technology itself is successfully ongoing by a German robotic company, subcontractor of TNL (USA, Breast cancer, established onsite) and UK (colon cancer; diabetes, established onsite). In addition, a new ALS Approach to more effective therapy of ALS is actively planned based on Schubert (Cytometry Part A, 87A: 696-703, 2015).

MPG

MPG has demonstrated that the nanoscale distribution of HER2 in biobanked resected tissues can be visualized and quantified using STED super-resolution microscopy. Robust experimental protocols were developed that principally could be transferred into a clinical setting. This could benefit patients with HER2 positive breast cancer by staging their disease more precisely than with established approaches. Moreover, this technology could be used to re-analyze stored breast cancer tissues in repositories to retroactively assess the specific response to anti-HER2 therapy. Establishing super-resolution microscopy in a clinical setting for diagnostic purposes is a potentially significant, yet currently unexplored market.

We are currently working together with clinicians from the University Medical Center of Goettingen to **extend the developed technology also to other biobanked tissues**. We intend to explore **potential licensing options**, possibly together with companies selling **STED-microscopes**. Currently, Leica Microsystem, Picoquant and Abberior Instruments are selling STED microscopes and are thus potential partners.

INO

The task of INO in the IMAGINT project was to identify new biomarkers implicated in breast cancer through the isolation and characterization of the protein-protein and RNA-protein interactions in AGO2 containing complexes. The successful development of tools to achieve this during IMAGINT has allowed INO to optimize the endogenous cellular complexes isolation yield and therefore, to obtain protein and RNA samples. This knowledge has been transferred to cellular cultures more complicated such as primary cell cultures. Furthermore, the application of those tools for identification of specific miRNAs and proteins implicated in the anti-HER2 therapies, allows INO to propose the generation of new cell models based on these newly identified molecules that would be further used for the screening of molecules that can further developed into drugs for the treatment of breast cancer or other diseases. Moreover this expertise is expected to raise the profile of INO and lead to increase business opportunities, in particular to **advertise INOs Protein Interaction Service**.

FLS

The **analysis pipeline** developed by FLS is currently in use for the analysis of data generated at FLS (genomic data, biomarker concentrations...) and build predictive models to discriminate different populations of patients. In the near future, it is planned to commercialize a **new user friendly software** including functionalities already present in the current version developed during the IMAGINT project. The target of such a software would be biologists with a basic knowledge in statistics who want to analyse large datasets generated by their experiments and linked to clinical outcomes.

The software in its current state is able to generate results from a large variety of data ranging from biomarker concentrations to gene expression levels (as demonstrated in the IMAGINT project). It is planned to improve the support of data generated by next generation sequencing instruments (Illumina).

Additionally, FLS plans to propose services to **customize the software** in order to allow the analysis of new types of data and the generation of new results based on needs specified by the customer. If the commercialization of such a software is deemed to be not profitable enough for FLS, a finalized version will be made available to the public and scientists on the FLS website.