ESMO Research Research Fellowship
(April 2021 – April 2023)

Dr. Long Viet Nguyen

FINAL REPORT

Host Institute: **CRUK Cambridge Institute, University of Cambridge (Cambridge, United Kingdom)**
Mentor: **Professor Carlos Caldas**
Project title: **Investigating the molecular mechanisms that regulate human breast cancer stem cell function in response to systemic therapy**

Home Institute: **Princess Margaret Cancer Centre, University Health Network, University of Toronto (Toronto, Canada)**

Introduction

Breast cancer heterogeneity. Breast cancers are a heterogeneous collection of diseases with different molecular features and treatment paradigms. Initial gene expression profiling studies described different "intrinsic subtypes": luminal A, luminal B, HER2-enriched, and basal-like[1]. Recently, the Caldas group identified 10 novel subgroups of breast cancer (IntClust 1-10) with distinct clinical outcomes based on copy-number aberrations (CNAs) and gene expression profiles[2]. While much effort has been made to characterize the molecular features of breast cancer, little is known about how these subtypes differ in their functional growth properties.

Cancer stem cells (CSCs) and cellular barcoding. CSCs are relatively rare tumour cells with the potential to regenerate, through the process of self-renewal, to produce daughter cells with similar regenerative potential[3]. CSC activity can be assayed in xenograft models that detect the ability of a transplanted cell to initiate new tumour formation, so-called "tumour-initiating cells" (TICs)[3]. Cellular barcoding is one clonal tracking method that can be used to measure the frequency of TICs and track their progeny *in vivo*. This method introduces unique, heritable, random DNA sequences (barcodes) into the genome of individual cells by viral transduction[4]. Variations of this approach use CRISPR/Cas9 technology to incorporate short insertions or deletions at specific target sites to generate these unique barcodes[5]. When coupled with PCR amplification and massively parallel sequencing (MPS) technology, cellular barcoding can be a powerful quantitative method to analyze the composition and growth dynamics of complex polyclonal samples. This approach differs from genomic and transcriptomic studies that depict tumour clonal evolution based on genotype, and not functional activity.

Single cell RNA sequencing (scRNAseq). Applied across tumour types, bioinformatic analysis of gene expression profiles generated from scRNAseq analysis reveal clusters of different cell types. Some cell clusters have high expression of proliferation genes or epithelial-to-mesenchymal transition, and thus may be primed for disease progression or metastasis[6]. However, this type of data has not been linked directly to quantitative measures of clonal growth. Therefore, I propose to create an expressible barcoding system compatible with scRNAseq in order to derive both gene expression and clonal growth information simultaneously on a single cell level.

Rationale and Aim

Rationale – While much effort has been made to characterize the molecular features of breast cancer, little is known about how these subtypes differ in their functional cancer stem cell (CSC) properties, and how these properties change following systemic therapy. Therefore, I propose to create an expressible barcoding system that will allow me to track individual clones as they grow *in vivo.* This system will also compatible with single cell RNA sequencing (scRNAseq) in order to derive both gene expression and clonal growth information simultaneously on a single-cell level.

AIM 1: To characterize the functional properties of CSCs in aggressive subtypes of human breast cancer before and after systemic therapy using a novel sequential cellular barcoding approach.

AIM 2: To elucidate and validate molecular pathways that regulate the functional activity of human breast CSCs using a combined cellular barcoding and scRNAseq approach.

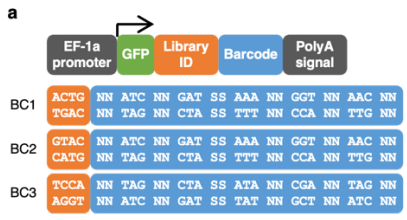
Experimental design

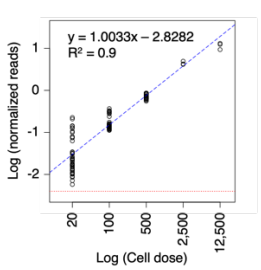
For this project, I relied on the valuable patient-derived tumour xenograft (PDTX) biobank established by the Caldas lab. These PDTX models were barcode labelled using the clonal tracking approach described above (and in Figure 1 below), and then injected subcutaneously into immunodeficient NSG mice in order to observe clonal growth with and without chemotherapy treatment (Carboplatin and Paclitaxel). Barcode clone analysis was performed using DNA-based amplicon sequencing on the Illumina MiSeq, as well as scRNAseq using the 10X Genomics platform.

Results, Conclusions and Future Perspectives

Result 1: Three lentiviral barcode libraries were constructed (Figure 1A). Barcoded cells were analyzed by DNA-based amplicon sequencing on the Illumina MiSeq, and clones could be quantitatively analyzed by correlating normal read count from sequencing ("fractional read value") with clone size in absolute cell numbers (Figure 1B). The sensitivity of clone detection using this approach was 87% for 10-cell clones, 99% for 20-cell clones, and 100% for clones containing 50 cells or more.

FIGURE 1:



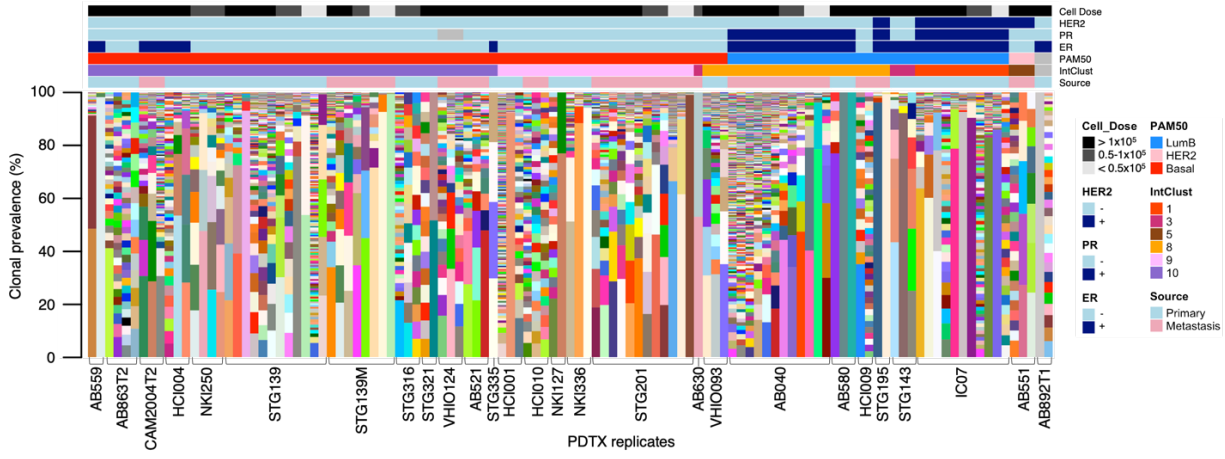


c

Spike-in control (cells)	Detected by sequencing
12,500	24 / 24 (100%)
10,000	30 / 30 (100%)
2,500	24 / 24 (100%)
1,000	30 / 30 (100%)
500	308 / 308 (100%)
100	308 / 308 (100%)
50	30 / 30 (100%)
20	275 / 278 (99%)
10	26 / 30 (87%)

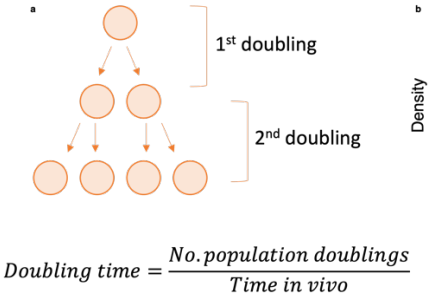
Result 2: A total of 113 tumours, from 27 different PDTX models were investigated using this clonal tracking cellular barcoding approach. The models chosen span many molecular subtypes of human breast cancer. In total, the clonal growth information from 20,000 barcode clones was captured showing that there is significant functional clonal heterogeneity across all subtypes and all models studied (Figure 2).

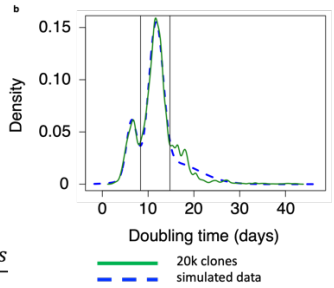
FIGURE 2:



Result 3: In collaboration with my colleagues at the Caldas lab in Cambridge, we established a framework with which to understand the functional heterogeneity of malignant clones in these human breast tumour models. This involves calculating the average population doubling time of the clones as shown in Figure 3A. Figure 3B shows the actual data (green line) and simulated data from mathematical modelling (blue dashed line), revealing that the doubling time for these clones separate into 3 distinct Gaussian distributions, allowing us to classify clones into those with fast, medium, and slow doubling times (Figure 3C). By examining clones in this way, this reveals subtype-specific differences (Figure 4). IntClust subtypes 8, 9 and 10 have the highest proportion of fast doubling time clones, whereas IntClust 1, 3, and 5 have almost exclusively slow doubling time clones. IntClust 10 consists mainly of triple-negative breast cancers, IntClust 5 consists mainly of HER2+ breast cancers, IntClust 1 and 9 are ER+ breast cancers with a higher risk of late-relapse, whereas IntClust 3 and 8 are ER+ breast cancers with a lower risk of late-relapse.

FIGURE 3:

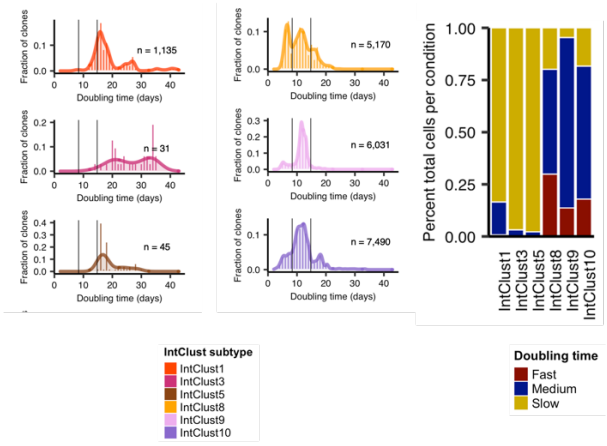




c

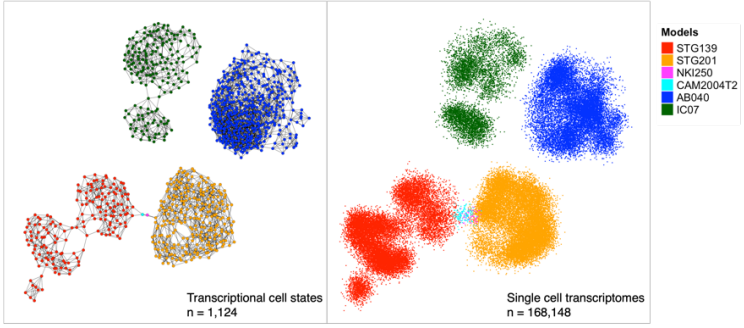
	Fast	Medium	Slow
Mean doubling time (days)	6.4	11.7	15.4
Range doubling time (days)	≤ 8	> 8 to 15	> 15
No. clones (%)	3,723 (18.7%)	12,523 (62.8%)	3.697 (18.5%)

FIGURE 4:



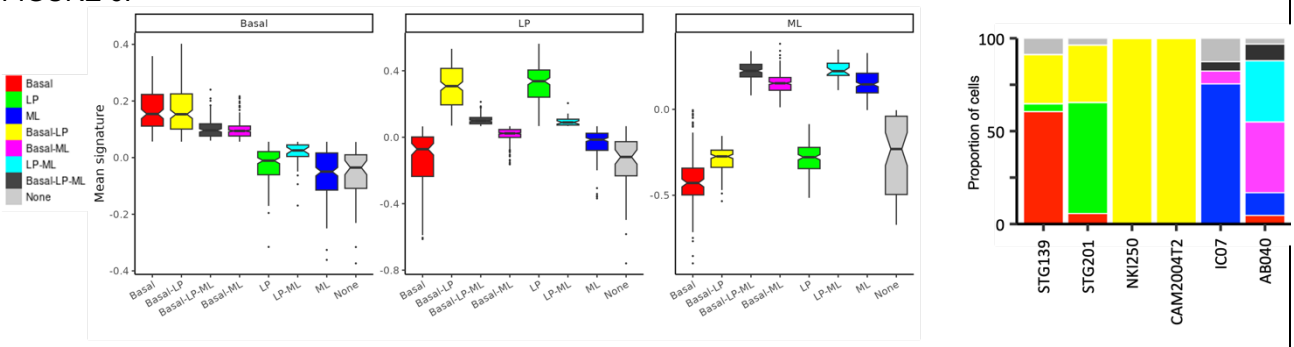
Result 4: scRNAseq was performed on 19 tumours, from 6 different PDTX models, including 4 triple-negative breast cancer models (STG139, STG201, NKI250, CAM2004T2), and 2 ER+/luminal B breast cancer models (AB040, IC07). In total ~170,000 single cell transcriptomic profiles passed quality-control metrics and were included in this dataset. Using the R package “metacell”, these cells were found to consist of 1,124 distinct transcriptional cell states (Figure 5).

FIGURE 5:



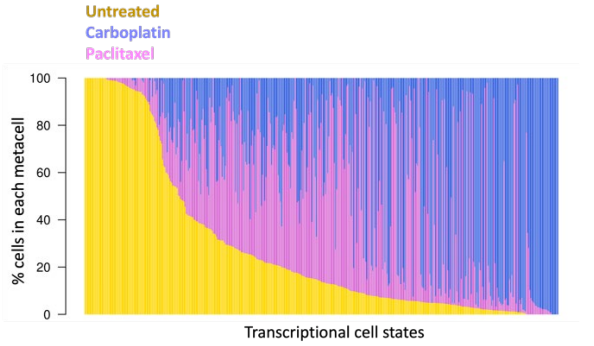
Result 5: In collaboration with my colleages at the Caldas lab in Cambridge, we established a framework with which to understand the transcriptional heterogeneity of malignant clones in these human breast tumour models. This involves applying molecular signatures derived from normal epithelial cell types (basal/myoepithelial, luminal progenitor/LP, and mature luminal/ML) to the transcriptional cell states identified in the scRNAseq dataset. Eight cell phenotypes were found: pure basal, pure LP, pure ML, basal-LP, basal-ML, LP-ML, basal-LP-ML, and none. The four triple-negative breast cancer models show cell phenotypes consistent with a pure basal, pure LP, or basal-LP signature, whereas the two ER+/luminal B breast cancer models show cell phenotypes consistent with a pure ML or basal-ML signature (Figure 6).

FIGURE 6:



Result 6: While work is still ongoing to understand how these different malignant clones respond to chemotherapy treatment, preliminary results are shown in Figure 7. This shows that there are clones that can resist chemotherapy treatment and survive due to their ability to adapt their transcriptional cell state. Further analysis will reveal the key molecular pathways that these clones rely on to resist chemotherapy.

FIGURE 7:



List of Publications and Presentations Resulting from the Translational Research Project "Investigating the molecular mechanisms that regulate human breast cancer stem cell function in response to systemic therapy"

Publications:

1. **Nguyen LV**, Caldas C. Functional genomics approaches to improve pre-clinical drug screening and biomarker discovery. ***EMBO Molecular Medicine*** 2021, PMID: 34254730.
2. **Nguyen LV**, Eyal-Lubling Y, Guerrero-Romero D, Chin S-F, Manzano Garcia R, Lui A, Lerda G, Bardwell H, Kania K, Coupland P, Greenwood W, Esmaeilishirazifard E, Aparicio S, Rueda O, Caldas C. Single cells regenerate the functional and transcriptional diversity of human breast tumours. Manuscript in preparation for submission, 2023.

Invited talks:

1. **Nguyen LV**, Caldas C. Phenotype and function of disease-propagating clones in human breast PDX models. Invited seminar presentation: October 4, 2022 at the ***Princess Margaret Cancer Centre, University Health Network, University of Toronto.***
2. **Nguyen LV**, Caldas C. Phenotype and function of disease-propagating clones in human breast PDX models. Invited seminar presentation: August 18, 2022 at the ***Lady Davis Institute, Jewish General Hospital, McGill University.***
3. **Nguyen LV**, Eyal-Lubling Y, Lui A, Guerrero Romero D, Greenwood W, Esmaeilishirazifard E, Chin SF, Caldas C. Expressed cellular barcoding system reveals differences in clonal growth of basal and luminal human breast tumours. Presented June 22, 2022 at the ***European Association for Cancer Research Congress in Seville, Spain.***
4. **Nguyen LV**. From normal stem cells to cancer stem cells – a look at clonal heterogeneity in the breast using lentiviral-based cellular barcoding. Invited seminar presentation: February 24, 2022 at the ***Sunnybrook Health Sciences Centre, University of Toronto***.

Awards:
Travel Award, European Association for Cancer Research Congress in Seville, Spain. June 2022.

List of Publications and Presentations resulting from other projects during the fellowship period (if applicable)

N/A

Selection of Courses and Workshops Attended During the Fellowship

1. CRUK Bioinformatics Summer School, July 21-23, 2021 & July 26-27, 2021
2. Animal work, Personal Licences course, University Biomedical Services (UBS), University of Cambridge, May 2021

Acknowledgements

This work would not have been possible without the support of the ESMO Translational Research Fellowship, and I am extremely appreciative of the opportunity to do this work and the impact it has had on my career progression because of the support from ESMO.

I am also immensely thankful for the unwavering support of my clinical and research mentor Professor Caldas, who provided critical resources, support and insights to make this project a success. I am also thankful to all the members of the Caldas lab for their contributions to this project, especially to Yaniv Eyal-Lubling who wrote custom scripts to analyze this data and also derived the epithelial molecular signatures, and Daniel Guerrero Romero and Oscar Rueda for their help with the mathematical modelling.

Personal Statement (not mandatory)

The two years I was able to spend in Cambridge because of the support from the ESMO Translational Research Fellowship has had an immeasurable impact on my career. I had the opportunity to present my work internationally and meet many scientists and clinicians who I will continue to work with in the future. Because of this time, I have now been able to secure a position as a Clinician Scientist and Staff Medical Oncologist at the Princess Margaret Cancer Centre in Toronto, Canada, as well as an academic appointment as Assistant Professor in the Department of Medicine at the University of Toronto where I look forward to continuing my research and have this work positively impact patient care.

References
1. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *PNAS* **98**, 10869–10874 (2001). 2. METABRIC Group *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012). 3. Nguyen, L. V., Vanner, R., Dirks, P. & Eaves, C. J. Cancer stem cells: an evolving concept. *Nat Rev Cancer* **12**, 133–143 (2012). 4. Bramlett, C. *et al.* Clonal tracking using embedded viral barcoding and high-throughput sequencing. *Nat Protoc* **15**, 1436–1458 (2020). 5. Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat Biotechnol* **36**, 442–450 (2018). 6. Lim, B., Lin, Y. & Navin, N. Advancing Cancer Research and Medicine with Single-Cell Genomics. *Cancer Cell* **37**, 456–470 (2020). 7. Nguyen, L. V. *et al.* Clonal Analysis via Barcoding Reveals Diverse Growth and Differentiation of Transplanted Mouse and Human Mammary Stem Cells. *Cell Stem Cell* **14**, 253–263 (2014). 8. Lu, R., Neff, N. F., Quake, S. R. & Weissman, I. L. Tracking single hematopoietic stem cells *in vivo* using high-throughput sequencing in conjunction with viral genetic barcoding. *Nature Biotechnology* **29**, 928–933 (2011). 9. Nguyen, L. V. *et al.* DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. *Nat Commun* **5**, 5871 (2014).

SIGNATURES	
Award Recipient full name	***Signature and Date***
Dr. Long Viet Nguyen	 April 2, 2023

Research Mentor full name	***Signature and Date***
Prof. Carlos Caldas	 April 3, 2023

Insert photo of yourself and/or colleagues at the host institute

GOOD SCIENCE
BETTER MEDICINE
BEST PRACTICE

UNIVERSITY OF CAMBRIDGE

This ESMO Translational Fellowship Research Project was supported by an educational grant from ESMO