Spatial and temporal age-related spectral alterations in benign human breast tissue

Georgios Theophilou, Simon W. Fogarty, Júlio Trevisan, Rebecca J. Strong, Kelly A. Heys, Imran I. Patel, Helen F. Stringfellow, Pierre L. Martin-Hirsch, Francis L. Martin*

Centre for Biophotonics, LEC, Lancaster University, Lancaster LA1 4YQ, UK

Correspondence to: Prof Francis L Martin, Centre for Biophotonics, LEC, Lancaster University, Lancaster LA1 4YQ, UK; Email: f.martin@lancaster.ac.uk; Tel.: +44(0)1524 510206

Highlights

- SR-FTIR microspectroscopy to examine benign breast tissue
- PCA-LDA of IR spectra show inter-individual and age-related epithelial differences
- A subset of spectrally-distinct epithelial cells point to putative stem cells
- Spectral changes that may occur several years prior to onset of breast cancer
- Novel approach to shed insights into aetiology of breast cancer

Abstract

Epidemiological evidence suggests that cancers attributable to exogenous carcinogenic agents may appear decades after initial exposures. Environmental factors including lifestyle and/or diet have been implicated in the aetiology of breast cancer. Breast tissue undergoes continuous molecular and morphological changes from the time of thelarche to menopause and thereafter. These alterations are both cyclical and longitudinal, and can be influenced by several environmental factors including exposure to oestrogens. Research into the latent period leading to breast carcinogenesis has been mostly limited to when hyperplastic lesions are present. Investigations to identify a biomarker of commitment to disease in normal breast tissue are hindered by the molecular and histological diversity of disease-free breast tissue. Benign tissue from reduction mammoplasties provides an opportunity to study biochemical differences between women of similar ages as well as alterations with advancing age. Herein, synchrotron radiation-based Fourier-transform infrared (SR-FTIR) microspectroscopy was used to examine the terminal ductal lobular epithelium (TDLU) and, intra- and inter-lobular epithelium to identify spatial and temporal changes within these areas. Principal component analysis (PCA) followed by linear discriminant analysis of mid-infrared spectra revealed unambiguous inter-individual as well as age-related differences in each histological compartment interrogated. Moreover, exploratory PCA of luminal and myoepithelial cells within the TDLU indicated the presence of specific cells, potentially stem cells. Understanding alterations within benign tissue may assist in the identification of alterations in latent pre-clinical stages of breast cancer.

Introduction

The clinical manifestation of breast cancer is the final expression of a complex sequential process that begins with exposure to a causative agent [1]. Tumour formation involves temporal alterations in genetic morphology or expression, which directly or indirectly disturbs normal cellular regulation of proliferation and growth inhibition, leading to malignancy [2]. The period from an initiating event to tumour formation is termed the "latency period" [3]. This definition implies that cancers in which environmental exposures play a role arise several years after initiating exposures [1, 4]. This latency period may be of different lengths depending on the type, timing and length of exposure as well as inherent predisposition to the particular type of cancer [5].

Exposure to carcinogens will certainly vary significantly between individuals as will their response to such agents [6, 7]. Factors that predispose women to a risk of breast cancer include early menarche, late menopause, nulliparity or delayed parity and, use of contraception and hormone replacement therapy [8, 9]. Such characteristics are associated with increased oestrogen exposure. It is now accepted that "Westernized" lifestyle either through immigration or adoption of Western diet are likely causative factors for breast or other hormone-dependent cancers [1, 10].

Little is known regarding the molecular changes that may develop before the appearance of pre-clinical or clinical breast cancer [11]. Changes that appear at the initiation stage or during the latency period may provide useful biomarkers for early identification of at-risk women. Also, these changes may be temporary, regressive, permanent or progressive [11, 12]. Biomarkers that could identify lesions with a low risk of progression towards malignancy or even the chance of regression would be

advantageous [13-15]. Women exhibiting high-risk alterations might be encouraged to make appropriate lifestyle alterations to "rectify" such changes [16].

To facilitate understanding of pathological processes involved in breast carcinogenesis, we first need to identify physiological differences within tissues from similarly-aged women as well as age-related alterations. The areas wherein these variations are most likely to occur are within the terminal ductal lobular unit (TDLU) along with the supporting intra- and inter-lobular stroma (Figure 1a). These areas are thought to be responsible for cancer initiation processes [11, 17, 18]. The TDLU consists of terminal ductules ending in acini, bounded by luminal epithelial cells, which are surrounded by myoepithelial cells (Figure 1b). TDLUs have different compositions depending on their developmental stage from pre-puberty to menopause (Figure 1c). The pre-pubertal "simple" TDLU consists of one central ductule with three or four branches. After menarche, the TDLU's morphology depends on the stage of the menstrual cycle with luminal cells growing in size as the cycle progresses from the follicular to the luteal phase. During pregnancy or lactation, the TDLU hypertrophies and remains in a similar state to the luteal phase. Post-menopausally, the lobule has fewer ductules and a denser intralobular stroma. With advancing age, the TDLU undergoes complete atrophy but the branching duct tree remains. In cancer, the micro-architecture of the TDLU is disturbed.

Breast tissue from reduction mammoplasties provides an opportunity to study spatial and temporal variations that may exist within the TDLU and surrounding areas of the mammary gland when there is no evidence of malignant or pre-malignant changes. The cancer risk in this population is comparable or marginally reduced relative to the general population [19, 20]. Biospectroscopy techniques can lend novel insights into structural alterations in target cells [21, 22]. This approach has been

employed to detect alterations associated with cancer in various tissues [23-27]. Its potential in identifying biomarkers that can be used in cancer screening has also been examined [28, 29].

Synchrotron radiation-based Fourier-transform infrared (SR-FTIR) microspectroscopy is superior to conventional bench-top systems in that it gives enhanced spectral signal-to-noise ratio (SNR) with greater spatial resolution. The main reason is that a synchrotron source emits a collimated light beam that is more brilliant than that of a typical globar source found in a bench-top spectrometer. This provides an excellent SNR that is 1000 times greater to that of conventional IR sources and allows spatial resolutions as small as 10 μm [23, 30].

Interrogation of biological tissues by IR spectroscopy can results in thousands of spectra, necessary due to the complex chemical composition of cells. Such a large amount of data obtained in an increasingly typical spectrochemical experiment may be analysed using multivariate analysis [31, 32]. This allows data simplification for visual representation and exploratory analysis. Two of the commonly utilised multivariate analysis approaches are: principal component analysis (PCA) and linear discriminant analysis (LDA) or a combination of both [33]. This study aimed to identify spectral differences in breast tissue of women of similar ages as well as changes with increasing age. This could be a first step towards the recognition of the initiation/early promotion stages of breast cancer.

Materials and Methods

Sample preparation

Human breast tissue was obtained from eleven patients who had undergone reduction mammoplasty for indications other than breast-related pathology. Consent was taken with ethical approval (Research Ethics Committee reference: 10/H0308/75) according

to the Declaration of Helsinki. Five individuals were aged 20 to 29 y, three were aged

30 to 39 y and three were aged 40 to 49 y. Breast tissue samples obtained were

formalin-fixed and paraffin-embedded. Ten-μm-thick tissue sections were floated

onto 1 cm × 1 cm BaF_2 slides (Photox Optical Systems). These were de-waxed by

serial immersion in sequential fresh xylene baths (×3) for 5 min and washed in an

acetone bath for another 5 min. Resulting samples were allowed to air-dry and then

placed in a desiccator until processing. Four-μm-thick parallel tissue sections were

stained with H&E. These assisted with correct identification of the different cell types

when overlaying mapping grids on the micrographs visualized through the SR-FTIR

microscope (Figure 1d).

Synchrotron radiation-based FTIR micro-spectroscopy

Spectral images were acquired using a Bruker Vertex 80v spectrometer coupled to a

Bruker Hyperion 3000 microscope containing a mercury cadmium telluride detector

cooled with liquid nitrogen, on the Multimode InfraRed Imaging and

Microspectroscopy (MIRIAM) Beamline B22 at the Diamond Light Source Ltd, UK

(www.diamond.ac.uk). Spectra were collected in transmission mode *via* a 36×

objective lens employing an aperture of 10 μm × 10 μm with a step size of 10 μm

intervals, 256 co-additions were acquired; all maps were generated within a total

acquisition time of ≈6 h. Background spectra were taken every 10 spectra to

compensate for beam and atmospheric alterations. Spectra were then converted to

absorbance using OPUS 8 software from Bruker Optics.

Spectral pre-processing

Absorbance spectral images were converted to suitable digital files (.txt) for input to

Matalab software. Computational analysis was carried out using in-house written

scripts for Matlab [34]. The wavenumber regions inputted were between 4,000 cm^{-1}

and 600 cm^{-1}. Spectra were then cut to include the regions between 1,800-900 cm^{-1} as

this is the spectral region most associated with biologically active molecules [35].

They were smoothed using the first derivative Savitzky-Golay filter, rubberband

baseline corrected and normalized to Amide I (1650 cm^{-1}).

Computational analysis

Pre-processed spectral data were explored using PCA. This was carried out as an

unsupervised technique using the first 10 principal components (PCs). Generally, the

first 10 PCs account for approximately 99% of the variance within a sample

population, without introducing excessive noise [33]. The output from PCA was

inputted into LDA. LDA is a supervised technique that maximizes inter-category

variance. The software analyses 90% of the data while using 10% of the data to train

itself. This process is repeated 10 times in a cyclical fashion, so that all data are used

for both analysis and system supervision. The statistical significance of each PC and

LDA contributing to class segregation was determined by the *ANOVA* test in

Graphpad 7 when >2 classes were present. For visualization purposes, scores plots

and loadings curves were generated.

Scores plots, derived from PCA-LDA allow visualization of a spectrum as a

single point, whose coordinates are its scores on a number of axes. This simplifies

visualization of potential differences between the particular classes as well as identify

co-clustering of similar spectral signatures. Loading curves allow identification of

distinguishing wavenumbers between classes. The *x*-axis represents wavenumbers

from 900 to 1800 cm^{-1}. The *y*-axis represents the absorbance coefficient; the highest

peaks and troughs on this axis identify the wavenumbers that are most responsible for

separation between selected classes. The 6 greatest absorbance coefficient deviations

were selected. These wavenumbers were then compiled onto tables alongside

tentative assignments. The resulting tables point to biochemical entities responsible for class segregation.

Results

Breast tissue sections from 11 patients were interrogated using SR-FTIR microspectroscopy. This allowed a high SNR with apertures close to the diffraction limit; the spatial resolution was 10 μm × 10 μm. A raster scan approach was applied to include TDLU regions within the specimens. The high resolution allowed separation of histologically different layers of the TDLUs and surrounding regions. A photomicrograph of involved areas with overlaid markers was used for identification of specific regions from which spectra were acquired. Following interrogation, spectral differences were apparent between the various location-derived spectra (Figure 1d). These differences allowed classification into inter-lobular stroma, intra-lobular stroma and TDLUs; the latter were further classified into myoepithelial cell and luminal cell layers. Correlation with parallel H&E tissue sections ensured correct selection of different cell classes.

In total there were $n=539$ spectra within the inter-lobular stromal class, $n=442$ spectra within the intra-lobular stromal class and $n=591$ spectra within the TDLU class. Within the TDLU class there were $n=155$ spectra from the luminal cell layer and $n=436$ from the myoepithelial layer. The very large number of individual spectra would impede easy visualization in scores plots. Therefore, for most classes every two, three or five spectra in chronological order were averaged (Table 1). The different areas were interrogated individually to identify putative spatial and temporal differences within or between the age groups. Furthermore, the myoepithelial and luminal cell layers within TDLUs were separately examined to identify the possible presence of cells with divergent spectral signatures indicative of disease [36].

9

Inter-individual variations

IR spectra derived from synchrotron maps were extracted and assigned to their pre-defined histological location (TDLU, inter-lobular stroma and intra-lobular stroma). IR spectra from each individual woman represented a separate class. These classes were allocated to their designated age group: 3rd, 4th or 5th decade of life. Each location was analysed separately for each individual within every age group to investigate the existence of inter-individual variation and identify responsible wavenumbers. PCA-LDA cascade analysis was used to reduce each spectrum to a single point at the same time as maximizing inter-category differences. Scores plots identified clustering of spectra taken from the same individual independent of location (Figure 2). Moreover, they identified clear separation between individuals of similar ages for all histological classes. This separation was highly significant and associated with different wavenumbers for different ages and histological locations.

Figure 3 shows the loadings curves containing the 6 principal discriminating wavenumbers for each category. Table 2 illustrates the discriminating wavenumbers for each location and age category alongside tentative assignments. Different spectral signatures were responsible for maximum segregation between individuals for the different histological classes. Two spectral biomarkers were responsible for inter-individual variation in all age groups: righthand side (RHS) Amide I (1,630 cm^{-1}) was responsible for segregation between TDLUs and DNA/RNA (1,080 cm^{-1}) for segregation between intra-lobular stromata of individuals of all ages.

Temporal variations

Following the observation that discriminating wavenumbers segregating TDLUs and intra-LS between individuals were consistent over all age groups, IR spectra were examined to determine the existence of variation between ages for the selected

histological areas. This was performed using the same parameters and method of analysis (*i.e.*, PCA-LDA using the first 10 PCs). Pre-processed spectral data was classified according to allocated histological locations and age groups (instead of individuals within a group). Resulting scores plots (Figure 4) reveal clustering within age groups but not as much segregation as might be expected. Despite overlap, the classes were significantly segregated ($P <00001$).

Corresponding loadings plots (Figure 4) highlight the 6 wavenumbers most responsible for segregation between age groups for each histological location. Another spectral biomarker corresponding to RHS Amide II (1,456 cm^{-1}) was a major segregating wavenumber for TDLUs. Therefore, RHS Amide II was responsible for segregation between individuals as well as between age groups within TDLUs. In the case of Intra-LS, DNA/RNA (1,080 cm^{-1}) alterations were not found to be responsible for segregation between age groups. Instead the principal segregating wavenumber was associated with ring base (1,554 cm^{-1}). All age groups were segregated in every histological class by glycogen (1040 cm^{-1}).

Inter-individual and temporal variations within the TDLU

Concentrating on TDLUs, spectra derived from point maps acquired *via* the 10 μm × 10 μm beam aperture were extracted and assigned as either myoepithelial or luminal cell classes. These spectra were classified initially according to individuals within age groups and analysed using PCA-LDA to identify wavenumbers responsible for segregation between these cell types. Despite some overlap, there was apparent segregation between the different cell classes, and this was significant ($P <0.0002$) in all age groups (Figure 5).

After confirmation that luminal and myoepithelial cells are segregated, each location was separately analysed to identify discriminating wavenumbers between age

11

groups (Figure 6). Amide I again featured as one of the significant discriminating wavenumbers in between the two cell layers. Interestingly, within the myoepithelial layer a principal discriminating wavenumber was for $v_sPO_2^-$ (1,094 cm^{-1}). This wavenumber has been found to be associated with stem cells in a variety of tissues including TDLUs in previous studies [19, 37].

Exploratory PCA within the myoepithelial and luminal layers

To further investigate the existence of spectral discriminating factors within the myoepithelial and luminal layers, IR spectra taken *via* the 10 μm × 10 μm aperture were extracted from the image maps of all individuals. These spectra were not averaged. They underwent exploratory PCA using the first 10 PCs, which accounted for ≈99% of variance. Three-dimensional scores plots were extracted representing the 3 first PCs (Figure 7a). Each point in the resulting scores plots represent a single spectral point in the image maps. Both plots identify spectral points that segregated away from the majority of clustered spectra. These "outliers" were particularly obvious along the PC3 axis in the case of myoepithelial cells. Loadings plots for PC3 identified the wavenumbers responsible for the separation of "outliers" from the clustered spectra (Figure 7b). In a similar fashion, PC1 was most responsible for these spectral points segregating away from clustered spectra in PCA scores plots for luminal cells. Loadings plots for PC1 were used to extract the top 6 segregating wavenumbers. The major discriminating factor in both cases was Amide I.

Discussion

The purpose of this study was to identify spectral differences within normal breast tissue of women of similar ages as well as age-related alterations. Using the IR radiation beam of a synchrotron facility, normal breast tissues from eleven healthy women were examined. Specifically the areas interrogated were TDLUs and surrounding intra-lobular and inter-lobular stroma. Morphological and molecular

12

alterations within these areas have been shown to be directly associated with breast carcinogenesis [11, 38]. Spectroscopic analysis of these areas in healthy individuals may reveal specific molecular causes of the vast heterogeneity that exists within breast tissues. It may also reveal alterations that predate carcinogenesis such as primary pre-cancerous changes including hyperplasia, premalignant changes and carcinoma *in situ* [38]. SR-FTIR spectra obtained were from the mid-IR region (900 to 1800 cm^{-1}) within which most bio-molecules can be identified [35].

Computational analysis of spectra highlighted significant differences between similarly aged individuals. Some of the discriminating wavenumbers responsible for this variation are also responsible for variation between age groups. Namely, Amide I was responsible for inter-individual variability in TDLUs while DNA/RNA (O-P-O stretching) was responsible for spectral separation of INTRA-LS in all age categories. Other spectral bio-molecular signatures were associated with separation of specific histological locations only in one age group (Table 2). Also, certain wavenumbers could identify inter-individual variations within age groups while others could identify inter-individual variation in all age groups, illustrating the vast heterogeneity that exists. Factors that contribute to this heterogeneity include: previous history of breast cancer, positive family history with or without *BRCA* mutations, nulliparity, late parity, high body mass index, use of hormonal contraception or hormone replacement therapy and menopausal status [39-41].

When analysing spectral signatures of the same histological areas for temporal variations between the three defined age groups it was noted that there was co-clustering of spectra from the same age groups and segregation between groups for all areas (Figure 4). Histologically, breast tissue undergoes several changes with age. These changes start at thelarche with the branching of the lactiferous ducts and reach

maturity in puberty with the formation of the adult TDLU. Changes continue in a cyclical fashion with menstrual cycles. Pregnancy and lactation cause hyperplasia of the TDLU. After menopause the TDLU involutes but the pattern of involution is different for nulliparous and parous women. The morphological and functional differentiation of the mammary epithelium is directly dependent on systemic hormones (mainly oestrogen and progesterone) but also by local signalling from the adjacent stroma [42]. Our study identified potential spectral alterations that may be associated with age-related histological changes. Some alterations were unique to particular histological areas (TDLU, Inter-LS or Intra-LS) with some responsible for age-related differences (Table 3). Spectral alterations within TDLUs and surrounding stroma may provide evidence to support age-related changes in functional interactions between these areas. Furthermore, it will be interesting to determine if such alterations may be associated with breast carcinogenesis [17, 42, 43, 44].

It is widely accepted that that the first morphological changes associated with cancer occur in the bi-layered TDLU epithelium [45]. It has also been hypothesized that micro-anatomical changes predating pre-cancerous changes reside in the same areas [46]. In order to further examine the role of IR spectroscopy in identifying such changes within each layer of the TDLU, IR spectra taken from this location were reclassified into luminal and myoepithelial cell categories. These spectra were analysed using multivariate analysis as before. When investigating inter-individual variations, the resultant 1-D scores plots revealed significant separation between the two layers in all age groups. Related loadings curves identified the responsible wavenumbers and their corresponding molecules are presented in Figure 6. These spectral variations may be associated with morphological differences that are specific

to a particular cellular layer of the TDLU rather than the whole TDLU structure. They may be used to pinpoint the cells associated with the increase in breast cancer risk.

The same two layers were examined for spectral variation between age groups. 2-D scores plots identified some segregation between age groups for both cellular layers. Amide I featured as a major discriminatory molecule for both cell types as well as for the whole TDLU. Unsupervised exploratory PCA of the luminal and myoepithelial cells identified aberrant spectral signatures in both layers. These signatures may represent multi-potent or uni-potent stem cells responsive are either age related or hormone dependent alterations. Indeed, there is expanding evidence that FTIR is capable of identifying stem cells in several tissues including cornea, epidermis and intestine [47-51]. Similarly stem cells within the mammary gland may undergo continuous differentiation under hormonal or micro-environmental influences and account on the diversity of breast tissue [42]. They may also represent pluripotent progenitor cells whose abnormal differentiation under oxidative stress in adjacent stroma can lead to carcinogenesis [42, 44, 45, 52].

Conclusion

With this study we demonstrated that ST-FTIR micro-spectroscopy coupled with multivariate computational analysis might be to identify discriminating biomarkers for both inter-individual and temporal variation within breast tissue. We also demonstrated the histological locations where this variation potentially occurs. This is particularly important as it demonstrates the potential interplay between external environmental influences, endogenous hormonal control and micro-environmental communication with the epithelial cells of the TDLU similarly to other tissues [17, 44, 53, 54, 55]. Although specific molecular changes associated with the vast variability

encountered in the mammary gland remains elusive spectral imaging is able to identify classes of molecules that may be used in the search for biomarkers associated with the initiation of breast disease. In the future, FTIR spectroscopy may be able to track molecular changes within particular cell layers involved in disease to produce a database of related biomarkers [56]. Moreover spectroscopy involves non-destructive procedures that do not produce oxidative radiation, as is the case with mammography [57]. Therefore it can be used to obtain molecular profiles of cell populations *in situ* [37]. Identification of such biomarkers may be followed by the application of bio-spectroscopic techniques in clinical practice. Spectral alterations associated with increased risk of breast cancer in a healthy population, may be used as biomarkers in potential population-screening programs without the need to identify high-risk individuals for inclusion to such a program. Further research in the field is required to assess spectroscopic applications in the search for biomarkers for screening for breast disease. Once specific biomarkers are established similar techniques may translate into clinical practice for the evaluation biomarkers within live breast tissue.

References

[1] P.L. Grover, F. L. Martin, Carcinogenesis 23 (2002) 1095-1102.

[2] L.G. Pedraza-Farina, (2006) Yale J. Biol. Med. 79 (2006) 95-103.

[3] K.J. Rothman, Am. J. Epidemiol. 114 (1981) 253-259.

[4] C.E. Land, M. Tokunaga, K. Koyama, M. Soda, D.L. Preston, I. Nishimori, S. Tokuoka, Radiat. Res. 160 (2003) 707-717.

[5] H.K. Armenian H. K. (1987). J. Chronic Dis. 40 Suppl 2 (1987) 9S-15S.

[6] C. Kahlenborn, F. Modugno, W.B. Severs, Mayo Clin. Proc. 81 (2006) 1290-1302.

[7] K.C. Johnson, A.B. Miller, N.E. Collishaw, J.R. Palmer, S.K. Hammond, A.G. Salmon, K.P. Cantor, M.D. Miller, N.F. Boyd, J. Millar, F. Turcotte, Tob. Control 20 (2011) e2.

[8] M.J. Broeders, A.L. Verbeek, Q J Nucl. Med. 41 (1997) 179-188.

[9] J. Wohlfahrt J., H. Mouridsen, P.K. Andersen, M. Melbye, Int. J. Cancer 81 (1999) 49-55.

[10] J.D. Yager, N.E. Davidson, N. Engl. J. Med. 354 (2006) 270-282.

[11] D.C. Allred D. C., S.K. Mohsin, S.A. Fuqua, Endocr. Relat. Cancer 8 (2001) 47-61.

[12] G. Arpino, R. Laucirica, R.M. Elledge, Ann. Intern. Med. 143 (2005) 446-457.

[13] M. Buyse, S. Loi, L. van't Veer, G. Viale, M. Delorenzi, A.M. Glas, M.S. d'Assignies, J. Bergh, R. Lidereau, P. Ellis, A. Harris, J. Bogaerts, P. Therasse, A.

Floore, M. Amakrane, F. Piette, E. Rutgers, C. Sotiriou, F. Cardoso, M.J. Piccart,

TRANSBIG Consortium, J. Natl. Cancer Inst. 98 (2006) 1183-1192.

[14] L.J. Esserman, D.H. Moore, P.J. Tsing, P.W. Chu, C. Yau, E. Ozanne, R.E.

Chung, V.J. Tandon, J.W. Park, F.L. Baehner, S. Kreps, A.N. Tutt, C.E. Gillett, C.C.

Benz, Breast Cancer Res. Treat. 129 (2011) 607-616.

[15] C. O'Donoghue, L. Esserman, Br. J. Cancer 108 (2013) 2200-2201.

[16] G.L. Blackburn, K.A. Wang, Am. J. Clin. Nutr. 86 (2007) s878-s881.

[17] L. Ronnov-Jessen, O.W. Petersen, M.J. Bissell, Physiol. Rev. 76 (1996) 69-125.

[18] M.C. Adriance, J.L. Inman, O.W. Petersen, M.J. Bissell, Breast Cancer Res. 7

(2005) 190-197.

[19] M.H. Brown, M. Weinberg, N. Chong, R. Levine, E. Holowaty, Plast. Reconstr.

Surg. 103 (1999) 1674-1681.

[20] F.E. Hassan, M.D. Pacifico, Aesthetic Plast. Surg. 36 (2012) 1105-1113.

[21] H.Y. Holman, M.C. Martin, E.A. Blakely, K. Bjornstad, W.R. McKinney,

Biopolymers 57 (2000) 329-335.

[22] J.R. Mourant, Y.R. Yamada, S. Carpenter, L.R. Dominique, J.P. Freyer, Biophys.

J. 85 (2003) 1938-1947.

[23] M.J. Tobin, M.A. Chesters, J.M. Chalmers, F.J. Rutten, S.E. Fisher, I.M.

Symonds, A. Hitchcock, R. Allibone, S. Dias-Gunasekara, Faraday Discuss. 126

(2004) 27-39; discussion 77-92.

[24] M.J. German, A. Hammiche, N. Ragavan, M.J. Tobin, L.J. Cooper, S.S. Matanhelia, A.C. Hindley, C.M. Nicholson, N.J. Fullwood, H.M. Pollock, F.L. Martin Biophys. J. 90 (2006) 3783-3795.

[25] J.L. Pichardo-Molina, C. Frausto-Reyes, O. Barbosa-García, R. Huerta-Franco, J.L. González-Trujillo, C.A. Ramírez-Alvarado, G. Gutiérrez-Juárez, C. Medina-Gutiérrez, Lasers Med. Sci. 22 (2007) 229-236.

[26] K. Gajjar, L.D. Heppenstall, W. Pang, K.M. Ashton, J. Trevisan, I.I. Patel, V. Llabjani, H.F. Stringfellow, P.L. Martin-Hirsch, T. Dawson, F.L. Martin, Anal. Methods 5 (2012) 89-102.

[27] K. Gajjar, J. Trevisan, G. Owens, P.J. Keating, N.J. Wood, H.F. Stringfellow, P.L. Martin-Hirsch, F.L. Martin, Analyst 138 (2013) 3917-3926.

[28] M.J. Walsh, M.J. German, M. Singh, H.M. Pollock, A. Hammiche, M. Kyrgiou, H.F. Stringfellow, E. Paraskevaidis, P.L. Martin-Hirsch, F.L. Martin, Cancer Lett. 246 (2007) 1-11.

[29] A.T. Harris, A. Lungari, C.J. Needham, S.L. Smith, M.A. Lones, S.E. Fisher, X.B. Yang, N. Cooper, J. Kirkham, D.A. Smith, D.P. Martin-Hirsch, A.S. High, Head Neck Oncol 1 (2009) 34.

[30] L.M. Miller, P. Dumas, Biochim. Biophys. Acta 1758 (2006) 846-857.

[31] F.L. Martin, J.G. Kelly, V. Llabjani, P.L. Martin-Hirsch, I.I. Patel, J. Trevisan, N.J. Fullwood, M.J. Walsh, Nat. Protoc. 5 (2010) 1748-1760.

[32] M.J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H.J. Butler, K.M. Dorling, P.R. Fielden, S.W. Fogarty, N.J. Fullwood, K.A. Heys, C. Hughes, P. Lasch, P.L. Martin-

Hirsch, B. Obinaju, G.D. Sockalingum, J. Sulé-Suso, R.J. Strong, M.J. Walsh, B.R. Wood, P. Gardner, F.L. Martin Nat. Protoc. 9 (2014) 1771-1791.

[33] F.L. Martin, M.J. German, E. Wit, T. Fearn, N. Ragavan, H.M. Pollock, J. Comput. Biol. 14 (2007) 1176-1184.

[34] J. Trevisan, J., P.P. Angelov, A.D. Scott, P.L. Carmichael, F.L. Martin, Bioinformatics 29 (2013) 1095-1097.

[35] Z. Movasaghi, S. Rehman, Dr. I. ur Rehman, Appl. Spectrosc. Rev. 43 (2008) 134-179.

[36] O.W. Petersen, T. Gudjonsson, R. Villadsen, M.J. Bissell, L. Rønnov-Jessen, Cell Prolif. 36 Suppl 1 (2003) 33-44.

[37] I.I. Patel, W.J. Harrison, J.G. Kerns, J. Filik, K. Wehbe, P.L. Carmichael, A.D. Scott, M.P. Philpott, M.D. Frogley, G. Cinque, F.L. Martin, Anal. Bioanal. Chem. 404 (2012) 1745-1758.

[38] X.R. Yang, J.D. Figueroa, S.M. Hewitt, R.T. Falk, R.M. Pfeiffer, J. Lissowska, B. Peplonska, L.A. Brinton, M. Garcia-Closas, M.E. Sherman, Breast Cancer Res. Treat. 137 (2013) 837-847.

[39] J.L. Kelsey, D.B. Fischer, T.R. Holford, V.A. LiVoisi, E.D. Mostow, I.S. Goldenberg, C. White, J. Natl. Cancer Inst. 67 (1981) 327-333.

[40] S.P. Helmrich, S. Shapiro, L. Rosenberg, D.W. Kaufman, D. Slone, C. Bain, O.S. Miettinen, P.D. Stolley, N.B. Rosenshein, R.C. Knapp, T. Leavitt Jr, D. Schottenfeld, R.L. Engle Jr, M. Levy M. Am. J. Epidemiol. 117(1983) 35-45.

[41] J.L. Kelsey, M.D. Gammon, E.M. John, Epidemiol. Rev. 15 (1993) 36-47.

[42] A.R. Howlett, M.J. Bissell, Epithelial Cell Biol. 2 (1993) 79-89.

[43] L. Ronnov-Jessen, M.J. Bissell, Trends Mol. Med. 15 (2009) 5-13.

[44] M. Hu, K. Polyak, Curr. Opin. Genet. Dev. 18 (2008) 27-34.

[45] T. Vargo-Gogola, J.M. Rosen, Nat. Rev. Cancer **7** (2007) 659-672.

[46] J.D. Figueroa, R.M. Pfeiffer, D.A. Patel, L. Linville, L.A. Brinton, G.L. Gierach, X.R. Yang, D. Papathomas, D. Visscher, C. Mies, A.C. Degnim, W.F. Anderson, S. Hewitt, Z.G. Khodr, S.E. Clare, A.M. Storniolo, M.E. Sherman, J. Natl. Cancer Inst. 106 (2014) dju286.

[47] M.J. Walsh, T.G. Fellous, A. Hammiche, W.R. Lin, N.J. Fullwood, O. Grude, F. Bahrami, J.M. Nicholson, M. Cotte, J. Susini, H.M. Pollock, M. Brittan, P.L. Martin-Hirsch, M.R. Alison, F.L. Martin, Stem Cells 26 (2008) 108-118.

[48] M.J. Walsh, A. Hammiche, T.G. Fellous, J.M. Nicholson, M. Cotte, J. Susini, N.J. Fullwood, P.L. Martin-Hirsch, M.R. Alison, F.L. Martin, Stem Cell Res 3 (2009) 15-27.

[49] J.G. Kelly, T. Nakamura, S. Kinoshita, N.J. Fullwood, F.L. Martin, Analyst 135 (2010) 3120-3125.

[50] I.I. Patel, D.A. Shearer, S.W. Fogarty, N.J. Fullwood, L. Quaroni, F.L. Martin, J. Weisz, Cancer Biol. Ther. 15 (2014) 225-235.

[51] S.W. Fogarty, I.I. Patel, J. Trevisan, T. Nakamura, C.J. Hirschmugl, N.J. Fullwood, F.L. Martin, Analyst 138 (2013) 240-248.

[52] J.P. Medema, L. Vermeulen, Nature 474 (2011) 318-326.

[53] G.R. Cunha, W.A. Ricke, Differentiation 82 (2011) 168-172.

[54] E.A. Ricke, K. Williams, Y.F. Lee, S. Couto, Y. Wang, S.W. Hayward, G.R. Cunha, W.A. Ricke, Carcinogenesis 33 (2012) 1391-1398.

[55] N.C. Purandare, I.I. Patel, J. Trevisan, N. Bolger, R. Kelehan, G. von Bünau, P.L. Martin-Hirsch, W.J. Prendiville, F.L. Martin, Analyst 138 (2013) 3909-3916.

[56] J. Trevisan, P.P. Angelov, P.L. Carmichael, A.D. Scott, F.L. Martin, Analyst 137 (2012) 3202-3215.

[57] M.J. Yaffe, J.G. Mainprize, Radiology 258 (2011) 98-105.

Legends to Figures

Figure 1 Study overview. (a) Diagram representing a lactiferous duct with an aggregation of TDLUs. The space between acini within the TDLU is occupied by intra-lobular stroma while the space between different TDLUs is occupied by inter-lobular stroma. (b) Diagrammatic interpretation of the terminal ductal lobular unit (TDLU), illustrating the types of cells that surround an acinus. (c) Diagrammatic description of the developmental progression of the TDLU with advancing age. (i) pre-pubertal, (ii) pubertal, (iii) Mature, (iv) lactating, (v) post-menopausal. (d) Example of principal component analysis and linear discriminant analysis (PCA-LDA) of TDLU, intra-lobular stroma and inter-lobular stroma. The tissue section was selected from a subject within the 40-49 y age group. (1) Parallel sections were stained with H&E for histological representation. (2) Numbered grid overlays were added to micrographs of the sections to aid spectral selection. (3) Image maps were produced from which spectra were extracted. (4) Class means representing spectral differences between different cell types. (5) PCA-LDA scores plots of different cells where each spectral point is derived from the average of 5 IR spectra.

Figure 2. PCA-LDA scores plots showing separation in every histological compartment tested between individuals in all age groups. The *x*-axis represents LD1 and the *y*-axis LD2. Each spectral point is derived from PCA-LDA of the average of 5 spectra. TDLU, Intra-lobular stroma and Inter-lobular stroma were examined separately. Each individual was processed as an independent category.

Figure 3. Loadings (LD1) plots derived from PCA-LDA comparing individuals from each age group for differences within each histological compartment. The *x*-axis is cm^{-1} and the *y*-axis represents absorbance coefficient. The five wavenumbers contributing to the most segregation were derived from the points furthest away from the *x*-axis.

Figure 4. Temporal changes in breast tissue. (A) PCA-LDA scores plots showing separation between age groups in every histological compartment. The *x*-axis represents LD1 and the *y*-axis LD2. (B) Loadings plots (LD1) showing the principal discriminating wavenumbers in graphical form. The *x*-axis is cm^{-1} and the *y*-axis represents absorbance coefficient.

Figure 5. Distinguishing cell types in TDLUs. (A) PCA-LDA 1D scores plots showing some separation between luminal and myoepithelial cells within the TDLU for the different age groups. The *x*-axis represents LD1 (B) loadings plots showing the principal discriminating numbers. The *x*-axis is cm^{-1} and the *y*-axis represents absorbance coefficient. (C) Tables of discriminating wavenumbers with their corresponding biochemical markers.

Figure 6. Temporal changes in individual cell populations in breast tissue. (A) PCA-LDA scores plots showing separation between different age groups for the TDLU compartment with a table of the principal discriminating wavenumbers and corresponding biochemical markers. (B) PCA-LDA scores plots showing separation between different age groups for the luminal and myoepithelial cell regions. The *x*-axis represents LD1 and the *y*-axis LD2. (C) Tables show the 5 major segregating wavenumbers with their corresponding biochemical markers.

Figure 7. Distinguishing putative stem cells. (A) 3-D scores plots of exploratory PCA using 10 first PCs examining luminal and myoepithelial cells separately to identify outliers which may represent residing stem cells. (B) Corresponding loadings plots for PC1 for luminal cells and PC3 for myoepithelial cells identify the major segregating wavenumbers. (C) Tables show the 6 major segregating wavenumbers with their corresponding biochemical markers.

Table 1. Spectra for each histological compartment for most patients were averaged by a factor of 2, 3 or 5 to aid visualisation in scores plots.

Patient/Averaging factor	TDLU	Inter-LS	Intra-LS	Myo-epithelial	Luminal
P1	2	1	1	2	1
P2	3	2	2	2	1
P3	3	2	2	2	1
P4	5	5	5	2	1
P5	5	5	5	2	1
P6	5	5	5	2	2
P7	5	5	5	2	2
P8	1	1	1	1	1
P9	2	1	5	1	1
P10	5	5	5	2	1
P11	1	2	1	1	1

Table 2. Principal segregating wavenumbers between individuals of each age group for each histological compartment. All wavenumbers were tested for significance using the *ANOVA* test with *P* <0.0001

Wavenumber (cm^{-1})/ Area	20-29 y			30-39 y			40-49 y		
	TDLU	Intra-LS	Inter-LS	TDLU	Intra-LS	Inter-LS	TDLU	Intra-LS	Inter-LS
1,780*						X		X	
1,709: C=O					X				
1,693: Amide I (C=O)									
1,671: Amide I (C=C)		X					X		
1,650: Amide I (C=O stretch and C-N stretch)							X		X
1,630: RHS Amide I (C=O stretch and C-N stretch)	X			X	X		X		X
1,550: Amide II (N-H bend and C-N stretch)			X	X					X
1,470-1,473: CH$_2$ bend	X	X		X	X				
1,451: CH$_3$	X				X				X
1,375: (C-N: cytosine, guanine)		X				X		X	
1,232: $\nu_{as}PO_2^-$					X				X

26

1,225: DNA $\nu_{as}PO_2^-$	X				X				X
1,200-1,210: RNA Ribose (C-O stretching)			X				X		X
1,140: Phosphate/ oligosaccharides			X			X		X	
1,080: DNA/RNA ($\nu_sPO_2^-$)		X			X	X	X	X	
1,053-1,063: (C-O: carbohydrates)		X					X		
1,040: Glycogen (C-O-H bond)		X					X		
1,018: (C-O, C-C, OCH)		X	X	X					
922: (Left Handed DNA)						X		X	

Table 3. Principal segregating wavenumbers between age groups for each histological compartment. All wavenumbers were tested for significance using the Anova test with $P < 0.0001$.

Wavenumber (cm^{-1})	TDLU	Intra-LS	Inter-LS
1,680: Amide I			X
1,650: Amide I (C=O stretch and C-N stretch)	X	X	
1,550: Amide II (N-H bend and C-N stretch)		X	X
1,495: C=C			X
1,456- 1,460: CH$_3$ methyl groups	X	X	
1,375: C-N cytosine guanine		X	
1,238-1,242: $\nu_{as}PO_2^-$	X		X
1,225: DNA (O-P-O asymmetric stretch)	X		
1,219: $\nu_{as}PO_2^-$			X
1,200: RNA Ribose (C-O stretching)			
1,080: DNA/RNA (O-P-O stretching)	X		
1.061: C-O deoxyribose			X
1,040: Glycogen (C-O-H bond)	X	X	

Figure 1

(a) Intralobular stroma, TDLU, Duct, Interlobular stroma

(b) Myoepithelial cell, Lactiferous duct, Luminal cell, Acinus, Terminal Ductal Lobular Unit

(c) (i)Prethelarche (ii)Prepupertal (iii)Mature (iv)Lactating (v)Involuting

(d) H&E, Micrograph, Image map, Spectrum, Scores Plot

Inter-LS
Intra-LS
Myo-epithelial
Luminal
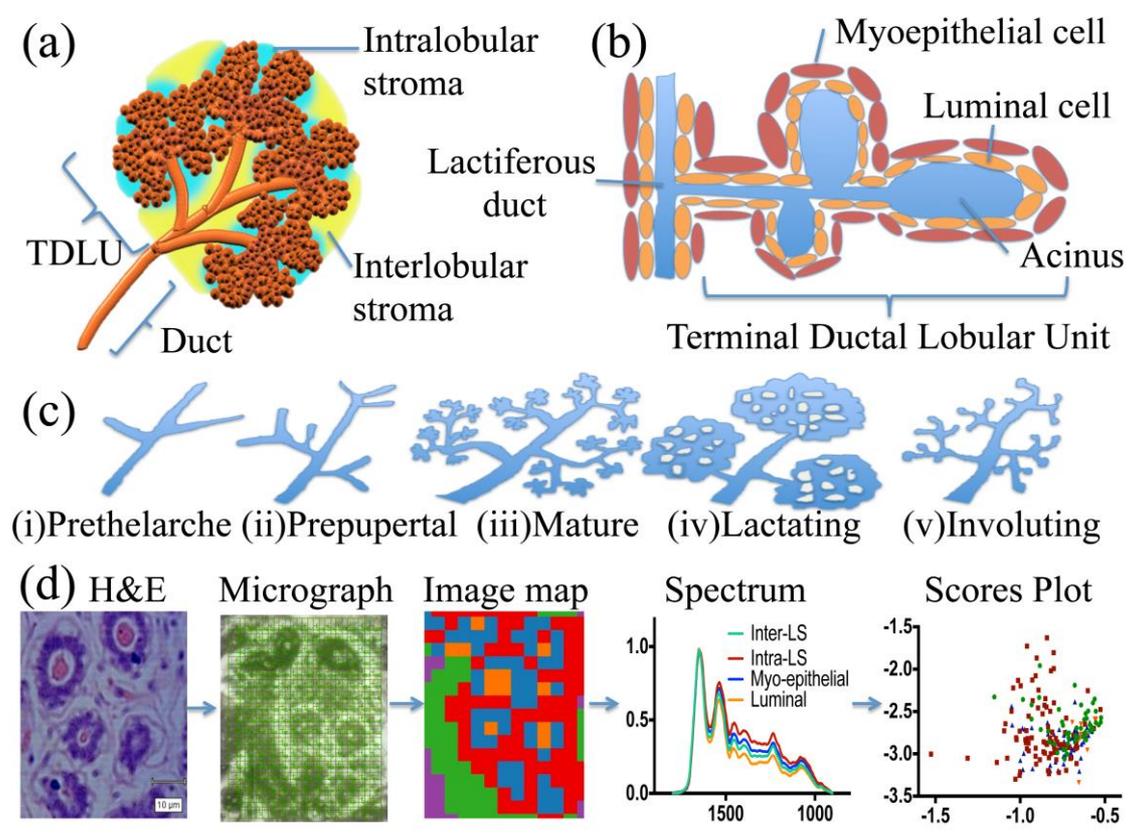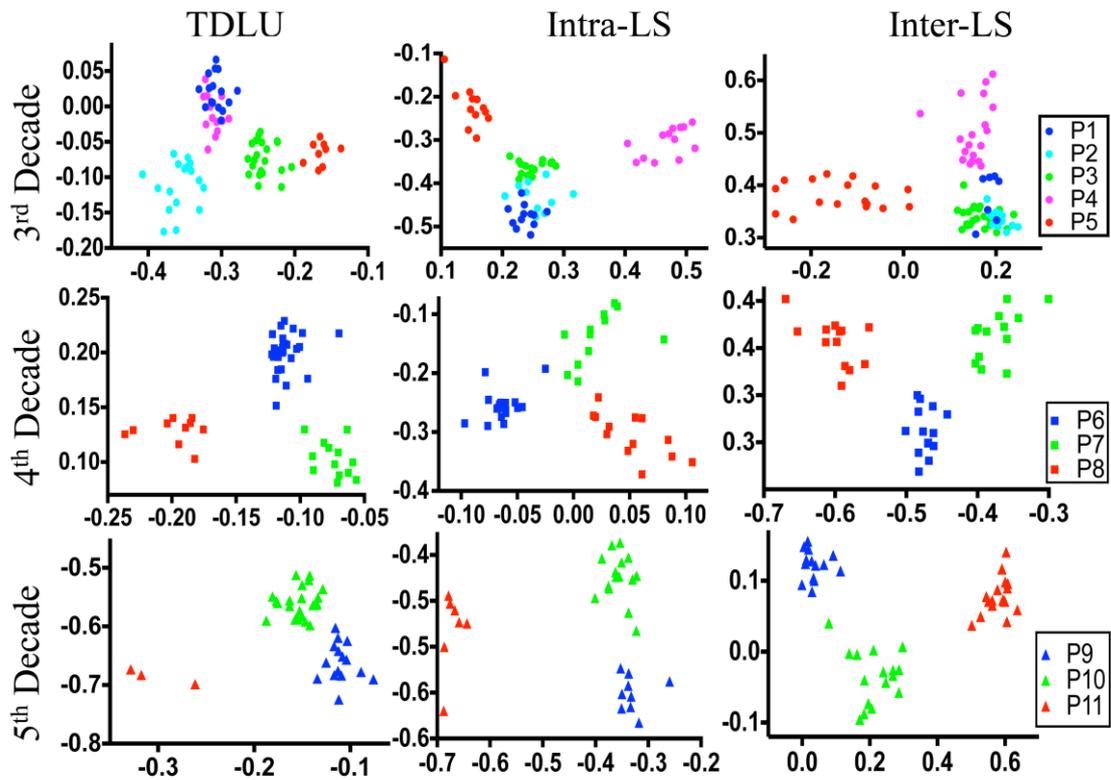
Figure 2

Figure 3

Figure 4

Figure 5

Wavenumber (cm-1)	Biochemical marker
1630	Amide I
1500	Amide II
1711	C=O bond
1448	Protein asymmetric CH$_3$ bending
939	Carotenoid
991	C-O ribose

Wavenumber (cm-1)	Biochemical marker
1550	Amide II
1456	CH$_3$ bending in proteins
1647	Amide I
1018	Polysaccharides
908	Phosphodiesters
1751	Fatty acids

Wavenumber (cm-1)	Biochemical marker
1456	CH$_3$ bending in proteins
1676	Amide I
1082	Phosphate band
1626	Amide I
1230	Nucleic acid phosphate vibration
1196	Phosphate band

Figure 6

(A) TDLU

(B) Luminal   Myoepithelial

(C) TDLU

Wavenumber (cm-1)	Biochemical marker
1051	DNA-RNA (C-O-C stretching)
1647	Amide I
1020	DNA
1236	$v_{as}PO_2$ of nucleic acids
1456	CH_3 in lipids and proteins

Luminal   Myoepithelial

Wavenumber (cm-1)	Biochemical marker	Wavenumber (cm-1)	Biochemical marker
1660	Amide I	1053	Phosphate (P-O-C, C-OH stretching)
1460	$\delta_{as}CH_3$ collagen		
1539	Amide II	1649	Amide I
1063	Stretching C-O Ribose	1238	PO_2 stretching
		1547	Amide II
943	Carotenoid	1094	PO_2 -

Figure 7

Luminal cells ← TDLU → Myoepithelial cells

(A)

(B)

(C)

Wavenumber	Biomarker
~1630	Amide I
~1535	Amide II
~1480	Amide II
~1710	C=O bond
~1590	C=N bond
~1425	Polysaccharides

Wavenumber	Biomarker
~1643	Amide I
~1508	Amide II
~1569	Amide II
~1469	Protein
~1504	C=O bond
~1226	O-P-O as. stretch