1 Title Page

2 Title

3 Comprehensive profiling of liver x receptor splicing in triple negative breast cancer
4 reveals existence of novel splice variants that are prognostic for survival

5 Authors

6 Priscilia Lianto¹, J. Bernadette Moore¹, Thomas A. Hughes², James L. Thorne^{1,€}.

7 ¹School of Food Science and Nutrition, University of Leeds, Leeds, LS2 9JT, UK.

8 ²School of Medicine, University of Leeds, Leeds, LS9 7TF, UK.

9 €Corresponding author

10 | Highlights

- 11 • Expression of full length LXR α is associated with shorter disease-free survival of triple
12 negative breast cancer patients
- 13 • A systematic evaluation of cell lines and primary tumour samples indicates LXR splicing
14 is extensive in breast cancer
- 15 • Confirmation of three new LXR splice variants at transcript and protein level
- 16 • Expression of full length LXR β or LXR α splice variants that harbour truncated ligand
17 binding domains are associated with better prognosis

- Expression of LXR target genes is positively correlated with LXR α in relapsed patients and inversely correlated with LXR β in survivors.

Abstract

The liver x receptors (LXR) alpha and beta are ligand-responsive transcription factors that link homeostatic control of lipid metabolism with cancer pathophysiology and prognosis. LXR activity is elevated in triple negative breast cancer relative to other breast cancer subtypes, driving gene signatures associated with drug resistance and metastasis. The loci encoding LXR α and LXR β produce multiple alternatively spliced proteins, but the true range of variants and their relevance to cancer remain poorly defined. Seven splice variants of LXR α or LXR β were detected. Three have not been recorded previously and five were prognostic. High expression of full length LXR α was associated with shorter disease-free survival but splice variants harbouring truncations of the ligand binding domain were prognostic for improved survival. All LXR β variants were associated with longer disease-free survival. Mechanistically, while full length LXR α positively correlated with target gene expression in primary samples, LXR β was inversely correlated. We conclude that canonical LXR α function is an oncogenic driver of triple negative tumour pathophysiology that can be countered by high expression of truncated splice variants and/or full length LXR β .

1. Introduction

Breast cancer (BCa) is the most commonly diagnosed cancer in women in the UK and is the cause of ~600,000 cancer death worldwide [1]. The triple-negative BCa (TNBC) subtype has higher rates of recurrence in the first three years after prognosis and increased mortality rates [2]. TNBC is defined by low oestrogen (ER), progesterone (PR), and Her2 receptor expression. Due to this lack of expression of therapeutic molecular targets primary disease can only be systemically treated with cytotoxic chemotherapy and TNBC remains a cancer of unmet clinical need; novel targeted therapies are urgently needed.

Nutritional and epidemiological studies indicate that cholesterol metabolism may play a role in the aetiology and severity of TNBC [3-6]. The liver x receptors (LXR α /NR1H3; LXR β /NR1H2) are homeostatic regulators of cholesterol metabolism and as such have been suggested as potential therapeutic targets. LXR α activation by hydroxylated cholesterol (oxysterols) has been linked to poor survival by driving multi-drug resistance in TNBC patients [7], and increased metastasis in mouse models [8, 9]. On the other hand, LXR β activation by histamine conjugated oxysterols such as dendrogenin A can induce lethal autophagy in breast cancer [10]. Intriguingly, ER-negative and ER-positive BCa subtypes respond differently to LXR ligands [11] even though oxysterol concentrations are similar between BCa subtypes [12]. Collectively, these data imply that the differential control over LXR's transcriptional regulation between subtypes is not simply at the level of ligand type or concentration.

LXR α and LXR β are each thought to be translated from their own single main transcript variants, producing 447 and 460 amino acid proteins respectively [13, 14].

These ‘full-length’ isoforms harbour domains that provide distinct functions, including activation function 1 (AF1), hinge region (H), DNA binding domain (DBD) and ligand binding domain (LBD) [13], as is typical for many ligand dependent nuclear receptors.

Aside from the canonical full-length LXR α 1, four additional LXR α splice variants have been reported in human cell lines or tissue samples to date [15, 16], and a single report found one alternative splice variant to the major LXR β 1 isoform [17]. These studies found alternative splicing of LXR disrupts the integrity of the DNA and ligand binding domains, rendering the proteins with reduced or even absent response to ligand [13, 14]. In the case of LXR β , a single known splice variant functions as an RNA co-activator [17]. Genomic and proteomic databases, such as NCBI, TCGA Splicing Variants database (TSVdb), ENSEMBL, and UNIPROT indicate that both LXR α and LXR β are far more extensively spliced than has been reported. An analysis of splicing in the nuclear receptor superfamily predicted LXR α has 62 different transcript variants [18], making it the most extensively spliced nuclear receptor.

However, experimental evidence for the existence of such an array of isoforms is lacking. Given the strong links between cholesterol metabolism, for which the LXRs play vital regulatory roles, and TNBC aetiology, the aim of this study was to describe the LXR splice repertoire and clinical significance in human triple negative breast cancer.

2. Materials and Methods

2.1 Systematic transcript variant analysis in public databases

The NR1H3/LXR α and NR1H2/LXR β spliced variant sequences were assessed in NCBI and ENSEMBL databases. Sequence annotation records of LXR α and LXR β spliced variant read from databases were stored in a feature library created with apE plasmid editor. This feature library can then be used to scan any nucleotide and/or amino acid sequence that are complementary to all alternative forms of the LXR α and/or LXR β sequences recorded in this library. The number of nucleotides in each exon and intron were counted in ApE plasmid editor to determine the proportions for drawing squares and lines represented exons and introns. The schematic diagrams of LXR variant structures were drawn using AutoCAD (Autodesk, US) software are shown in **Fig1**, **SF1**, and naming convention used in this study is described in Supplementary Information Section 1: Isoform naming rationale. Variants are listed in **TableS1 (LXR α)** and **TableS2 (LXR β)**.

2.2 The Cancer Genome Atlas (TCGA) Splicing Database Analyses

TCGA splicing variants database (TSVdb) was searched and LXR splicing expression data in BCa tumour tissue, reported as normalised RNA-Seq by Expectation Maximisation (RSEM) values, were downloaded using the TSVdb webtool (<http://www.tsvdb.com>) [19]. Unreported and/or missing clinical status of the deposited TCGA BCa tumour samples were replenished by assessing patient tumour sample information from cBioportal (<http://cBioportal.org>) [20].

2.3 Human samples: ethical approval, collection, and processing

Thirty-eight frozen tumour samples were obtained with ethical approval from the Leeds Breast Research Tissue Bank (15/HY/0025 and LBTB_TAC_1/17). The clinicopathological features and selection criteria of these cohorts have been described previously [7, 11, 12, 21] (**Table-S3**). The Leeds TNBC tumour cohorts were grouped based on their disease-free survival (DFS) status. TNBC patients who did not have disease relapse during follow-up (median follow-up time was 96 months) were categorised to 'No Event' group. Meanwhile, TNBC patients who had relapsed and/or died due to their disease were grouped to an 'Event' group (median follow-up time was 20 months).

2.4 Cell culture

The HEPG2, MCF7, BT474, MDA.MB.231, MDA.MB.468, MDA.MB.453, and MDA.MB.157 cell lines were originally obtained from ATCC (Manassas, USA). Cell lines were routinely cultured in Dulbecco's Modified Eagle Medium (DMEM; Thermo Fisher, UK, Cat. 31966047), supplemented with 10% Fetal Bovine Serum (FBS; Thermo Fisher, UK, Cat. 11560636) and incubated in 37°C, 5% CO₂. Cells were periodically checked and confirmed to be mycoplasma free.

2.5 siRNA transfection

siRNA transfection was performed as previously reported [22]. Briefly, 1x10⁵ cells were plated in 6-well plates and incubated overnight. Lipofectamine RNAiMAX (Thermo Fisher, UK, Cat. 13778030), siLXR α (Origene, USA, Cat. SR322981) and siLXR β (Origene, USA, Cat. SR305039) or the scrambled siRNA (Origene, USA, Cat.

SR30004) were diluted in OptiMeM (Thermo Fisher, UK, Cat. 31985062), and added to the cells at a final concentration of 30 nM. The cells were incubated and after 20 h the media was removed and fresh DMEM added. Knockdown was confirmed at the protein level 48 h post-transfection.

2.6 mRNA extraction, cDNA synthesis, and qPCR

Total RNA was isolated from approximately 5×10^5 cells using the ReliaPrep™ RNA Cell Miniprep System (Promega, UK, Cat. Z6012) following product guidelines, including on column digestion of gDNA. Cytoplasmic RNA was isolated from approximately 2×10^6 cells using the Cytoplasmic & Nuclear RNA Purification Kit (Norgen, Belmont, CA, USA, Cat. 21000). The concentration of isolated RNA was determined spectrophotometrically, and RNA purity was evaluated by 260/230 nm and 260/280 nm ratios using a CLARIOstar plate reader (BMG LABTECH, Germany). cDNA was synthesised from 2 µg RNA using the GoScript™ Reverse Transcription kit (Promega, UK, Cat. A5003). Primers were designed using NCBI BLAST primer design or, where transcript variants were not available in the NCBI database, primer 3 software was used [23]. Genomic DNA was digested, and cytoplasmic RNA was extracted to ensure intron retention was not confused with gDNA or incompletely spliced transcript. Amplicon size was restricted between 80 to 150 bp and amplicons were required to span an exon-exon boundary. Primer efficiency was measured from qPCR standard curves, and primers were redesigned if amplification efficiency did not fall within 90-110% and/or a single peak was not observed in melting temperature analysis. A description of primer design strategy is provided in Supplementary

Information Section 2: Primer design strategy. Primer positions are shown in **SF2** and sequences provided in **Table-S4**.

2.7 Protein lysate extraction

Five mg fresh-frozen tumour biopsy samples were homogenised with a 1 mL Dounce tissue grinder (SLS, UK, Cat. HOM3580). Cell pellets and/or tumour samples were lysed in RIPA buffer (10 mM Tris-HCl pH 8, 140 mM NaCl, 0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate, 1 mM EDTA, 0.5 mM EGTA), with 1 mM PMSF (Thermo Fisher, UK, Cat. 78440) added fresh prior to use. The lysed pellets were incubated on ice for 5 min and centrifuged at 11,500xg at 4°C for 10 min. The protein lysates concentrations were determined using a BCA kit (Thermo Fisher, UK, Cat. 23227) and lysates were kept in -80°C until further analyses. Each tumour sample was split in two and duplicate extractions were made. For each protein quantified (see 2.9) both duplicates were run and the average of each was used for analysis. Concordance of duplicates was high and is shown in **SF3**.

2.8 Cytoplasmic and Nuclear Protein Extraction

For the cytoplasmic and nuclear protein extractions, the REAP (Rapid Efficient And Practical) protocol [24] was followed with a slight modification. In brief, 1×10^7 cells were suspended in 1 mL ice cold PBS with 1 mM PMSF. 200 μ L of cell lysate was taken, put into a new chilled Eppendorf tube and centrifuged at 1,500 rpm for 3 min at 4°C. Supernatant was removed. The cell pellet (from 200 μ L cell lysate) was resuspended in RIPA buffer with 1 mM PMSF (Thermo Fisher, UK, Cat.

78440) added fresh prior to use and labelled as a “whole cell fraction”. The remaining 800 μ L cell lysate was centrifuged at 11,500xg at 4°C for 10 s. The supernatant was removed. The cell pellet (from 800 μ L cell lysate) was resuspended in 200 μ L ice-cold PBS+0.1 NP40 with 1 mM PMSF and centrifuged at 11,500xg at 4°C for 10 s. The supernatant was taken into a new chilled Eppendorf tube and labelled as a “cytoplasmic fraction”. The cell pellet was then resuspended in 200 μ L ice-cold PBS+0.1 NP40 with 1 mM PMSF and centrifuged at 11,500xg at 4°C for 10 s. The supernatant was removed and cell pellet resuspended in 50 μ L RIPA buffer with 1 mM PMSF. Cell lysate was sonicated with a water-bath sonicator (Diagenode Bioruptor Pico, Belgium, Germany) with cycles of 10 s on and 10 s off for five cycles, followed by centrifugation at 11,500 g at 4°C for 10 min. The supernatant was taken, put into a new chilled Eppendorf tube, and labelled as a “nuclear fraction”. The protein lysates concentrations were determined using a BCA kit and lysates were kept in -80°C until further analyses.

2.9 Immunoblotting

Forty-five micrograms of protein lysate combined with NUPAGE LDS sample loading buffer (Thermo Fisher, UK, Cat. NP0007) and DTT reducing agent (Thermo Fisher, UK, Cat. NP0004) was heated at 70°C for 10 min. For LXR variant expression, the protein lysate was loaded onto a 10% SDS polyacrylamide gel, electrophoresed at constant 80 V for 150 min, and transferred onto a PVDF membrane (Merck, UK, Cat. IPFL00010). The membrane was then blocked with TBS Odyssey Blocking Buffer (LI-COR Biosciences, UK, Cat. 92750000) for 1 h. Proteins were probed with anti-LXR α

(R&D Systems, USA, Cat. PP-PPZ0412-00, dilution 1/1000), anti-LXR β (Active Motif, Germany, Cat. 61177, dilution 1/1000), and anti-HPRT (Santa Cruz Biotechnology, Cat. sc-376938, dilution 1/100) overnight at 4°C. The membrane was then blocked and probed with LICOR secondary antibodies (IRDye 800CW goat anti-mouse Cat. 926-68170, IRDye 680RD goat anti-rabbit Cat. 926-68071; dilution 1/15,000, LI-COR Biosciences, UK) for 1 h and signal was visualised using the Odyssey system (LI-COR Biosciences, UK). Densitometry was performed using Image Studio™ Lite (LI-COR Biosciences, UK) software.

2.10 Immunoprecipitation

One mg protein sample was precleared with Dynabeads Protein A (Thermo Fisher, UK, Cat.100002D) and 1 μ g isotype IgG2a (Cell Signalling, USA, Cat. 61656S). Immunoprecipitation was performed by incubating 40 μ L of Dynabeads Protein A coupled to 2 μ g of anti-IgG2a or 2 μ g of anti-LXR α using bis(sulfosuccinimidyl)suberate (Thermo Fisher, UK; Cat. A39266) with 1 mg protein sample overnight at 4°C. Dynabeads were washed 3 times with 10 mM Tris-HCl, 50 mM KCl (pH 7.5) and sample eluted in 20 μ L of NUPAGE LDS sample loading buffer containing 100 mM DTT, then heated at 70°C for 10 min 10 at 70°C. The supernatant was transferred to a new Eppendorf tube after being separated from Dynabeads using a magnetic separator (Promega, UK, Cat. CD4002).

2.11 In-silico peptide mass prediction

The amino acid sequences of LXR variants downloaded from NCBI, ENSEMBL, and/or UNIPROT databases were subjected to the PeptideMass tool (<http://www.expasy.org/tools/peptide-mass.html>; Swiss Institute of Bioinformatics, Switzerland) [25] to theoretically trypsin digest the protein sequence in silico. A set of peptides from each LXR variant were then subjected to Diagram Venn (<http://bioinformatics.psb.ugent.be/webtools/Venn>; Bioinformatics & Evolutionary Genomics, Belgium) in order to find the unique peptides of each LXR variant.

2.12 S-Trap column coupled Mass Spectrometry (MS)

Protein samples were processed using the S-TRAP Micro column (PROTIFI, NY, USA) following the manufacturer's instructions. Proteins were fully solubilised by adding 20 μ L of 10% SDS solution to 20 μ L sample in RIPA buffer. Reduction and alkylation were then performed. DTT was added to a final concentration of 20 mM before heating to 56 °C for 15 min with shaking. The sample was left to cool for 5 min then iodoacetamide was added to a final concentration of 40 mM, before heating to 20 °C for 15 min with shaking in the dark. Phosphoric acid was then added to a final concentration of 1.2%, to ensure destruction of all enzymatic activity and maximise sensitivity to proteolysis. Samples were then diluted with S-Trap binding buffer (100 mM TEAB pH 7.1 in methanol), and 1 μ g of trypsin, reconstituted in 50mM triethylammonium bicarbonate (TEAB), was added before the sample was quickly loaded onto the S-trap column. Proteins were captured within the submicron pores of the three-dimensional trap. Proteins captured within the trap present exceptionally high surface area allowing them to be washed free of contaminants. The

S-trap was washed by adding 150 μ L binding buffer before being spun at 4000xg for 30 s. 30 μ L of 0.02 μ g/ μ L Trypsin (Promega, WI, USA) was then added to the top of the S-trap. S-traps were loosely capped and placed in a 1.5mL Eppendorf and heated to 46 °C for 15 min with no shaking. Digested peptides were eluted by first spinning the S-trap at 4,000 g for 1 min. Further elution was performed in 40 μ L 50mM TEAB, 40 μ L 0.2% formic acid, and 30 μ L 50% acetonitrile with 0.2% formic acid prior to centrifugation. Eluates were combined then dried down prior to resuspension in 0.2% formic acid.

2.13 Statistical analyses

All statistical analyses were performed using GraphPad Prism v8. The expression of LXR splice variants in TCGA tumours compared with the adjacent normal tissues was assessed using a Mann-Whitney two-tailed U test. Differential LXR splicing expression levels in both TCGA BCa and Leeds TNBC tumour samples were established using multiple t-tests with Holm-Sidak for multiple correction. The knockdown siRNA experiment was analysed using a two-tailed one-way ANOVA. The relationship between protein isoform variants and their mRNA transcripts were assessed using Spearman's correlation and linear regression. ROC curves were used to establish the expression cut offs for high and low expression levels of each LXR variant. Tumour expression of each LXR variant was then assessed alongside patient survival to assess whether expression is predictive of survival in Kaplan Meier graphs. Patient survival was analysed using log-rank test.

3. Results

3.1 Database analysis of LXR transcript splice variance in breast cancer

To evaluate the variety of LXR splice variants, the NCBI and ENSEMBL databases were mined for mRNA transcript variants and UNIPROT database for protein variants. In total, this analysis indicated there were 64 LXR α transcript variants, of which 48 could code for 26 LXR α protein variants. We found 11 LXR β transcript variants that could code for 9 LXR β proteins (**SF1**). These included all nine variants (α 1- α 5, β 1- β 4) later observed in breast cancer cells and/or breast tumour tissue (summarised in **Fig1**).

To investigate the expression of LXR splice variants in BCa, we first examined RNA-seq data from 1,103 BCa tumours in the TCGA Pan-Cancer Atlas utilising the TSVdb web-interface [19]. These data showed that out of six LXR α and three LXR β splice variants with RSEM reads [26], the α 1.1 (median RSEM value=231.52) and β 1.1 (median RSEM value =1158.53) variants were most highly expressed in primary BCa samples (**Fig2A**). The α 3.1 (median RSEM value=115.42), α 1.2 (median RSEM value=31.88) and β 2 (median RSEM value=21.69) transcripts were expressed at low levels. No other transcript variants were detected in this database (**Fig2A**).

We further examined expression of the four variants detected at highest levels in matched tumour and normal samples from ER+ and TNBC subtypes. Both α 1.1 and α 1.2 were expressed at significantly higher levels in ER+ and TNBC tumour tissue compared to adjacent normal tissue ($P < 0.01$ for all; **Fig2B**). α 3.1 expression was lower

in ER+ tumour tissue than adjacent normal ($P=0.011$; **Fig2B**) but there was no difference in expression of this isoform between tumour and normal tissue in TNBC disease. There was no difference in $\beta 1.1$ expression in either ER+ or TNBC tumours relative to adjacent normal ($P>0.05$; **Fig2B**). There was no difference in expression of any individual LXR variants between ER+ and TNBC tumours (**SF4**).

We then examined if transcript variant expression identified in TSVdb was associated with disease-free survival (DFS) in ER+ (**Fig2C**) or TNBC patients (**Fig2D**). High $\alpha 1.2$ expression was associated with shorter DFS in ER+ patients ($p=0.029$), while in TNBC patients, high $\alpha 1.1$ expression was associated with shorter DFS ($p=0.04$). Neither $\alpha 3.1$ or LXR β expression was associated with DFS. These data suggest that elevated LXR $\alpha 1$ ($\alpha 1$) may be linked to poor prognosis in breast cancer patients.

3.2 LXR expression in breast cancer cell lines

Having established evidence for LXR splice variant expression in the TSVdb dataset, we next examined transcript expression in a panel of BCa cell lines. HEPG2 cells express relatively high levels of both LXR α and LXR β [27] so were included as a positive control. Total LXR α and LXR β RNA expression was determined using SYBR green primers that target the exon 9-10 junction of LXR α and exon 8-9 junction for LXR β . These exons are spliced together in every previously reported variants and in all variants described in this study (**Table-S1+S2**; **SF1-A-B**) and were used to estimate the total pool of LXR transcripts and for normalisation of variant expression.

The claudin high TNBC cell line MDA.MB.468 had highest expression of LXR α of all BCa cell lines ($P < 0.05$). MDA.MB.468, MDA.MB.453 and BT474 had highest LXR β (**Fig3A-B**). In all cell lines analysed, the expression of LXR α mRNA was significantly lower than LXR β (all $p < 0.05$; **SF5**), recapitulating the observations from the TSVdb (**Fig2A**). Across the cell lines, the transcripts measured accounted for between 91-100% of LXR α (exon 9-10) and 94-102% of LXR β (exon 8-9) exon-exon boundaries (**Fig3C-D**), indicating other variants were expressed at very low levels or did not contain the exon 9-10/7-8 boundaries. Across all cell lines we detected five mRNA species ($\alpha 1.1$, $\alpha 1.3$, $\alpha 2.3$, $\alpha 3.1$, $\alpha 5.3$; **Fig3C**) predicted to code for four LXR α protein variants ($\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 5$), and two coding for LXR β ($\beta 1$, $\beta 4$; **Fig3D**). LXR $\alpha 5$ and LXR $\beta 4$ have not been reported in the literature previously.

$\alpha 1$ transcripts were dominant (50-80% of all LXR α) in HEPG2 and TNBC cells (**Fig3C**: note black and dark grey sections corresponding to $\alpha 1.1$ and $\alpha 1.3$ respectively). In the ER+ cell lines $\alpha 5$ (40-50%) and $\alpha 3$ (30-40%) were the majority species (**Fig3C**: note light grey and white sections corresponding to $\alpha 5.3$ and $\alpha 3.1$ respectively). $\alpha 2$ comprised between 5-20% depending on the cell line (**Fig3C**: hatched sections). LXR $\beta 1.1$ was detected in all BCa cell lines, with a very small amount of $\beta 4$ transcript found in HEPG2 and the ER+ cell lines only (**Fig3D**). We successfully detected over 90-100% of all the LXR variants (harbouring the exon 9-10 and 7-8 junctions) present in the cell lines. From these data, we concluded that full-length $\alpha 1$ was the dominantly expressed variant in TNBC cells, truncated $\alpha 5$ was the dominant LXR α isoform in ER positive cells, and $\beta 1$ was the dominant LXR β transcript across all cell types.

We next performed immunoblotting to establish the range of protein variants present. Representative blots and densitometry analysis show bands corresponding to predicted sizes of LXR α 1 ($\approx$ 50kDa) and LXR β 1 ($\approx$ 55kDa) were robustly and reproducibly detected in all cell lines (**Fig3E** and **Fig3F** respectively). However, optimisation of immunoblotting experimental conditions, including antibody choice, indicated multiple additional bands were present when probing for LXR α (**Fig3E**). These bands closely corresponded to sizes of proteins that would be coded by the transcripts identified above, namely: α 1 (50 kDa), α 2 (44 kDa), α 3 (46 kDa), and α 5 (39 kDa). The cell line specific pattern of protein expression matched that of the transcript expression; α 1 RNA and protein were highly expressed in HEPG2 and the TNBC cell lines while α 5 was dominant in the ER+ cells (**Fig3G**). For LXR β only a single isoform β 1 was present at both RNA and protein level (**Fig3H**).

We concluded that in TNBC cell lines full-length LXR α 1 is the most abundant isoform comprising 50-80% of LXR α protein (depending on cell line), with two α 1 transcript variants accounting for 51-75% of LXR α transcript. LXR α 3, which lacks part of the AF-1 domain and has diminished response to ligand [15] was the next most abundant producing 15-30% of the protein and 8-27% of the transcript. LXR α 2, which is non-responsive to ligand [15], made up 8-20% of the total LXR α protein coded by two transcripts producing 7-12% of the LXR α coding RNA (see **Supplementary Information Section 1** for variant nomenclature). Only a small amount of LXR α 5 was detected in TNBC (<10% in MDA.MB.453 cells and undetected in other TNBC lines). Interestingly, this 'double-truncation' variant that lacks both the AF-1 region of α 3 and

the LBD region of $\alpha 2$, was the most abundant isoform in ER+ cells at protein (35-70%) and transcript level (43-54%).

3.3 LXR splice variant expression and clinical significance in triple negative breast cancer

To evaluate LXR variant expression and significance in clinical samples, LXR protein expression was measured in a cohort of 38 fresh-frozen TNBC tumour samples (mean follow-up of three years), which has been reported on previously [11, 12] (Table-S3). Patients were dichotomised into two groups: 'No Event' patients defined as those who were alive and disease free at the time of last reporting or had died from unrelated reasons; 'Event' patients were those who had died from their disease and/or had disease recurrence. Representative blots are shown in **Fig4A**. Event patients had significantly higher expression than No Event patients of both full-length isoforms ($\alpha 1$: $p < 0.0001$; $\alpha 4$: $p = 0.001$; **Fig4B**). Together $\alpha 1$ and $\alpha 4$ comprised 70% of the total LXR α protein in the No Event group, but this rose to 93% in the Event group (**Fig4C**). No significant difference was observed in level of $\alpha 2$, $\alpha 3$, $\alpha 5$, $\beta 1$, or $\beta 4$ protein variants between groups (**Fig4B**).

When patients were dichotomised using ROC analysis, high protein expression of the full-length variants were, as expected, found associated with significantly shorter disease-free survival ($\alpha 1$: $p = 0.0005$; $\alpha 4$: $p = 0.0079$; **Fig4D**). Interestingly, in this more nuanced analysis, high expression of $\alpha 5$ ($p = 0.044$), LXR $\beta 1$ ($p = 0.0023$), and LXR $\beta 4$ ($p = 0.037$) was associated with significantly longer DFS (**Fig4E**). These observations

were replicated when dichotomising based on RNA expression (**SF6A**). Furthermore, $\alpha 3.1$, which codes for intact LBD but harbouring a deletion in the AF1 domain, was also associated with shorter DFS ($p=0.022$; **SF6A**). The $\alpha 2.3$ transcript, which codes for a protein with the same LBD deletion as $\alpha 5$ was, like $\alpha 5$, associated with longer DFS ($p=0.031$; **SF6A**). $\alpha 2$ and $\alpha 3$ protein however were not associated with DFS (**SF6B**). We concluded that full length LXR α isoforms may exacerbate disease severity, whereas those lacking the full LBD or LXR β isoforms were associated with reduced disease severity.

3.4 Validation of isoform identity

3.4.1 Bands representing LXR variants were reduced by targeted siRNA

To confirm if the protein bands were LXR α variants, we first performed siLXR α treatments in cell lines. Using previously validated [7] siRNA duplexes (Origene trisilencers) against LXR α and LXR β we found that the protein bands predicted by size to be $\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 5$, and $\beta 1$ were all significantly reduced by targeted siRNA in all cell lines tested (all $p<0.05$; **SF7**). Note: $\alpha 4$ and $\beta 4$ were not expressed in cell lines so could not be validated in this way. Furthermore, siLXR α did not reduce LXR β expression and siLXR β did not reduce expression of any LXR α isoform, and, as is characteristic of LXRs [15, 28], all protein variants were localised in the nucleus (**SF8**).

3.4.2 Unique peptides representing $\alpha 4$ and $\beta 1$ were identified by MS S-trap

Mass Spectrophotometry

As was the case for cell line analyses, confirmation of the identity of the observed protein variants in tumour tissue was important, especially having identified the $\alpha 4$ variant which has not previously been reported in the literature. As siRNA was not possible on the tumour samples, we performed S-trap [29] coupled mass spectrometry (St-MS) in cell line and tumour lysates samples selected to represent as much of the diversity in isoforms as possible (**Table1**, **Table2**, **ST5**, **ST6**). This included two tumours in duplicate, ligand treated MDA.MB.468 cells (**SF9A**), and in siCON, siLXR α , siLXR β (**SF9B**). We found 15 peptides uniquely corresponding to $\alpha 4$ BLAST sequence alignment indicated they were high-quality hits (100% identity), including the two first exons coding for the alternative AF1 domain in this isoform (**SF9C** – boxed amino acids) supporting our prior conclusion that $\alpha 4$ was expressed in some tumour samples. Unique peptides of $\beta 1$ were also detectable in our tumour sample (**SF9D** – boxed amino acids). However, unique peptides corresponding to $\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 5$, or $\beta 4$ were not produced by our enzymatic cleavage and in silico analysis indicated these isoforms would not be distinguishable from each other (**Table3**). Our next approach was to perform immunoprecipitation. We found this successfully enriched each isoform (**SF9E**) but limited amounts of tumour meant bands could not be visualised for excision on coomassie gels. We concluded that $\alpha 4$ and $\beta 1$ were present in tumour tissue.

3.4.3 Transcript-protein correlation analysis

Linear correlation analysis was performed in the seven cell lines comparing expression of each transcript variant against each protein band for which it could

potentially code. There was a strong and significant positive correlation between transcript variants and the protein isoform for which they were predicted to code: $\alpha 1$ protein correlated with $\alpha 1.1$ ($R^2=0.97$; $p<0.0001$; **SF10** top row) and $\alpha 1.3$ ($R^2=0.78$; $p=0.0084$; **SF10** top row). $\beta 1$ protein correlated with $\beta 1.1$ transcript ($R^2=0.99$; $p<0.0001$; **SF10** bottom row). During primer design it became clear that due to complexity arising from the large number of LXR coding transcripts, several transcripts were not distinguishable from each other due to sequence homology (see **Supplementary Information Section 1 and 2**, and denoted with asterix in **Fig3C**). Specifically, indistinguishable transcripts were: $\alpha 1.4$ and $\alpha 2.3$; $\alpha 3.1$ and $\alpha 5.1$; $\alpha 3.3$ and $\alpha 5.3$. We tested all potential protein isoforms for correlation with the ambiguous primer pairs: $\alpha 1.4/\alpha 2.3$ correlated with $\alpha 2$ protein ($p=0.0022$; $R^2=0.97$; **SF10** row 2) but not $\alpha 1$ protein ($p>0.05$); $\alpha 3.1/\alpha 5.1$ correlated with $\alpha 3$ protein ($p=0.034$; $R^2=0.63$; **SF10** row 3) but not $\alpha 5$ protein ($p>0.05$); $\alpha 3.3/\alpha 5.3$ correlated with $\alpha 5$ protein ($p=0.029$; $R^2=0.65$; **SF10** row 4) but not $\alpha 3$ protein ($p>0.05$). We also designed primers against the exon 5 and 7 junction to simultaneously recognise the $\alpha 2$ and $\alpha 5$ variants that lack exon 6 and a portion of the LBD. These $\alpha 2/\alpha 5$ primers detected RNA that correlated with $\alpha 5$ protein ($p=0.0065$; $R^2=0.8$; **SF10** fourth row) but not with $\alpha 2$ ($p>0.1$; **SF10** second row).

Next, although cytoplasmic extraction was not possible in frozen tumour samples as they had lost cellular substructure in the freezing process, given the number of replicates was larger ($n=38$) we considered testing for correlations a reasonable approach (**SF11**). As for cell lines, both $\alpha 1.1$ and $\alpha 1.3$ correlated with $\alpha 1$ protein

($p < 0.05$ for both); $\alpha 2.3$ correlated with $\alpha 2$ protein ($p = 0.036$); $\alpha 3.1$ correlated with $\alpha 3$ protein ($p = 0.056$); $\alpha 4.1$ correlated with $\alpha 4$ protein ($p = 0.0037$; note, this variant was not detected in cell lines); $\alpha 5.3$ ($p = 0.039$) and $\alpha 2/\alpha 5$ ($p = 0.054$) correlated with $\alpha 5$ protein. For LXR β , $\beta 1.1$ transcript correlated with $\beta 1$ protein ($p = 0.022$). $\beta 3$ and $\beta 4$ are almost identical in size (46 and 47 kDa respectively) so are indistinguishable by immunoblotting. $\beta 4$ transcript ($p = 0.031$) but not $\beta 3$ ($p > 0.05$) correlated with the band corresponding to $\beta 3/\beta 4$ protein (**SF11** bottom row) confirming this as LXR $\beta 4$. In combination with the siRNA in the cell line studies above and MS experiments, these correlative observations indicate five LXR α variants and two LXR β variants are expressed in TNBC.

3.5 LXR splice variants are differentially correlated with expression of target genes in Event and No Event patients

We previously reported that LXR α expression is positively correlated to target gene expression in ER-negative patients who had relapsed and/or died due to their disease (Event patients) but not those who survive disease free (No Event patients) [7]. Here, we tested the hypothesis that differential LXR splice variant expression between cancers may contribute to disease aetiology via their ability to control expression of gene targets. LXR splice variant expression was tested for correlations with established LXR target genes, ABCA1 and ABCB1 [7]. Expression of both full-length isoforms ($\alpha 1$ and $\alpha 4$) positively correlated with expression of both ABCA1 or ABCB1, but interestingly only in Event patients ($p < 0.05$ for all; **Fig5**). Strikingly, expression of $\beta 1$ was inversely correlated with both ABCA1 and ABCB1 but only in No Event patients

($p < 0.05$; **Fig5**). Statistically significant correlations between $\alpha 2$, $\alpha 3$, or $\alpha 5$ with target genes were not observed (**SF12**). From these data we concluded that while full length LXR α is associated with activation of gene targets in patients who relapse or die, LXR β is associated with inhibition of the same target genes in TNBC survivors.

Discussion

The objective of this study was to establish the repertoire, expression levels, and pathophysiological significance of LXR splice variants in breast cancer. In TNBC we found evidence that five LXR α and two LXR β splice variants are expressed at RNA and protein levels. Three of these isoforms, $\alpha 4$, $\alpha 5$, and $\beta 1$, are new to the literature and do not have validated UNIPROT records. Our data demonstrate that LXR splice variants have clinical significance in TNBC; full length LXR α ($\alpha 1$ and $\alpha 4$) is associated with shorter DFS, while LXR β and LXR α variants with truncated ligand binding domains were associated with longer DFS.

We provide several lines of evidence that the LBD is required for LXR α 's oncogenic potential. Firstly, high expression of the full-length $\alpha 1$ was strongly associated with shorter DFS at both transcript and protein level. High expression of two other isoforms, $\alpha 3$ (transcript only) and $\alpha 4$ (transcript and protein) were also associated with shorter DFS. Although both have disruptions to their AF1 domains, they are otherwise homologous to full length $\alpha 1$, including, importantly, an uninterrupted LBD. $\alpha 5$ skips the same exons as $\alpha 2$ and $\alpha 3$ and has the same deletions; 60 amino acids are missing from the LBD, and 45 amino acids of the AF1 domain are missing, respectively. These

truncations have been shown separately to diminish ($\alpha 3$) or completely abolish ($\alpha 2$) transcriptional activity [15]. Interestingly, high transcript expression of $\alpha 2$ was, like $\alpha 5$, linked to longer DFS. High $\alpha 3$ transcript on the other hand, like full length $\alpha 1$ and $\alpha 4$ was linked to shorter DFS. These observations suggest that splice variants with disrupted LBD are advantageous, but the protein can tolerate disruption or even partial deletion of the AF1 domain and still be associated with worse prognosis. $\beta 1$ (P55055-1) and $\beta 4$ (M0R2F9) are generated by the mutually exclusive inclusion or exclusion of exon 7, again leading to a change in the LBD. With LXR β however, high expression of either $\beta 1$ and $\beta 4$ was linked to longer DFS. The cellular role of this altered LBD remains to be determined.

Belorusova et al. reported peptides of LXR α comprising the H3 and partial part of H5 correlated to high ABCA1 expression [30]. In addition, the differential relationship between target genes and LXR α isoforms in different patient groups also points to an oncogenic role for LXR α and a tumour suppressor role for LXR β in TNBC tumours. Indeed, a beneficial role for LXR β has been reported previously. Both dendrogenin A and RGX-104 are tumour suppressors and selective LXR β agonists [10, 31, 32].

To the best of our knowledge, $\alpha 4$ (NM_001251934; NP_001238863; B4DXU5), $\alpha 5$ (NM_001363595.2; NP_001350524.1; B5MBY7), and $\beta 4$ (M0R2F9) have not been reported in the literature before. $\alpha 4$ has an alternative 21 amino acids in the start of the AF1 domain producing a unique peptide mass signature that was detected with MS. These data support its inclusion in UNIPROT with a Q13133 prefix. As this variant

has not been investigated before and we perform no functional studies herein, the significance of this partial AF1 substitution remains unknown. $\alpha 5$ was not detected by MS but its presumed protein band (39 kDa) was lost after siLXR α treatment, its transcript was robustly expressed, and protein and transcript expression were strongly correlated. Although a unique peptide for $\beta 4$ was predicted from *in silico* analyses it was not detected with MS, however, its transcript was robustly expressed, and protein and transcript expression were strongly correlated.

Some of our results are at odds with prior observations. We were unable to detect expression of XP_02430405X (previously termed $\alpha 4$ [16]), in any cell lines or tissues, at RNA or protein level, including in HEPG2 or MCF7 cell lines that were previously shown to express this isoform [16, 33]. Intriguingly, a variant previously reported to be $\alpha 5$ [16, 33] did not correspond to any of the 64 LXR α transcripts from NCBI, nor any LXR α UNIPROT entries (**SF2A**). In this study [16], the measurement of XP_02430405X and previously reported $\alpha 5$ relied on PCR primers that detected a retained intron between exon 6-7 and a retained intron between exon 7-8, respectively (**SF2A**). We went to extensive lengths to ensure primary RNA or gDNA did not contaminate our cDNA libraries, steps that surprisingly have not been reported in LXR splicing studies previously. Cytoplasmic RNA is preferred over total cell isolates when examining splice expression and this has been reported previously to increase sensitivity of splice junction detection [34]. Previous detection of XP_02430405X may therefore be due to amplification of incompletely processed transcripts or contaminating gDNA. During our primer design stage, we found that the PCR primers

previously used to measure $\alpha 1$ at the exon 2-3 junction [15, 16, 35] also detect $\alpha 2$ (**SF3**). In addition, primers previously used to measure $\alpha 2$ at the exon 5-7 junction [15, 16, 35] also detect the previously unreported $\alpha 5$ variant (**SF2**). Our correlation analyses in cell lines and primary tumour samples suggest the majority of exon 6 skipping transcripts actually code for $\alpha 5$ not $\alpha 2$. We also found $\alpha 5$ is significantly (20-fold) higher than $\alpha 2$ in MCF7 and equal to $\alpha 2$ in HEPG2, the same cell lines that in prior studies were used to show $\alpha 2$ was strongly expressed. It likely that previous reports have over-estimated the contribution of $\alpha 2$ and under-estimated the contribution of $\alpha 5$ to the LXR pool.

In summary, this study provides critical insight into the pathophysiology of LXR splicing in BCa. Our study clarifies the relative roles of the LXR α and LXR β isoforms and our data are consistent with the hypothesis that while full length LXR α is oncogenic, LXR β and truncated LXR α variant ($\alpha 5$) can act as tumour suppressors. We propose the existence of two new LXR α isoforms: $\alpha 4$ containing an alternative AF1 region, and $\alpha 5$ which lacks large sections of the AF1 and LBD, probably rendering it with dominant negative function.

Triple negative breast cancer is defined as a category of exclusion. The data presented here suggest that measuring LXR $\alpha 1$, LXR $\alpha 5$, and LXR β at RNA or protein level may allow sub-stratification of TNBC patients for increased accuracy of prognosis. Perhaps more importantly, given the array of pharmacological and dietary modulators of the LXR pathway, defining gene networks regulated by the opposing

oncogenic and tumour suppressor LXR variants will aid in the development of lifestyle and pharmacological interventions to reduce incidence and improve survival in this challenging cancer of unmet clinical need.

Figure Legends

Figure 1. Schematic diagram of (A) LXR α and (B) LXR β splice variants detected at RNA and protein level in this study. AF1=Activation Function 1; DBD=DNA Binding Domain; H=hinge; LBD=Ligand Binding Domain. The numbers right below the isoform domain boxes are represented the position of amino acids.

Figure 2. LXR Splicing events in BCa validated by TCGA Splicing Variants database (TSVdb). (A) Heatmap visualization of normalized RSEM (RNA-Seq by Expectation Maximization) value expression of LXR α and LXR β in 1,103 TCGA BCa samples. Note: Transcript uc002prv was only reported in TSVdb and is not recorded in NCBI, ENSEMBL, or UNIPROT databases. (B) Comparison of LXR transcript variant expression, for α 1.1, α 1.2, α 3.1, and β 1.1, in matched tumour and normal tissues grouped by BCa subtype, ER+ (n=78) and TNBC (n=18). Statistical analysis by Mann-Whitney two tailed U tests. $P \leq 0.05$ was considered significant. (C, D) Kaplan-Meier survival curves plotting disease free survival of TCGA BCa patients. The TCGA BCa samples (downloaded from TSVdb TCGA splicing database) were grouped by BCa subtype, (C) ER+ (n=803) and (D) TNBC (n=101). Survival curves of each subtype were separated into two groups, no event (ER+=675, TNBC=79) and event (ER+=128;

TNBC=22) based on their overall survival data reported in cBioportal. P-value is based on the log-rank test and data groups were considered significantly different if $p < 0.05$.

Figure 3. Diversity of LXR splice variants in breast cancer cell line panel. RNA analyses of total LXR α (A) and LXR β (B) expression. The stacked bar graphs show LXR α (C) and LXR β (D) transcript variants normalised to ubiquitously expressed exon-exon junctions. Asterix denote ambiguous amplicons. Representative blots and densitometry of LXR α (E) and LXR β (F) protein variants observed in cell lines. Pie charts show contribution of each transcript variant and protein variant to the total amount of LXR α (G) and LXR β (H). All data shown are mean of three independent replicates with SEM. Statistical significance was measured by two-tailed one-way ANOVA and significant differences are denoted with different letters if $p \leq 0.05$.

Figure 4. LXR α 1 and LXR α 4 are highly expressed in TNBC patients with poor survival. Representative blots showing the LXR splicing events in tumours derived from 38 TNBC patients, who were either free from disease (no event; $n=23$) or who had relapsed or died (event; $n=15$) after follow-up (A). Differential expression of LXR α and LXR β protein variants in No Event and Event groups were plotted in box and whiskers charts (B). The line shows the median value, the box shows 10th to 90th percentile and whiskers show minimum to maximum values. Statistical analysis was established using multiple *t*-tests with Holm-Sidak for multiple correction, $p \leq 0.05$ was considered significant. Pie charts show contribution of each transcript variant and protein variant to the total amount of LXR α in No Event and Event patients (C). Kaplan-Meier survival curves plotting disease free survival of TNBC patients dichotomised based on protein

expression of full length LXR α 1 and LXR α 4 transcripts (D), or truncated LXR α 5 and LXR β 1 and LXR β 4 (E). All protein measured relative to HPRT. Data derived from the mean of two different slices of tumours. Significance determined by the Log-rank (Mantel-cox) test where $p \leq 0.05$ was considered significant.

Figure 5. Protein expression of LXR variants is differentially associated with LXR target gene expression in TNBC tumours. TNBC patients (n=38) were divided into two groups, no event (n=23) and event (n=15) based on disease-free survival status. Linear regression was used to determine if the slope of the lines were significantly different from 0 ($p \leq 0.05$). Line of best fit are shown if the relationship was significant. Circles represent mean protein expression of two tumour slices for individual TNBC patients.

Acknowledgments

The mass spectrometry was performed and analysis supported by Dr James Ault and Dr Rachel George at the Biomolecular Mass Spectrometry Facility, Faculty of Biological Sciences, University of Leeds.

Funding Statement

Ms Priscilia Lianto was jointly funded by a Leeds International Doctoral Scholarship and the University of Leeds School of Food Science and Nutrition.

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: a cancer journal for clinicians* 68(6) (2018) 394-424.
- [2] C.A. Hudis, L. Gianni, Triple-negative breast cancer: an unmet medical need, *Oncologist* 16 (2011).
- [3] WCRF/AICR, Continuous Update Project Report: Diet, Nutrition, Physical Activity and Cancer, World Cancer Research Fund International/American Institute for Cancer Research 2017.
- [4] B. Liu, Z. Yi, X. Guan, Y.X. Zeng, F. Ma, The relationship between statins and breast cancer prognosis varies by statin type and exposure time: a meta-analysis, *Breast Cancer Res Treat* 164(1) (2017) 1-11.
- [5] L. Jiang, X. Zhao, J. Xu, C. Li, Y. Yu, W. Wang, L. Zhu, The Protective Effect of Dietary Phytosterols on Cancer Risk: A Systematic Meta-Analysis, *Journal of Oncology* 2019 (2019) 11.
- [6] G. Cioccoloni, C. Soteriou, A. Websdale, L. Wallis, M.A. Zulyniak, J.L. Thorne, Phytosterols and phytostanols and the hallmarks of cancer in model organisms: A systematic review and meta-analysis, *Crit Rev Food Sci Nutr* (2020) 1-21.
- [7] S.A. Hutchinson, A. Websdale, G. Cioccoloni, H. Roberg-Larsen, P. Lianto, B. Kim, A. Rose, C. Soteriou, A. Pramanik, L.M. Wastall, B.J. Williams, M.A. Henn, J.J. Chen, L. Ma, J.B. Moore, E. Nelson, T.A. Hughes, J.L. Thorne, Liver x receptor alpha drives chemoresistance in response to side-chain hydroxycholesterols in triple negative breast cancer, *Oncogene* (2021).
- [8] E.R. Nelson, S.E. Wardell, J.S. Jasper, S. Park, S. Suchindran, M.K. Howe, N.J. Carver, R.V. Pillai, P.M. Sullivan, V. Sondhi, 27-Hydroxycholesterol links hypercholesterolemia and breast cancer pathophysiology, *Science* 342(6162) (2013) 1094-1098.
- [9] A.E. Baek, A.Y. Yen-Rei, S. He, S.E. Wardell, C.-Y. Chang, S. Kwon, R.V. Pillai, H.B. McDowell, J.W. Thompson, L.G. Dubois, The cholesterol metabolite 27 hydroxycholesterol facilitates breast cancer metastasis through its actions on immune cells, *Nature communications* 8(1) (2017) 1-11.
- [10] M. Poirot, S. Silvente-Poirot, The tumor-suppressor cholesterol metabolite, dendrogin A, is a new class of LXR modulator activating lethal autophagy in cancers, *Biochem Pharmacol* 153 (2018) 75-81.
- [11] S.A. Hutchinson, P. Lianto, H. Roberg-Larsen, S. Battaglia, T.A. Hughes, J.L. Thorne, ER-Negative Breast Cancer Is Highly Responsive to Cholesterol Metabolite Signalling, *Nutrients* 11(11) (2019) 2618.
- [12] S. Solheim, S.A. Hutchinson, E. Lundanes, S.R. Wilson, J.L. Thorne, H. Roberg-Larsen, Fast liquid chromatography-mass spectrometry reveals side chain oxysterol heterogeneity in breast cancer tumour samples, *The Journal of steroid biochemistry and molecular biology* 192 (2019) 105309.
- [13] P.J. Willy, K. Umesono, E.S. Ong, R.M. Evans, R.A. Heyman, D.J. Mangelsdorf, LXR, a nuclear receptor that defines a distinct retinoid response pathway, *Genes & development* 9(9) (1995) 1033-1045.
- [14] D.M. Shinar, N. Endo, S.J. Rutledge, R. Vogel, G.A. Rodan, A. Schmidt, NER, a new member of the gene family encoding the human steroid hormone nuclear receptor, *Gene* 147(2) (1994) 273-276.

- [15] M. Chen, S. Beaven, P. Tontonoz, Identification and characterization of two alternatively spliced transcript variants of human liver X receptor alpha, *J Lipid Res* 46(12) (2005) 2570-9.
- [16] K. Endo-Umeda, S. Uno, K. Fujimori, Y. Naito, K. Saito, K. Yamagishi, Y. Jeong, H. Miyachi, H. Tokiwa, S. Yamada, M. Makishima, Differential expression and function of alternative splicing variants of human liver X receptor alpha, *Mol Pharmacol* 81(6) (2012) 800-10.
- [17] K. Hashimoto, E. Ishida, S. Matsumoto, N. Shibusawa, S. Okada, T. Monden, T. Satoh, M. Yamada, M. Mori, A liver X receptor (LXR)- β alternative splicing variant (LXRBSV) acts as an RNA co-activator of LXR- β , *Biochemical and biophysical research communications* 390(4) (2009) 1260-1265.
- [18] A.J. Annalora, C.B. Marcus, P.L. Iversen, Alternative Splicing in the Nuclear Receptor Superfamily Expands Gene Function to Refine Endo-Xenobiotic Metabolism, *Drug Metab Dispos* 48(4) (2020) 272-287.
- [19] W. Sun, T. Duan, P. Ye, K. Chen, G. Zhang, M. Lai, H. Zhang, TSVdb: a web-tool for TCGA splicing variants analysis, *BMC genomics* 19(1) (2018) 405.
- [20] E. Cerami, J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C.J. Byrne, M.L. Heuer, E. Larsson, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data, *AACR*, 2012.
- [21] R.V. Broad, S.J. Jones, M.C. Teske, L.M. Wastall, A.M. Hanby, J.L. Thorne, T.A. Hughes, Inhibition of interferon-signalling halts cancer-associated fibroblast-dependent protection of breast cancer cells from chemotherapy, *British Journal of Cancer* 124(6) (2021) 1110-1120.
- [22] J.L. Thorne, S. Battaglia, D.E. Baxter, J.L. Hayes, S.A. Hutchinson, S. Jana, R.A. Millican-Slater, L. Smith, M.C. Teske, L.M. Wastall, T.A. Hughes, MiR-19b non-canonical binding is directed by HuR and confers chemosensitivity through regulation of P-glycoprotein in breast cancer, *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* (2018).
- [23] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, S.G. Rozen, Primer3—new capabilities and interfaces, *Nucleic acids research* 40(15) (2012) e115-e115.
- [24] K. Suzuki, P. Bose, R.Y. Leong-Quong, D.J. Fujita, K. Riabowol, REAP: A two minute cell fractionation method, *BMC research notes* 3(1) (2010) 294.
- [25] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. Wilkins, R. Appel, A. Bairoch, *The Proteomics Protocols Handbook: Protein Identification and Analysis Tools on the ExPASy Server*, Humana Press, 2003.
- [26] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC bioinformatics* 12(1) (2011) 1-16.
- [27] B.A. Laffitte, S.B. Joseph, R. Walczak, L. Pei, D.C. Wilpitz, J.L. Collins, P. Tontonoz, Autoregulation of the human liver X receptor α promoter, *Molecular and cellular biology* 21(22) (2001) 7558-7568.
- [28] K. Prüfer, J. Boudreaux, Nuclear localization of liver X receptor α and β is differentially regulated, *Journal of cellular biochemistry* 100(1) (2007) 69-85.
- [29] K. Hayoun, D. Gouveia, L. Grenga, O. Pible, J. Armengaud, Evaluation of sample preparation methods for fast proteotyping of microorganisms by tandem mass spectrometry, *Frontiers in microbiology* 10 (2019) 1985.

- [30] A.Y. Belorusova, E. Evertsson, D. Hovdal, J. Sandmark, E. Bratt, I. Maxvall, I.G. Schulman, P. Åkerblad, E.-L. Lindstedt, Structural analysis identifies an escape route from the adverse lipogenic effects of liver X receptor ligands, *Communications biology* 2(1) (2019) 1-13.
- [31] G. Segala, M. David, P. de Medina, M.C. Poirot, N. Serhan, F. Vergez, A. Mougél, E. Saland, K. Carayon, J. Leignadier, Dendrogenin A drives LXR to trigger lethal autophagy in cancers, *Nature communications* 8(1) (2017) 1-17.
- [32] M.F. Tavazoie, I. Pollack, R. Tanqueco, B.N. Ostendorf, B.S. Reis, F.C. Gonsalves, I. Kurth, C. Andreu-Agullo, M.L. Derbyshire, J. Posada, S. Takeda, K.N. Tafreshian, E. Rowinsky, M. Szarek, R.J. Waltzman, E.A. McMillan, C. Zhao, M. Mita, A. Mita, B. Chmielowski, M.A. Postow, A. Ribas, D. Mucida, S.F. Tavazoie, LXR/ApoE Activation Restricts Innate Immune Suppression in Cancer, *Cell* 172(4) (2018) 825-840.e18.
- [33] J. Bunay, A. Fouache, A. Trousson, C. de Joussineau, E. Bouchareb, Z. Zhekun, A. Kocer, L. Morel, S. Baron, J.-M.A. Lobaccaro, Screening for Liver X Receptor modulators: where are we and for what use?, *British Journal of Pharmacology* (2020).
- [34] A. Zaghlool, A. Ameer, L. Nyberg, J. Halvardson, M. Grabherr, L. Cavelier, L. Feuk, Efficient cellular fractionation improves RNA sequencing analysis of mature and nascent transcripts from human tissues, *BMC biotechnology* 13(1) (2013) 99.
- [35] C. Rondanino, L. Ouchchane, C. Chauffour, G. Marceau, P. Dechelotte, B. Sion, H. Pons-Rejraji, L. Janny, D.H. Volle, J.M. Lobaccaro, F. Brugnon, Levels of liver X receptors in testicular biopsies of patients with azoospermia, *Fertil Steril* 102(2) (2014) 361-371 e5.

Table 1. LXR α peptides detected by S-trap column coupled with Mass Spectrometry (MS). Amino acids position numbers based on LXR α 1 structure. The grey highlight indicated unique peptides of LXR variant detected by MS.

Sample	Total identified peptides	-10lgP [LXR α]	Coverage LXR α peptides	Supporting peptides [LXR α]	Unique	Amino acids position	
						start	end
Tumour 2 rep 1	3500	52.32	13%	Q.AQ(+.98)GGSSC(+57.02)ILR.E	No	32	41
				A.QGGSSC(+57.02)ILR.E	No	33	41
				R.MPHSAGGTAGV.G	No	46	56
				R.ASSPPQ(+.98)ILPQ.L	No	196	205
				R.ASSPPQILPQLSPEQ(+.98)LGMIEK.L	No	196	216
				R.AMN(+.98)ELQLN(+.98)DAEFALLI.A	No	345	360
Tumour 2 Rep 2	3971	58.66	22%	K.C(+57.02)RQ(+.98)AGM(+15.99)REEC(+57.02)VLSEEQ(+.98)IRLK.K	No	158	177
				R.QEEEEQAHATSLPPR.A	No	182	195
				R.ASSPPQILPQLSPEQ(+.98)LGMIEK.L	No	196	216
				L.SPEQ(+.98)LGMIEK.L	No	207	216
				G.M(+15.99)IEKLVAQQ(+.98)Q(+.98).C	No	213	223
				K.QLPGFLQLSR.E	No	274	283
				Q.VEFINPIFEFSR.A	No	333	344
				R.AMNELQLNDAEFALLI.A	No	345	360
Tumour 5 Rep 1	3713	61.52 [α 4 +LXR α homolog]	19%	MQ(+.98)QTSWN(+.98)PLGGTC(+57.02)K.Q	Yes [α 4]	-14	-1
		59.30 [LXR α homolog]	16%	R.AEPPSEPTEIRPQ(+.98)K.R	No	70	83
				KC (+57.02)RQAGMR.E	No	196	208
				R.ASSPPQILPQLSP.E	No	196	208
				R.ASSPPQILPQLSPEQ(+.98)LGMIEK.L	No	196	216
				G.MIEKLVAQQQC(+57.02)NRR.S	No	213	227
				K.TSAIEVMLLETSTRRYN(+.98)PG.S	No	292	309

Tumour 5 Rep 2	3412	38.73	9%	C.VLSEEQIR.L	No	168	175
				R.ASSPPQILPQLSPEQ(+.98)LGMIEK.L	No	196	216
				R.ASSPPQ(+.98)ILPQ.L	No	196	205
				R.EDQIALLK.T	No	284	291

Table 2. LXR β peptides detected by S-trap column coupled with Mass Spectrometry (MS). Amino acids position numbers based on LXR β 1 structure. The grey highlight indicated unique peptides of LXR variant detected by MS.

Sample	Total identified peptides	-10lgP [LXR β]	Coverage LXR β peptides	Supporting peptides [LXR β]	Unique	Amino acids position	
						start	end
Tumour 2 Rep 1	3500	46.60	7%	L.MIQQLVAAQ.L	No	226	234
				Q.Q(+.98)LVAAQLQC(+57.02)NK.R	No	229	239
				Q.VEFINPIFEFSR.A	No	346	357
				K.RPQDQ(+.98)LR.F	No	410	416
Tumour 2 rep 2	3971	39.62	10%	R.RSVVRGGAR.R	No	113	121
				K.EAGM(+15.99)REQ(+.98)C(+57.02)VLSEEQIRKK.K	No	151	168
				K.VTPWPLGADPQSR.D	Yes [β 1]	248	260
				K.RPQDQ(+.98)LR.F	No	410	416
Tumour 5 Rep 1	3713	56.33	14%	R.SVVRGGAR.R	No	113	121
				E.LM(+15.99)IQQ(+.98)LVAAQ.L	No	225	234
				V.QLTAAQ(+.98)ELMIQ(+.98).Q	No	218	228
				Q.Q(+.98)LVAAQLQC(+57.02)NK.R	No	229	239
				K.RSFSDQ(+.98)PK.V	No	240	247
				K.VTPWPLGADPQ(+.98)SR.D	Yes [β 1]	248	260
				A.LQQ(+.98)PYVEALLS.Y	No	394	404
				E.ALLSYTR.I	No	401	407
Tumour 5 Rep 2	3412	59.82	20%	R.RSVVRGGAR.R	No	113	121
				C.VLSEEQIR.K	No	159	166
				R.KQ(+.98)QQ(+.98)ESQSQSQSPVGPQG.S	No	172	189

E.GVQLTAAQ(+.98)ELMIQ.Q	No	216	228
E.LM(+15.99)IQQ(+.98)LVAAQ.L	No	225	234
Q.QLVAAQ(+.98)LQ(+.98)C(+57.02)N(+.98)K.R	No	229	239
K.RSFSDQ(+.98)PK.V	No	240	247
R.EDQIALLK.A	No	297	304
A.LQQ(+.98)PYVEALLS.Y	No	394	404
K.RPQDQ(+.98)LR.F	No	410	416

Table 3. Unique peptides of LXR variants. Amino acids position number based on $\alpha 1$ and/or $\beta 1$ amino acid numbering position. Amino acid numbering position with “-“ indicates the additional amino acid(s) coming before $\alpha 1$ and/or $\beta 1$ ’s amino acid position number 1. The numerical exons represent the first-discovered exons. The numerical exons followed with alphabets represent the later-discovered exons.

LXR	Also, recognise	Peptide mass	Amino acid position (based on LXRa1 structure)		Exon	Unique peptides (after trypsin digestion)
			start	end		
$\alpha 1$	$\alpha 2$	1607.8137	1	15	2	MSLWLGAPVPDIPPD
		1348.7140	293	300	6	SAIEVMLLETSR
$\alpha 2$	$\alpha 5$	1148.6343	235,236-297	300	5/7	VT-VMLLETSR
$\alpha 3$	$\alpha 5$	2307.1761	46	69	3	MPHSAGGTAGVGLEAAEPTALLTR
$\alpha 4$	-	1550.7090	-14	-1	-1d	MQQTSWNPLGGTCK
		554.3045	-19	-15	-1d	QPPGR

XP_0243040XXX -	2942.3842 1170.5538 1005.4669 741.3890 590.3079 510.2381	Additional 64 amino acids due to the presence of a retained intron in between exon 6 and 7		Exon -7a	GAEWDYLVWEGPPDIELGEP NLLGSR DEENRPPWK EVAGEGQGMK TSPPSPR RPCSK FAACV
β1	1423.7328 1475.7892	248 268	260 279	Exon6/ exon7abc	<u>VTPWPLGADPQSR</u> <u>FAHFTELAISVQ</u>
β3	1021.5564	249-280	286	Exon6/ exon7b	<u>V-TEIVDFAK</u>
β4	545.3405	249-293	299	Exon6/ Exon 7c	<u>V-TLGREDQ</u>

Figure 1

Figure 2

Figure 3

Figure 4

Target gene mRNA expression in TNBC tumours

Figure 5