AWARD NUMBER: W81XWH-14-1-0110

TITLE: A Molecular Framework for Understanding DCIS

PRINCIPAL INVESTIGATOR: Gregory Hannon

CONTRACTING ORGANIZATION: The Chancellor, Masters and Scholars of the University of Cambridge, The Old Schools, Cambridge CB2 1TN

REPORT DATE: October 2018

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2018	Annual	30 Sep 2017 - 29 Sep 2018

4. TITLE AND SUBTITLE

A Molecular Framework for Understanding DCIS

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0110

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Gregory Hannon PhD/ Clare Rebbeck

E-Mail: greg.hannon@cruk.cam.ac.uk

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

CRUK Cambridge Institute
Li Ka Shing Centre
Robinson Way
Cambridge
CB2 ORE

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project centers on creating a molecular framework of DCIS (ductal carcinoma in situ). DCIS is considered to be the precursor to Invasive Ductal Carcinoma (IDC), the most common form of breast cancer. IDC accounts for 80% of all breast cancers, predominantly affecting women aged 55 and older; however, at least a third of women with IDC are diagnosed before they reach 55. Not all patients with DCIS will develop IDC however, we are looking for ways to better predict those patients that need life-saving treatment, and separate these from those patients who are less at risk.

So far we have made close to 2000 RNA libraries and 410 whole genome libraries and dissected material from 167 freshly frozen patient biopsies. The tissue included for RNA and DNA analysis DCIS, IDC, stroma adjacent to DCIS/IDC and normal tissue.

15. SUBJECT TERMS

DCIS, IDC, LCM, RNAseq, DNAseq, evolution

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified		19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

A Molecular Framework for understanding DCIS

Award No. W81XWH-14-1-0110
Annual report year 3

1
Introduction

This project centers on creating a molecular framework of DCIS (ductal carcinoma in situ). DCIS is considered to be the precursor to Invasive Ductal Carcinoma (IDC), the most common form of breast cancer. IDC accounts for 80% of all breast cancers, predominantly affecting women aged 55 and older; however, at least a third of women with IDC are diagnosed before they reach 55.

Utilizing a unique bank of frozen mammary biopsies, containing samples with DCIS alone, and a combination of DCIS and IDC, we have started to profile both DCIS and related tissue components. It is our aim to sample the ~300 biopsies, and compare both by RNA seq, and whole genome amplification, DCIS lesions, within, and between patients, and see how these may be correlated with IDC lesions. We also intend to look for changes in the stroma between those patients that present with IDC and those that do not. This work aims to identify characteristics that may be suggestive of a patients' likelihood of progressing from DCIS to IDC, with the purpose of reducing the need for over treatment for this disease.

2
Keywords

Ductal carcinoma in situ, DCIS, Invasive ductal carcinoma, IDC, RNA, DNA, Copy number, Laser capture microscope, LCM

3.
Accomplishments

Aim 1. The evolution of DCIS.

Task 1. Sample collection and annotation

Task 2. Sample choice from frozen bank.

We have received 167 samples from the frozen bank now and have processed all of these so far. These include pure DCIS and also mixed DCIS and IDC samples. We have selected samples that had 5 or more DCIS legions for this Aim as these will be more informative for looking at the evolution of DCIS. In addition we also selected samples with 3 or more DCIS legions for additional information on their DNA and to cover more patients. We are using both CNV profiles and SNPs (variants are called using the stromal tissue).

Task 3. Laser capture of frozen samples for characterization

From each of the 167 samples we have dissected material for DNA, however we have material from 18 patients for characterization (based on having 5 or more DCIS legions). We have selected DCIS legions, IDC regions, normal epithelium, where present, Athypical epithelium, Solid DCIS, papillary DCIS, benign epithelium, and stroma (as far away from DCIS or IDC regions as possible). The table below represents the distribution across patients, with a total of 214 legions including the normal and variants of epithelium.

Number of DCIS legions	Number of samples	Number of samples with IDC	Number of IDC legions per sample
5	4	2	4,4
6	7	3	8,2,4
7	6	3	5,7,5
8	1	0	
9	1	1	5
10	1	1	6
11	0	0	
12	0	0	
13	1	0	
Total	**144**	**10**	**50**

Task 4. DNA sequencing

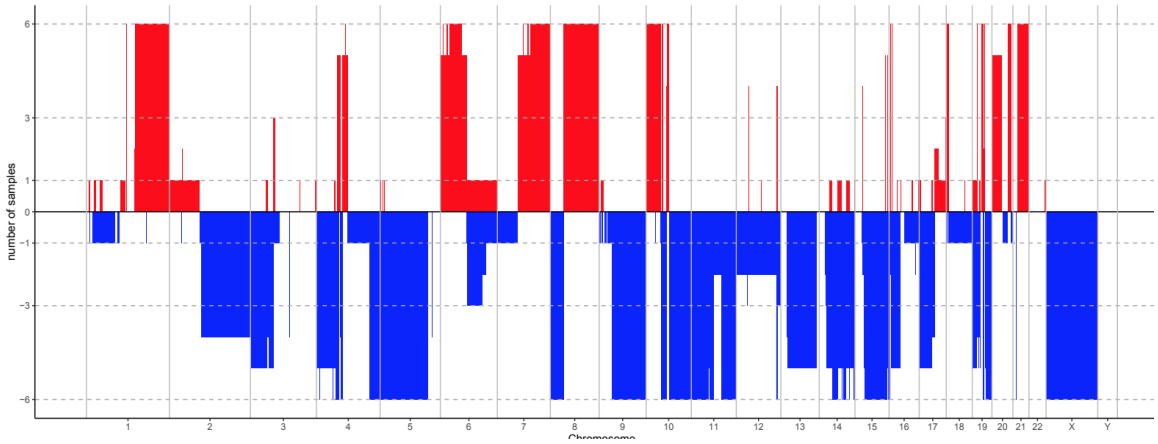
We initiated work on this using the Nextera Exome Capture kit, however on the couple of samples we used, this did not prove successful, as there was a very low distribution of probes represented. Having investigated the costs and what is needed to get deep enough coverage for accurately calling CNVs and SNVs, we decided to make use of the X10 sequencing machine at the NYGC and do whole genome sequencing instead. A trial run with this demonstrated that the Whole Genome Sequencing kit that we were using (and other kits on the market) was only compatible with sequencing machines after the DNA had been sheared (resulting in removal of end primers). This was not efficient with sequencing on the X10, as reads are generally longer and shearing would result in very short reads. We therefore established a new protocol, where by we enzymatically chewed the primers off the ends of DNA strands after amplification with the WGA kit, this then allowed us to attach the primers for sequencing (this was somehow hindered without removal of the WGA primers). This pipeline proved very effective and in addition to the 18 patients we selected for the evolution study, we have also sequenced an additional 40 patients, making a total of 59 patients. This amounts to 410 DNA X10

libraries, comprising of 81 IDC, 201 DCIS, 69 stroma and the remainder of normal epithelium, atypia and benign epithelium.

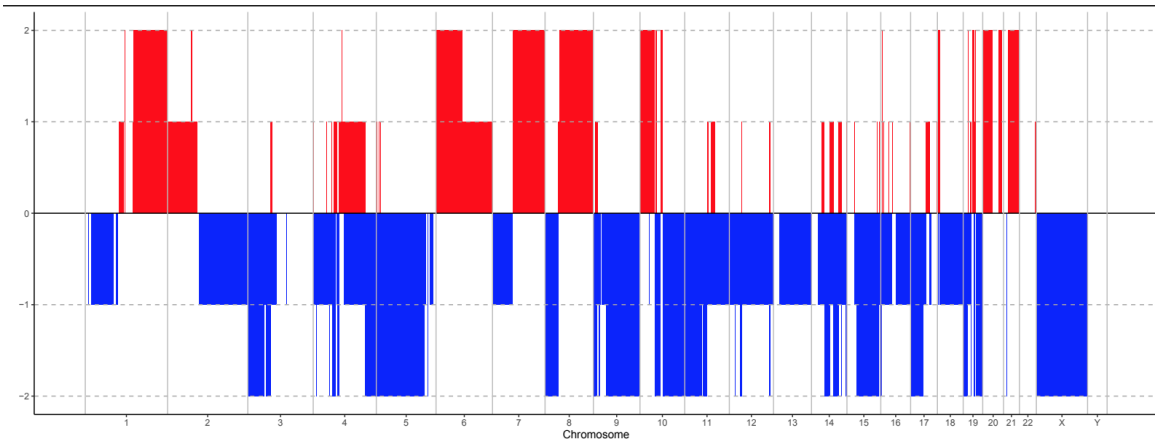
Task 5. Analyze DNA data

All DNA libraries have now been run and analyzed. We run a few quality control analyses on the samples once they have gone through a standard pipeline (this is done by the NYGC). Concordance analysis looks for any discrepancies between a "normal" sample and its paired "tumor" sample. Pairs generally have over 90% concordance, however this analysis has proved useful as it identified a misread tube label and thus allows us to correct such errors. Where samples have a low concordance that can not be corrected easily (ie. We have no other way to identify a mis labled sample) they are unfortunately put to one side for the time being. We also run analyses for "contamination" this could be from the tissue, or from other samples in the library prep. It is likely that the lower the quality or quantity then the higher the effects of any contamination are likely to be. Samples with very low coverage are also put to one side. After removing samples with low concordance, low contamination and low sequencing coverage we currently have data from 41 patients and 233 libraries.

Initial analysis on CNV data has been carried out on 11 patients thus far and shows that there are both similarities and differences to be seen between DCIS legions within the same patient. An example for one patient is below. The plot shows CNVs that are located among the 6 DCIS samples from this patient. You can see that some are shared by all 6 samples (sample number on the Y axis, chromosome number along the X axis) and some are unique to just one or two samples.



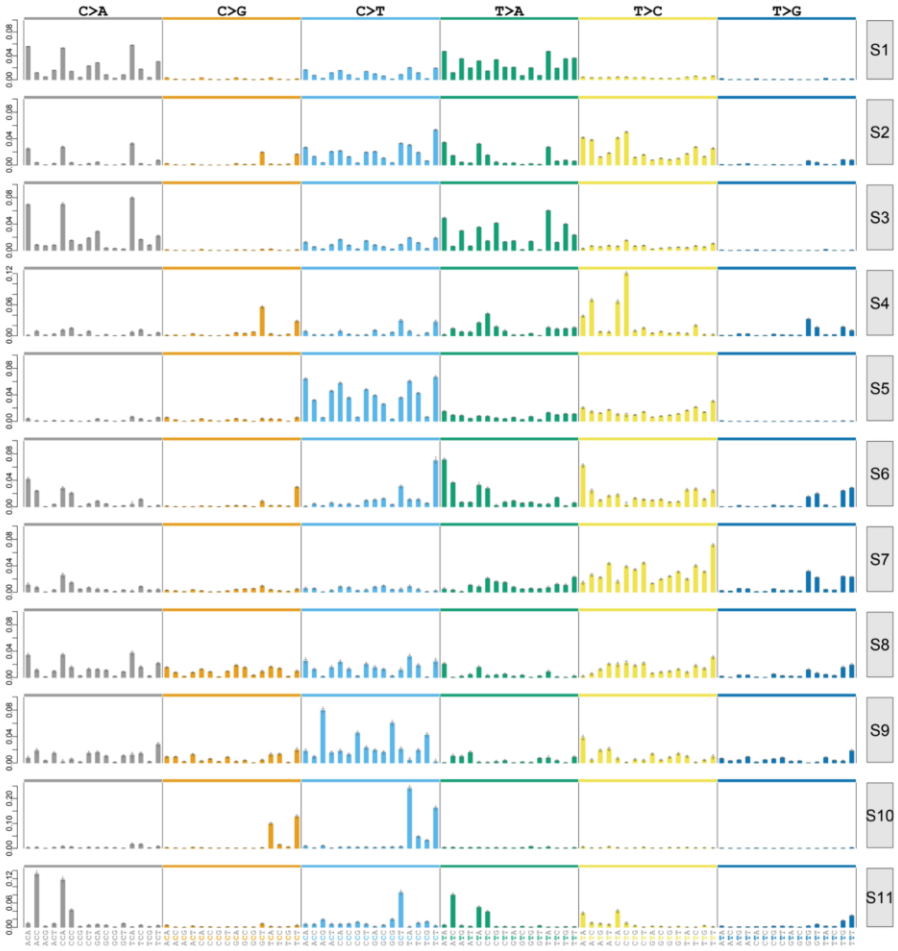
The plot below shows the CNVs that are shared or unique between the DCIS and the IDC of this same patient. The X axis represents a shared region (1 is the CNV is located in only a dcis sample, or just an IDC sample, 2 is that the cnv is found in both an IDC sample and a dcis sample).



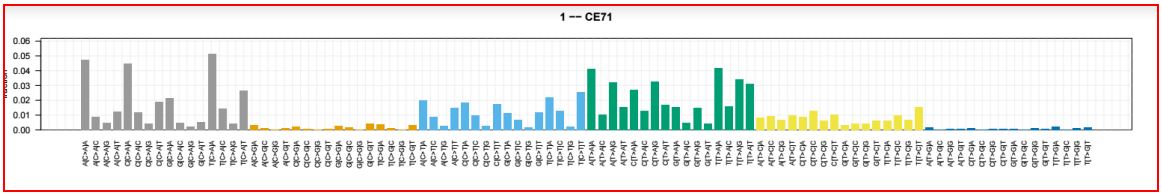
We are currently working on constructing phylogenetic trees using the program LICHEE, using the SNVs. The trees are fairly complex so we are working on creating a simpler/ singular tree using only regions of 2N as determined by the CNV profiles. We have also worked out a confidence cut off for the SNVs based on the number of reads per SNV and the VAF number. We have made this fairly stringent to minimize "noise" and for initial analysis we have used the cut off of 25 reads in both the "normal" (where we use the stroma tissue) and 25 reads in the "tumor", and a VAF of 0.3.

In addition to this we have carried out work on the mutational signatures of these samples to look for characteristic signature patterns.

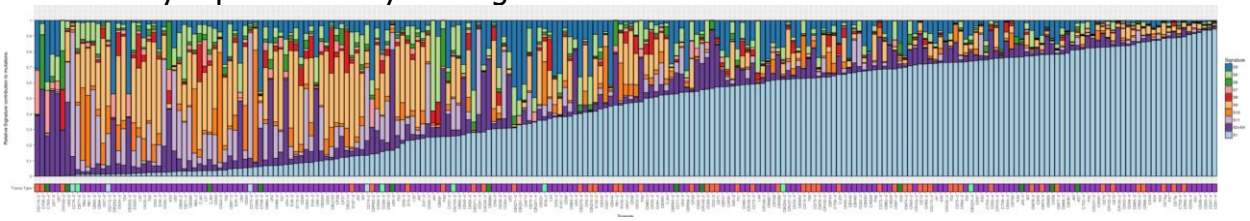
In addition to the more well know APOBEC signature (S10 below), we have also found some mutational signatures that seem to be unique to our data and initial work on these are proving interesting regarding their different characteristics.
Some examples of the signatures and how they segregate our samples are below. We have used a modified program previously used to create the "Cosmic" signatures to identify 11 different signatures, however, it must be noted that some signatures are components and others are used to assign a sample but that sample may contain more than one signature pattern that is not recognized as a unique pattern. Most do fall into their correctly assigned "signature".



S1. This signature seems to be prevalent in the IDC samples and much more interestingly also present in some (often just one, or two) DCIS from the same patient. It is mainly characterized by the Grey, light blue, and green bars, as the yellow and orange bars may vary if there is a different signature combined. Below is an example sample with this signature. And the distribution of all samples and their contribution from S1.

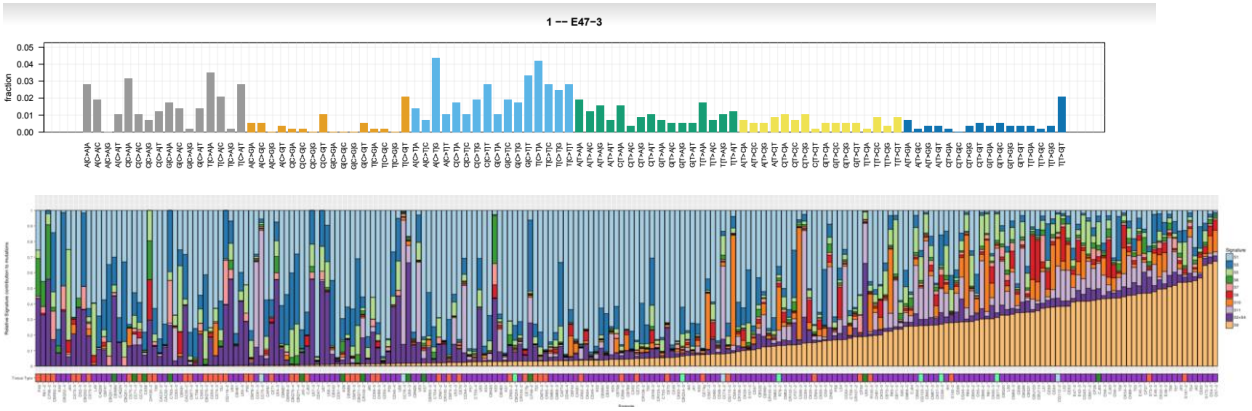


The X axis on the plot below show DCIS with purple Ids, IDC with Orange Ids and others with Green Ids. The black line is the point where the samples (left of the line) are not truly represented by this signature.

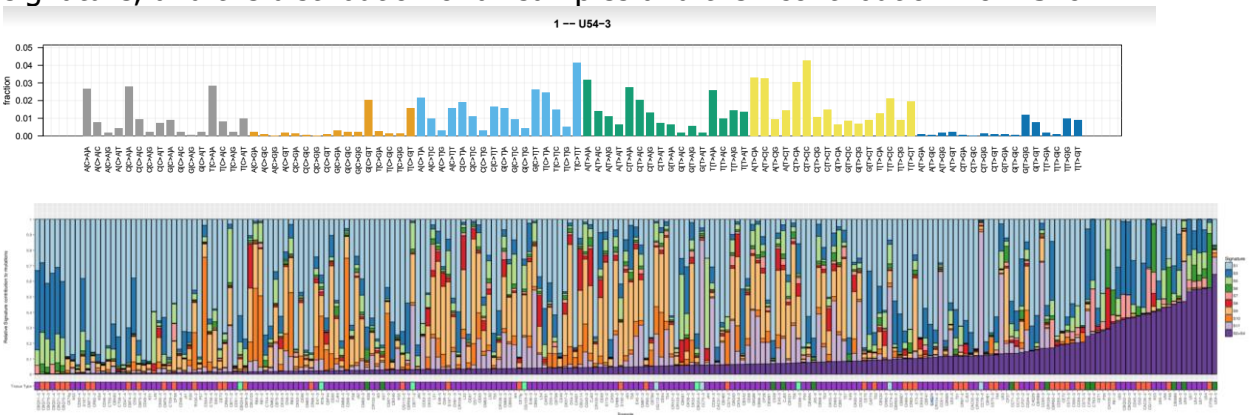


S9. This Signature seems to be prevalent in the DCIS samples and not really seen in the IDC samples (except for 1 patient exception). It is mainly characterized by the light blue

bars. Below is an example sample with this signature, and the distribution of all samples and their contribution from S9. It can be seen that S1 and S9 and generally exclusive of each other.



S2&4 This signature seems to be spread over both DCIS and IDC, and is characterized by the orange bars, yellow and dark blue bars. Below is an example sample with this signature, and the distribution of all samples and their contribution from S2&4.



We are currently trying to find possible mutational or transcriptional components to these signatures.

Aim 2. <u>A transcriptional landscape of early breast cancer.</u>

Task 6. Sample choice from frozen bank.
- choose samples for pure DCIS and DCIS with microinvasion/IDC

Task 7. Laser capture of frozen samples for characterization
We have processed all 167 patients so far. For each sample the following regions are annotated by Joe (the pathologist) and dissected in triplicate for RNA: DCIS, IDC, normal epithelium, Athypical epithelium, Solid DCIS, papillary DCIS, benign epithelium, areas of high immune infiltration, stroma adjacent to DCIS, stroma adjacent to IDC and stroma away (as far away from DCIS or IDC regions as possible). This has provided over 9500 legions.

Task 8. RNAseq library construction

Approximately 2000 RNA seq libraries have been sequenced, with useable quality data from ~1500. Currently we have been focusing on DCIS and IDC and normal and other epithelium and are prioritizing these for sequencing now. This task is still on going and

we are doing the 3rd tissue replicate in cases where we don't have 2 good replicates for a sample. We have decided to resection some patients again where the majority of libraries are not useable. Libraries sequenced now seem to be far better quality than those we initially started with, potentially due to kit improvements or simply due to increased experience in library prep with these difficult samples.

Task 9. Analyze RNAseq datasets

Thus far we have analyzed 1500 DCIS, IDC, benign/normal epithelium and stroma away libraries. For quality control samples with a Gene Assignment of < 15% with % of Uniquely mapped reads < 20, are removed from the group analysis.

We have carried out subtype analysis on the DCIS and IDC samples that we have data for using both the PAM50 and the AIMs methods. We have decided to use the output for the AIMs method rather than the PAM50, as we have found that the subtype profiles tend to change depending on which samples you add to the group. For the Aims method, this does not happen and each subtype is classified based only on the data for that sample. It seems that our subtype classification is generally inline with information sent to us with the tissue samples detailing their classification via standard histology.

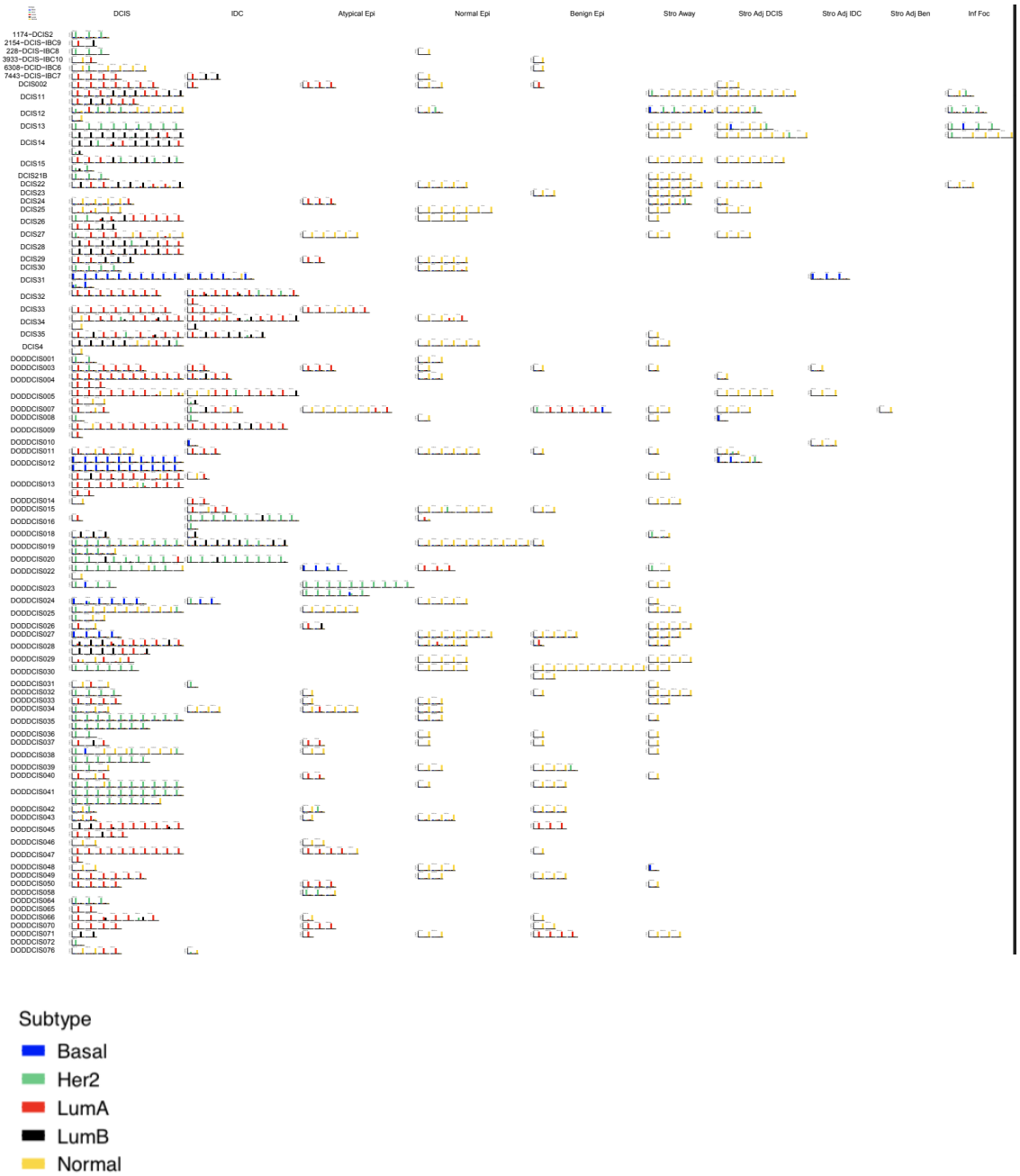
So far, when looking at the majority subtype within a patient the distribution we see is as following:

21 (green) Her2 - 29%
14 (yellow) Normal – 19%
25 (red) lumA – 35%
4 (blue) Basal – 5%
7 (black) LumB – 10%
It seems we have a lower than average (average reported is 15-25%) proportion of Basal subtype tumors. This could be due to the fact that basal tumors are more common in younger woman, and the samples we have come from mammagram guided biopses – typically done on older woman, and also it could be the locality and population biases in the area where the samples are collected from.

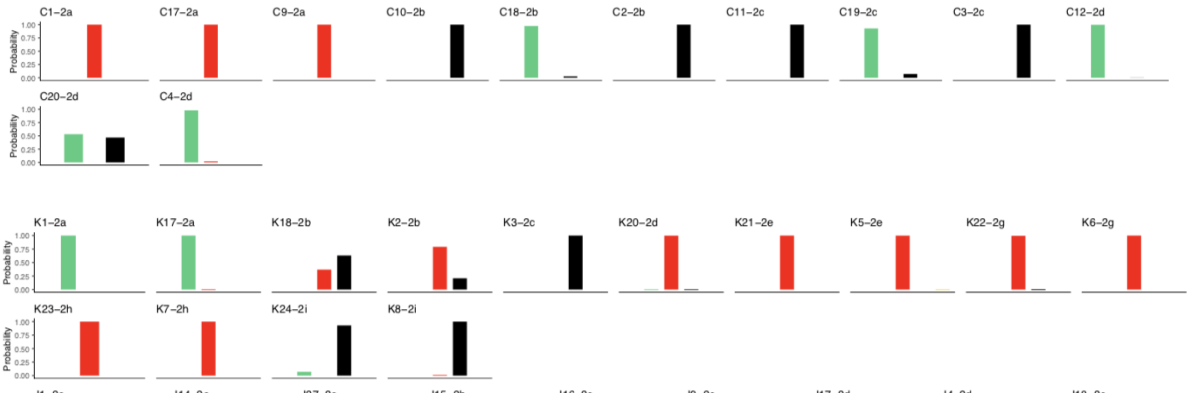
However, if we look at IDC legions only the distribution is slightly different

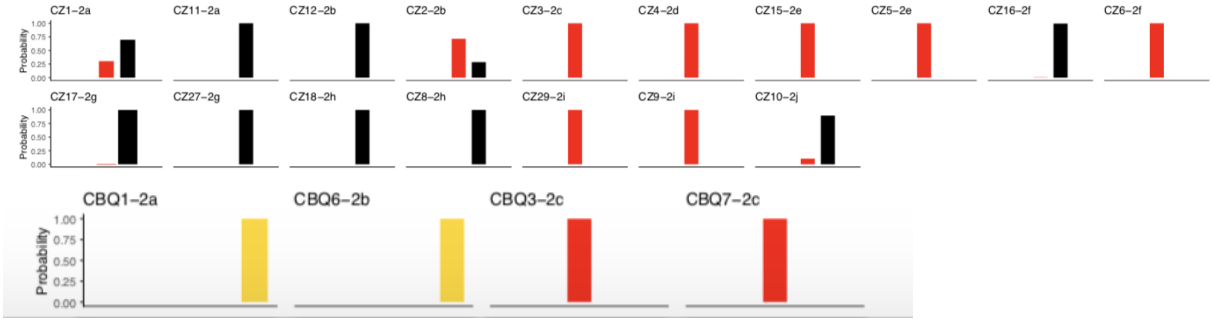
4 (green) Her2 - 16%
1 (yellow) Normal – 4%
11 (red) lumA – 44%
3 (blue) Basal – 12%
6 (black) LumB – 24%

Below is a table showing some patients so far and the subtype profile



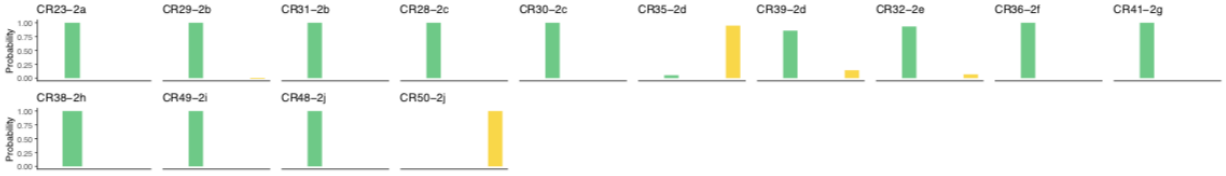
We find that replicates (donated by the letter eg. 2c, 2d) are fairly uniform in their subtype classification. – patient is denoted by the first letter eg. C, K, CZ, and that Basal subtypes appear to be very homogenous as reported by some other publications.



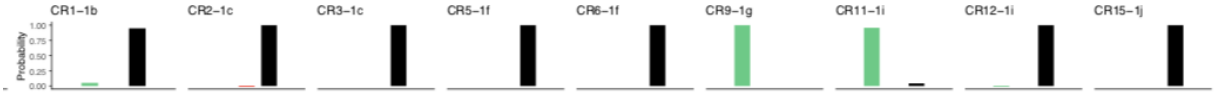


We have also found that the subtype of the IDC can be different from the same patient's DCIS legions, as show below. Patient CR has a Her2 and Normal DCIS subtype but their IDC is more LumB. And patient CE has more Lum A and Normal and their IDC shows subtypes Her2 and LumB present.

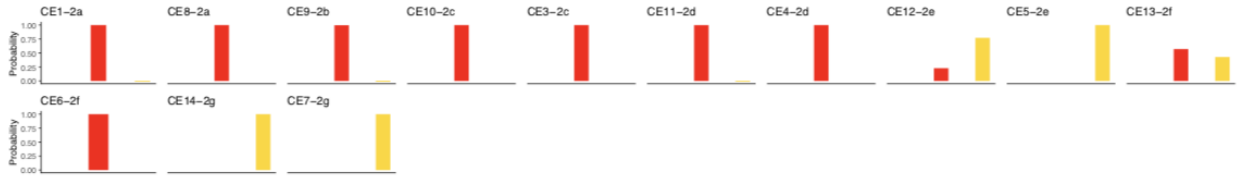
DCIS CR



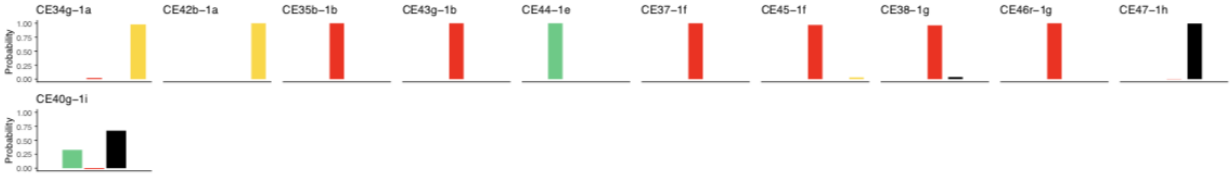
IDC CR



DCIS CE



IDC CE



We are doing some preliminary differential analysis on the RNA seq using different groups as defined by both the subtyping and the DNA data (ie, the different mutational signatures).

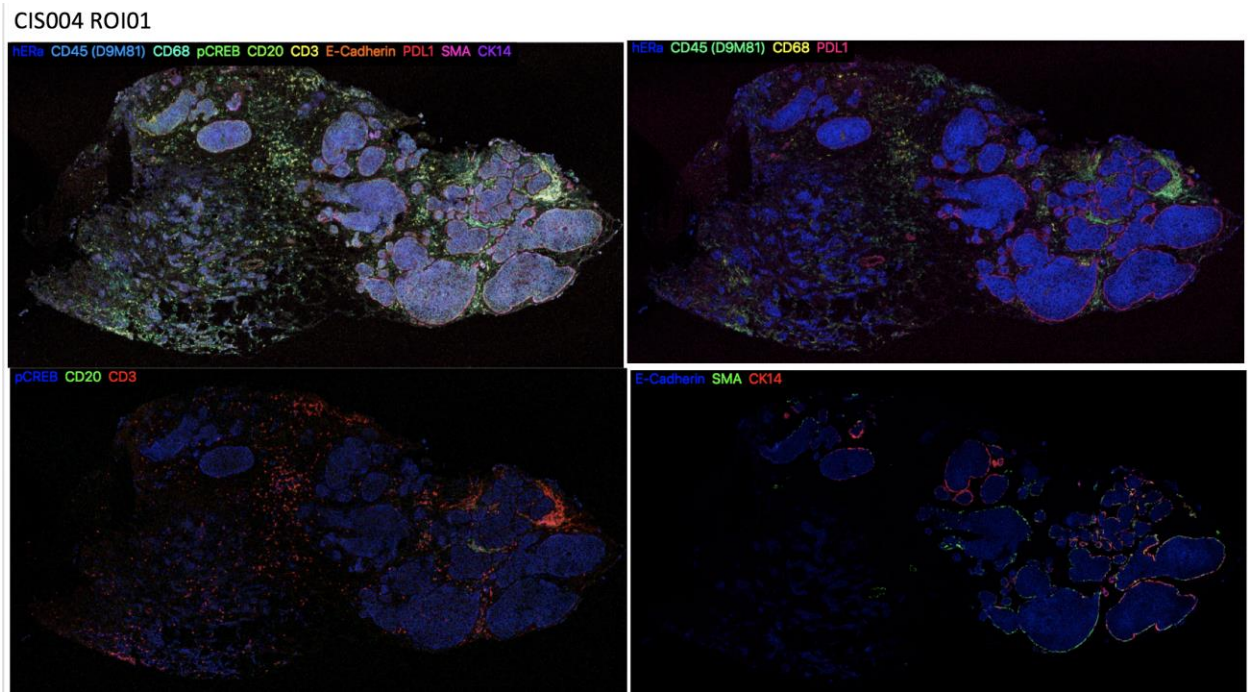
AIM 3: <u>Identify the main cellular contributors within the stromal compartments of DCIS lesions</u>

Imaging Mass Cytometry

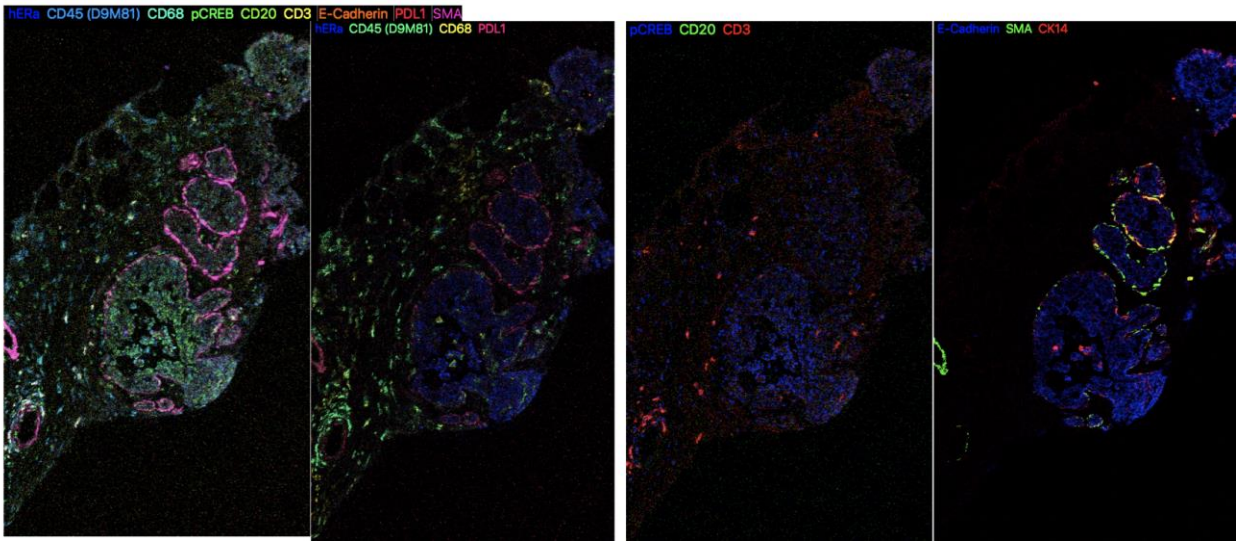
Task 10. Analyze the DCIS microenvironment

In addition to the RNA analysis, and in combination with the DNA analysis we are initiating data collection with IMC (Imaging Mass Cytometry). We have thus far imaged 16 patients. We are now working on improving antibody staining on slides that have previously been stained with H&E. We have developed methods to remove the H&E staining and some antibodies are able to provide signal. This is particularly useful as we can use the same slides that are adjacent to the DNA and RNA legions that we already collected, from the H&E slide that was used to annotate by the pathologist. Other wise we will have to take new sections further down the biopsy.

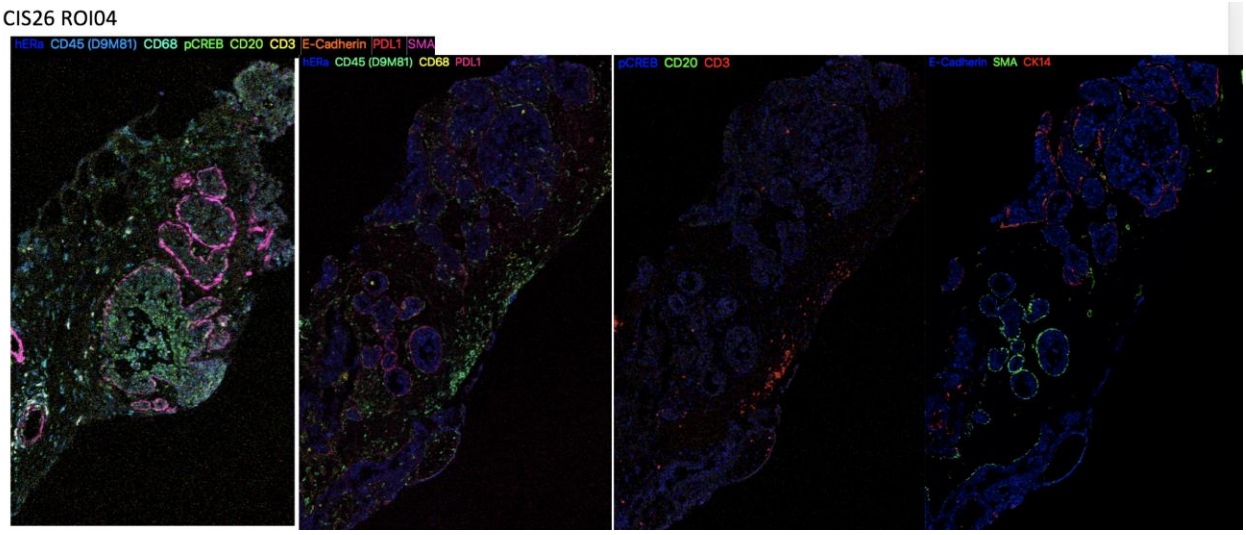
Below are a few images we have been working on. Of particular interest we see some DCIS legions that are surrounded by CK14+ cells and some legions surrounded by SMA. These markers are very interesting due to their relationship to EMT (epithelial-to-mesenchymal transition)



CIS26 ROI03



CIS26 ROI04



Opportunities for training and professional development

Nothing to report (not intended for training)

Results disseminated to communities of interest

In house seminar at CRUK.

4.
Impact

Nothing to report

5.
Changes / problems

IMC data collection is currently slow due to access to the machines required for scanning. This could be sped up considerably with the purchase of another scanning machine.

6.
Products

Nothing to report

7.
Participants & other collaborating organizations

Individuals worked on the project

Name : Greg Hannon
Project Role: Initiating PI – contributed to project design and liaising with bioinformatics team
Nearest person month worked: 1 CM (10% x 12 months)
Funding support: CR-UK and Royal society

Name : Clare Rebbeck
Project Role: Co-PI – contributed to project design, staining strategy, dissecting with the LCM, RNA and DNA library preparation and liaising with Bioinformatics team and pathologist and data analysis.
Nearest person month worked: 9 CM (75% X 12 months)

Name : Jian Xian
Project Role: senior research assistant - contributed to dissecting with the LCM and RNA and DNA library preparation and IMC preparation.
Nearest person month worked: 12 CM
Funding support: CR-UK

Name : Laurence de Torrente
Project Role: Post doc – bioinformatician
Nearest person month worked: 12 CM
Funding support: salaried via the NYGC subaward.

Name : Bassem Ben Cheikh
Project Role: Post doc – bioinformatician
Nearest person month worked: 4 CM
Funding support: salaried part by Duke subaward

Change in active support since last report

Nothing to report (this is the first reporting period)

Other organizations involved as partners

Duke university – collaboration to provide tissue samples and clinical annotation, as detailed in the grant application.

New York Genome Center – Collaboration with the bioinformatics team to analyse the data, as detailed in the grant application.