This work was supported by funds from Foundation for Science and Technology (FCT), European Regional Development Fund (FEDER) and Compete under project POCI-01-0145-FEDER-030562

“If I have seen further, it is by standing on the shoulders of Giants”

- Isaac Newton

AKNOWLEDGEMENTS

One of my favourite quotes states “If I have seen further, it is by standing on the shoulders of Giants”. This is a phrase often used in reference of scientific progress, a context that I believe couldn’t be more adequate, but I also relate to it in the context of the people that surround us. “No man is an island”, and I am no exception. This past year was particularly defying, on all levels, and has shown me, more than in any other period in my life, how lucky and thankful I am of the people I have by my side. This dissertation wouldn’t have been possible without then, as they have made me learn and grow every single day, and for that I am truly thankful.

This year has showed me how important a good work environment can be, because you will always be able to get out of bed with the motivation to lead another day of work. I was extremely lucky to work alongside some truly amazing people, who have inspired and thought me so much.

My biggest thanks go to my supervisor, Doctor Renata Freitas. I don’t think I have enough words to describe how thankful I am to have found such an amazing supervisor. Renata, you have guided me through the last three years, in a way that I wish everyone was as lucky to have. You always had a kind word to stay, a positive point of view, and the best motivational speeches. You have always pushed me to keep going forward, to give it my best, to be a better scientist, colleague and person. You always offered to help out, even if that meant marching on a protest with me. I will never forget all the opportunities, advice and help that you have given me throughout these years. It has been a true honour to be able to learn from such an inspiring scientist and woman, with such intelligence, and strong values. You have taught me so much, and have made a huge impact on the person and scientist I am today.

To Simone and Mafalda, I would like to thank for all the teachings and huge patience you had with me. I entered this lab, three years ago, a sloppy and uncoordinated person, and you have completely turned me around. Your method, your organization, your professionalism, and your advice really inspired me to follow the same steps, and if I was able to finish this dissertation, was due to all of the things that you have passed on to me (and your amazing lab books). I am so grateful to be surrounded by such inspiring women.

Still on the topic of strong women, I would like to thank Raquel. I can already picture her rolling her eyes, and now laughing because I have totally nailed her reaction, but something cheesy is coming. Throughout my time at the lab, you went from colleague to friend, to confidant, to the person I always shared laugh and despair with. You faced every situation with a joke and a huge smile on your face. No matter the struggle, you would always overcome it. You are one of the

bravest people I know, and I really look forward to the time when we are sitting in our rocking chairs, very very old, complaining about our husbands, and how much our hips hurt. Until then, just know that I will be here, to listen to you telling me “I told you so” numerous times, but also for everything you might need, whether that’s a really good laugh, or to remind you of your worth. I know great things are waiting for you, and you deserve every single one of them.

Another huge thanks to Francisco. When I say this could not have been done without you, I mean it. You have been my accountability partner, making sure that I was on track with my writing, but also a great friend. I know you are a person of few words, but you have always managed to say what I needed to hear, when I needed to hear it. For all the late nights working, memes, shared food, coffee breaks, deep and random conversations, gym sessions, rock climbing classes, blood drives and so much more, I owe you a big thank you. You have progressed so much over this year, and I hope you always know that you are capable of all of that, and so much more. If I haven’t said it already, I am really proud of you, and I am glad that we are ending this journey together.

To all of my other colleagues at the lab, I want to thank Miguel for all the funny stories over that chocolate “truffle square”, to Rita for all the shared moments at the lab and at the gym, and to the best and most patient statistics teacher I could have, João Mouta. Thank you so much for everything. To Ana, for being the kind, funny and sweet person she is. I know you are capable of so much. I will be here for anything that you need.

Regarding my family, the biggest of thanks to my grandma. You’ve been like a mother to me, but also a friend. I never thought I would have so much complicity with someone whose age is so far apart from my own, but you have always been ahead of your time, and have made me feel like I could always talk to you about anything. Thank you for all that you have done for me. I love you so much. To my sister, “Tata”, I have to thank you for being the most amazing sister a person could ask for. You make me proud, every single day. I love you. To the Matos family, who have taken me in as their own, who have made me feel welcomed, and celebrated every “victory” with me, no matter how small. Thank you for everything.

To all my amazing friends, who have always made me feel so loved and valued and who have always pushed me forward. A special thanks to Carrinho and Coutinho, for all of their plenitude, but above everything else, for being the best friends I could ask for. You are some of the most inspiring women I know, and make me feel proud every day to be able to share this journey with you. I am so “proud you were created” and of the women I saw you grow into. To Varela, that was always present, even on the other side of Europe. You always manage to make the hardest of

days better. To Inês for all the late nights writing, but most importantly for showing me that we have the power to break any curse and live the life we deserve. To Gabrielle, who has been a huge support, and the biggest example of strength I could imagine. You are capable of overcoming any situation, while still being kind to everyone, and that is one of the many many things I admire about you. I am so proud of you two, and I know that you will conquer every single one of your dreams. To Rute C., for being the only person I will ever be able to blindly trust with any work. To Ricardo, for always being able to top me on the worst jokes. To Jé, for ensuring that no one takes advantage of us. To Rute G., for every deep conversation and for noticing all of the things that I always seem to miss (and not only the colour of Ricardo's glasses), and to Rita, for all of her funny snarky comments at the right time. You have made the programme much more bearable.

Last but not least, to João. Words are never enough to thank you for everything you have done for me in the past five years. You have been my best friend, you have been my boyfriend and you have been family. I am so immensely proud of you, of the person you have grown into, and of everything that I know you will accomplish. Thank you for always being there for me, and for always reminding me that I am capable of everything I set my mind into. I can never thank you enough for giving me these years. I love you.

I have learned so much from all of you, but I have also learned that no mountain is too high for you to climb. So I will leave this small note here, for future Joana. If you ever feel alone, or that you can't do something, read this, and remember how lucky you are to be surrounded by these people, and that, if you have made it this far, you are capable of anything.

Thank you all.

ABSTRACT

Breast Cancer (BrCa) is the most incident cancer in women and the leading cause of cancer-related death in this group. This type of cancer is highly heterogeneous, which emphasises the need to find biomarkers to improve prognosis and therapy and ensure personalized treatment. Several *HOX* genes were found dysregulated in BrCa and related to poor prognosis. However, the causes of this dysregulation and the mechanisms affected remain largely uncharacterized.

In this work, we studied the causes of *HOXB7* dysregulation in BrCa by illustrating three epigenetic events involved in its aberrant expression in distinct BrCa cellular models: promoter methylation, histone deacetylation, and histone acetylation. For this purpose, we have subjected five breast cancer cell lines, representing distinct molecular subtypes, and a cell line representative of breast epithelium, to a hypomethylating compound (5-aza-2-deoxycytidine). We observed that, upon treatment, *HOXB7* expression increased in all the analysed lines (MCF-10A, MDA-MB-231, MDA-MB-468, MCF7, SKBR3, and BT-474), strongly suggesting that methylation plays a crucial role in *HOXB7* regulation in BrCa. Upon these results, we characterized the methylation profile of the CpG island in the *HOXB7* promoter in MCF-10A, MDA-MB-231 and SKBR3 to determine the role that this epigenetic event plays in *HOXB7* regulation. We were able to identify that the differences in *HOXB7* expression between these cell lines were not due to differences in promoter methylation, but might be due to other methylation events regarding other locations, in or outside the *HOXB7* gene itself. Additionally, we analysed the role of methylation on the expression regulation of *HOXB8* and *HOXB9*, as well as in *HOXB7* putative targets *CDH1* and *CTNNB1*, and found that methylation regulated these genes, whether by direct effect or indirectly, possibly through the regulation of upstream modulators. Furthermore, we assessed the role of histone deacetylation through TSA treatment in MCF-10A, MDA-MB-231, MCF-7, and MDA-MB-468, and concluded that this process is not responsible for *HOXB7* repression in normal cells, but may interfere with the regulation of this gene in breast cancer cells.

In this dissertation, we also evaluated the therapeutic potential of novel N-1,2,3-triazole-isatin hybrid drugs, performing also gene expression assays regarding their target proteins (acetyltransferase p-300 and acyl-protein thioesterase), and the effect of their inhibition on *HOXB7* expression. We found that these compounds were highly effective on all breast cells analysed, but needed a higher dosage to induce cytotoxicity on other cancer cell lines and in normal kidney cells (Hek293). Target inhibition was able to modulate *HOXB7* expression, particularly on MCF-7, in which p-300 inhibition strongly decreased *HOXB7* expression,

strappingly suggesting that histone acetylation is one of the reasons why this gene is overexpressed in this cell line.

Overall, we aimed to explore the mechanisms associated with *HOXB7* dysregulation in BrCa and contribute to uncover novel therapeutic agents and targets to improve prognosis, diagnosis, and therapy in BrCa patients.

Keywords: breast cancer, *HOXB7*, epigenetics, methylation, acetylation, *HOXB8*, *HOXB9*, *CDH1*, *CTNNB1*, triazol-isatins, p-300, APT-2, novel therapies

RESUMO

O cancro da mama (BrCa) é o mais incidente na população feminina e a principal causa de morte por doença oncológica neste grupo. Esse tipo de cancro é altamente heterogéneo, o que enfatiza a necessidade de encontrar biomarcadores para melhorar o prognóstico e a terapia de pacientes com esta patologia, e garantir um tratamento personalizado. Vários genes *HOX* foram encontrados desregulados no BrCa e associados a um mau prognóstico. No entanto, as causas dessa desregulação e os mecanismos afetados permanecem, na sua maioria, pouco caracterizados.

Neste trabalho, estudamos as causas da desregulação do *HOXB7* em BrCa, através da ilustração de três eventos epigenéticos envolvidos em sua expressão aberrante, em modelos celulares distintos de BrCa: metilação do promotor, desacetilação de histonas e acetilação de histonas. Para esse propósito, submetemos cinco linhas celulares de cancro de mama, representantes de diferentes subtipos moleculares, e uma linha representativa de tecido epiteliar mamário normal, a um composto hipometilante (5-aza-2-deoxicitidina). Observamos que, após o tratamento, a expressão de *HOXB7* aumentou em todas as linhas analisadas (MCF-10A, MDA-MB-231, MDA-MB-468, MCF7, SKBR3 e BT-474), o que sugere que a metilação desempenha um papel crucial na regulação do *HOXB7* em BrCa. Perante esses resultados, caracterizamos o perfil de metilação das ilhas CpG localizadas no promotor do *HOXB7* em MCF-10A, MDA-MB-231 e SKBR3, para determinar se papel que este evento epigenético desempenha na regulação de *HOXB7* é direto ou indireto. Isto permitiu concluir que as diferenças na expressão de *HOXB7* entre essas linhas celular não foram devido a diferenças na metilação do promotor, mas podem ser devido a outros eventos de metilação noutras localizações intregénicas, na região circundante ao gene do *HOXB7* ou em genes implicados na sua regulação. Além disso, analisamos o papel da metilação na expressão dos genes *HOXB8* e *HOXB9*, bem como nos alvos putativos de *HOXB7*, *CDH1* e *CTNNB1*, e descobrimos que a metilação tinha um papel na regulação destes genes, efeito direto ou indireto, possivelmente via regulação de moduladores destes genes. além disso Estas analises indicaram que....Complementarmente, avaliamos o papel da desacetilação de histonas através do tratamento com TSA em MCF-10A, MDA-MB-231, MCF-7 e MDA-MB-468, e concluímos que este processo não é responsável pela repressão do *HOXB7* em células normais, mas pode interferir na regulação deste gene em células cancerígenas.

Nesta dissertação, também foi realizado um estudo do potencial terapêutico de novos fármacos híbridos de N-1,2,3-triazol-isatina, juntamente com ensaios de expressão génica das suas proteínas-alvo (acetiltransferase p-300 e acil-proteína tioesterase), e o efeito de sua inibição na expressão de *HOXB7*. Descobrimos que esses compostos eram altamente eficazes em células da mama, mas que era necessária um dosagem mais elevada para induzir a citotoxicidade em linhas celulares representativas de outros tipos de cancro e em células renais normais (Hek293). A inibição das proteínas-alvo foi capaz de modular a expressão de *HOXB7*, particularmente em MCF-7, em que a inibição de p-300 diminuiu fortemente a expressão deste gene, sugerindo que a acetilação de histonas é uma das razões pelas quais este gene é superexpresso nesta linha celular.

Em síntese, o nosso objetivo foi explorar os mecanismos associados à desregulação do *HOXB7* em BrCa e contribuir para a descoberta de novos agentes e alvos terapêuticos com o objetivo de melhorar o prognóstico, diagnóstico e terapia em pacientes com esta doença oncológica.

Palavras-chave: cancro da mama, *HOXB7*, epigenética, metilação, acetilação, *HOXB8*, *HOXB9*, *CDH1*, *CTNNB1*, triazol-isatinas, p-300, APT-2, novas terapeuticas

TABLE OF CONTENTS

AKNOWLEDGEMENTS III

ABSTRACT VI

RESUMO VIII

TABLE OF CONTENTS X

LIST OF TABLES AND FIGURES XIII

LIST OF ABBREVIATIONS XX

1. INTRODUCTION 1

1.1. Breast Cancer 1

1.1.1. Epidemiology 1

1.1.2.Risk Factors 3

1.1.3.Diagnostic and Screening 5

1.1.4.Histopathology 6

1.1.5.Molecular subtypes and predictive factors 7

1.2.HOXOME 9

1.2.1.The *HOX* family 9

1.2.2.*HOX* genes in cancer 12

1.2.3.*HOX* genes in breast cancer 14

1.2.4.*HOX* gene epigenetic regulation in cancer 18

2. AIMS 23

3. MATERIALS AND METHODS 24

3.1.Cell lines and culture 24

3.2.Treatment with DNA methylation inhibitor 5-aza-2′-deoxycytidine (ZdCyd) 24

3.3.Total RNA extraction from cells 25

3.4.cDNA synthesis 26

3.5.Reverse Transcription Quantitative PCR 27

3.6.Bisulphite conversion of DNA 29
3.7.Gradient PCR 29
3.8.Cloning into pGEM®-T Easy Vector 31
3.9.Competent Bacteria DH5α Transformation 32
3.10.Blue-white screening and bacterial selection 32
3.11.Miniprep and colony PCR 33
3.12.DNA sequencing and Methylation Status Analysis 33
3.13.Treatment with inhibitor of histone deacetylase Trichostatin A (TSA) 34
3.14.MTT Anti-proliferative Assay upon treatment with N-1,2,3-Triazole-Isatin Derivatives 34
3.15.Treatment for gene expression analysis on MCF-7, MDA-MB-231 and MCF-10A 35
3.16.Statistical Analysis 35
4. RESULTS 36
4.1.Regulation of *HOXB7* expression through methylation 36
4.1.1.Variations on *HOXB7* expression upon 5-aza-2′-deoxycytidine treatment 36
4.1.2.Methylation profile of the *HOXB7* promoter region 38
4.2.*HOXB8* and *HOXB9* expression upon ZdCyd treatment 42
4.3.Expression of *HOXB7* putative downstream ZdCyd treatment: *CDH1* and *CTNNB1* 44
4.4.*HOXB7* expression regulation through histone deacetylation 46
4.5.Exploring targeted therapies to overcome *HOXB* dysregulation in BrCa 47
4.5.1.Anti-Proliferative effect of N-1,2,3-Triazole-Isatin Derivatives 49
4.5.2.Gene expression studies of the predicted biological targets 50
4.5.3.N-1,2,3-Triazole-Isatin Derivatives as modulators of *HOXB7* expression 52
5. DISCUSSION 55
5.1.Regulation of *HOXB7* through DNA methylation 55
5.2.Regulation of *HOXB8* and *HOXB9* through DNA methylation 58
5.3.Regulation of *CDH1* and *CTNNB1* through DNA methylation 60
5.4.Regulation of *HOXB7* through histone deacetylation (HDAC) 62

5.5.Therapeutic potential of N-1,2,3-Triazole-Isatin derivatives and *HOXB7* modulation 63

6. CONCLUSION 68

7. REFERENCES 70

8. SUPPLEMENTARY TABLES 87

LIST OF TABLES AND FIGURES

Table 1- Molecular Classification and Immunoprofile of the cell lines used in the course of this dissertation. Positive (+) and Negative (-) symbols stand for amplified and non-amplified receptors, respectively. Based on [149-152]. 24

Table 2-Reagents used for cDNA synthesis using the High-Capacity cDNA Reverse Transcription Kit. 26

Table 3- Reagents used in validatory PCR to validate cDNA synthesis. FW- Forward. RV-Reverse (*NZYTaq II 2x Green Master Mix, NZYtech) 27

Table 4- Sequence and product length of the primers used in qPCR analyses. [S]- self-made using Primer3 Online Software. Bp- base pair. FW - forward primer. RV- reverse primer.....28

Table 5- Components used in gradient PCR reaction. FW - forward primer. RV- reverse primer. (*QIAGEN Multiplex PCR Master Mix) 30

Table 6- Thermocycler Protocol for Gradient PCR. 30

Table 7- Components used in the ligation reaction for vector cloning. 31

Table 8- Calculated IC_{50} of the target compounds against MCF-7, MDA-MB-231, MCF-10A and Hek293 cell lines using the MTT assay. Enantiomeric purity was Determined using HPLC with a chiral column, and is presented in percentage of enantiomeric excess (%ee). IC_{50} values calculated from MTT data using the GraphPadPrism Software Inc. (version 8.0.1), and are presented in µM ± 95% Confidence Level Intervals (CI).....49

Supplementary Table 1- Calculated IC50 of the target compounds against DOHH2, VAL, OCI-LY-10, SU-DHL-2, SW480, A549 cell lines using the MTT assay. Enantiomeric purity was Determined using HPLC with a chiral column, and is presented in percentage of enantiomeric excess (%ee). IC50 values calculated from MTT data using the GraphPad Prism Software Inc. (version 8.0.1), and are presented in µM ± 95% Confidence Level Intervals (CI) 105

Figure 1- Worldwide cancer incidence (A) and mortality (B) rates among women. Source: GLOBOCAN 2020. Graph Production: Global Cancer Observatory – International Agency for

Research on Cancer (http://gco.iarc.fr). Extracted from the World Health Organization Website (accessed on 31/12/2020)..... 1

Figure 2- Breast cancer incidence (A) and survival (B) rates across EU-28 countries. A: Age-standardized incidence rates in European women, in 2018. B: Five-year survival rate in EU-28 citizens, in 2014. Adapted from [6]. 2

Figure 3- BC pathogenesis and histologic versus molecular subtypes, by Eric Wong and Joana Rebelo..... 8

Figure 4- Arrangement of the HOX clusters in the human genome and their connection to cancer. A. HOX gene structure highlighting the homeodomain. **B.** Thirty-nine genes dispersed by four clusters (HOXA, HOXB, HOXC, HOXD) located on four distinct chromosomes. In each cluster, genes are arranged in tandem sequence from 3' to 5', which is relevant for the temporal and spatial activation of each HOX gene during embryonic development. The white bar emphasizes that all HOX genes have been found related to cancer progression. HOX genes with the same number are referred to as 'paralogs', in [70]. 10

Figure 5 - Representation of the direct and indirect roles that HOX proteins potentially have in tumorigenesis. (i) After HOX gene transcription, mRNA is exported from the nucleus into the cytoplasm. (ii) HOX mRNA is then translated into HOX proteins. (iii) These proteins are translocated into the nucleus, affecting the expression of a set of genes that may include methyltransferases such as DNMT3B. (iv) mRNA from these genes are then exported to the cytoplasm and translated. (v) Methyltransferases are translocated into the nucleus, where they inhibit the expression of other genes by adding methyl group to their sequence. (vi) HOX genes may also regulate the expression of oncogenes or tumor-suppressor genes, in [70].....13

Figure 6 - Expression of HOX genes in the nine most common solid tumours, by Duan et al. 2015 [88]. 14

Figure 7- Hox gene expression and function in MG development. MG development throughout its six developmental stages: pre-puberty, puberty, pre-pregnancy, pregnancy, lactation, and involution. TEB: Terminal embryonic buds: AB: Alveolar buds; MA: mature milk alveoli. Adapted from [72]. 15

Figure 8- HOXB7 role in pathways critical to initiation and progression of solid cancers, namely, apoptosis, angiogenesis, proliferation, migration, drug resistance, invasiveness, DNA repair, and EMT. Adapted from [105]. 17

Figure 9- The influence of epigenetic modifications on gene expression and associated therapeutic agents. (A) DNA methylation is accompanied by transcriptional repression and is increased by DNA-methyltransferases (DNMT). (B) Histone methylation is facilitated by histone methyltransferases (HMT) and results in chromatin condensation, which may either inhibit or stimulate transcription**.** (C) Histone acetylation, which is promoted by histone acetyltransferases (HAT) enzymes, is linked with the opening of the chromatin mass and the initiation of transcription, whereas deacetylation, which is promoted by histone deacetylases (HDAC), is the inverse. Adapted from [122]. 22

Figure 10- HOXB7 relative expression upon ZdCyd treatment across different cell lines, analyzed by RT-qPCR. **A.** HOXB7 relative expression in MCF-10A upon ZdCyd treatment at 1µM ($p<0.0001$) and 5µM ($p=0.0030$). **B.** HOXB7 relative expression in SKBR3 upon ZdCyd treatment at 1µM ($p=0.0072$) and 5µM ($p=0.0615$). **C.** HOXB7 relative expression in MDA-MB-231 upon ZdCyd treatment at 1µM ($p=0.0325$) and 5µM ($p=0.0166$). **D.** HOXB7 relative expression in MDA-MB-468 upon ZdCyd treatment at 1µM ($p=0.0145$) and 5µM ($p=0.0210$). **E.** HOXB7 relative expression in MCF-7 upon ZdCyd treatment at 1µM ($p=0.6052$) and 5µM ($p=0.1067$). **F.** HOXB7 relative expression in BT-474 upon ZdCyd treatment at 1µM ($p=0.3439$) and 5µM ($p=0.0233$). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, $p \leq 0.05$, **, $p \leq 0.01$, ****, $p \leq 0.0001$). All experiments were performed in triplicate, using independent biological replicates.....37

Figure 11- Endogenous HOXB7 mRNA expression levels analyzed by qPCR in breast cancer cell lines SKBR3, MDA-MB-231, MDA-MB-468, MCF-7 and BT-474, compared with normal breast cells MCF10-A. MCF-10A represented as MCF10A, MDA-MB-231 represented as MDA231, MDA-MB-468 represented as MDA468, MCF-7 represented as MCF7, and BT-474 represented as BT474. Blank spaces in each column represent a break in the y scale, to facilitate reading. Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, $p \leq 0.05$, **, $p \leq 0.01$). All experiments were performed in triplicate, using independent biological replicates. Adapted from [81].....38

Figure 12- Predicted CpG islands within the promoter sequence of HOXB7. NCBI Reference Sequence AJ414528 for HOXB7 promoter was submitted fully to the EMBOSS CpGPlot website http://www.ebi.ac.uk/emboss/cpgplot website. The selected parameters to define a CpG island were set as follow: sequence of ≥ 200 bp with a percentage of C + G content ≥ 50% (**B**) and threshold of CpG observed/CpG expected ≥ 0.6 (**A**). Following these parameters, the program was able to identify one CpG island between nucleotides 1425 and 1810 (**C**), which are identified by the red arrows. **D.** Schematic representation of HOXB7 promoter sequence. Predicted CpG island in red. Selected region for sequence amplification between positions 1346 and 1796. ...39

Figure 13- Methylation profile of the selected HOXB7 promoter region in MCF-10A cell line. Bisulphite sequencing analysis revealed a methylation status of 50.9%, out of 289 CpG sites analysed from 10 independently sequenced clones. Each square represents one CpG site. Methylated CpG sites are represented by a black square, while unmethylated ones are represented by a grey square, and non-analysed CpG islands are identified by a white square. Each row represents a single clone.....40

Figure 14- Methylation profile of the selected HOXB7 promoter region in MDA-MB-231 cell line. Bisulphite sequencing analysis revealed a methylation status of 51.7%, out of 232 CpG sites analysed. Each square represents one CpG site. Methylated CpG sites are represented by a black square, while unmethylated ones are represented by a grey square, and non-analysed CpG islands are identified by a white square. Each row represents a single clone.....41

Figure 15- Methylation profile of the selected HOXB7 promoter region in SKBR3 cell line. Bisulphite sequencing analysis revealed a methylation status of 50.0%, out of 140 CpG site analysed. Each square represents one CpG site. Methylated CpG sites are represented by a black square, while unmethylated ones are represented by a grey square, and non-analysed CpG sites are identified by a white square. Each row represents a single clone.....41

Figure 16- Aligned sequences of the analyzed CpG island in the HOXB7 promoter region. The five analyzed clones were aligned and compared to a reference sequence of the promoter region, where all the cytosines had already been converted through bisulfite treatment. List of sequences are presented on the left, while nucleotides are presented on the right. To facilitate analysis, the represented sequence starts at the position corresponding to 172bp of the original sequence. Point mutation highlighted by red boundary. Adenine (A) is presented in green,

Thymine (T) by dark pink, Guanine (G) by purple and Cytosine (C) by blue. (*) correspond to conserved nucleotides. 42

Figure 17- HOXB8 relative expression upon ZdCyd treatment across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analyzed by RT-qPCR. **A-C.** HOXB8 relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon ZdCyd treatment at 1µM (p=0.0031 MCF-10A; p=0.0134 MDA-MB-231; p=0.0485 MCF-7) and 5µM (p=0.1987 MCF-10A; p=0.0158 MDA-MB-231; p=0.0747 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on analyses using Unpaired T-Test with Welch correction (*, p ≤ 0.05, **, p ≤ 0.01). All experiments were performed in triplicate, using independent biological replicates. .42

Figure 18- *HOXB9* relative expression upon ZdCyd treatment across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analyzed by RT-qPCR. **A-C.** *HOXB9* relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon ZdCyd treatment at 1µM (p=0.0006 MCF-10A; p=0.9685 MDA-MB-231; p=0.0006 MCF-7) and 5µM (p=0.0017 MCF-10A; p=0.0198 MDA-MB-231; p=0.0060 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, p ≤ 0.05, **, p ≤ 0.01, ***, p ≤ 0.001). All experiments were performed in triplicate, using independent biological replicates. 43

Figure 19- *CDH1* relative expression upon ZdCyd treatment across cell lines MCF-10A (A), MDA-MB-231 (B) and MDA-MB-468 (C), analyzed by RT-qPCR. A-C. *CDH1* relative expression in MCF-10A, MDA-MB-231 and MDA-MB-468 upon ZdCyd treatment at 1µM (p=0.0057 MCF-10A; p=0.0097 MDA-MB-231; p=0.0001 MDA-MB-468) and 5µM (p=0.0059 MCF-10A; p=0.0104 MDA-MB-231; p=0.0022 MDA-MB-468). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, p ≤ 0.05, **, p ≤ 0.01, ***, p ≤ 0.001). All experiments were performed in triplicate, using independent biological replicates. 44

Figure 20 - *CTNNB1* relative expression upon ZdCyd treatment across cell lines MCF-10A (A), MDA-MB-231 (B) and MDA-MB-468 (C), analyzed by RT-qPCR. A-C. *CTNNB1* relative expression in MCF-10A, MDA-MB-231 and MDA-MB-468 upon ZdCyd treatment at 1µM (p=0.0071 MCF-10A; p=0.2472 MDA-MB-231; p=0.0740 MDA-MB-468) and 5µM (p=0.0058 MCF-10A; p=0.0231 MDA-MB-231; p=0.5767 MDA-MB-468). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-

Test with Welch correction (*, p ≤ 0.05, **, p ≤ 0.01). All experiments were performed in triplicate, using independent biological replicates. 45

Figure 21- 2% Agarose gel representative of HOXB7 and GAPDH expression on MCF-10A cells upon treatment with 400nM TSA. (L) 100bp Ladder. **1,3,5,7,9,11:** Lanes corresponding to HOXB7 expression. **2,4,6,8,10,12:** Lanes corresponding to GAPDH amplification. **1,2:** Lanes corresponding to biological replicate 1 of DMSO control. **3,4:** Lanes corresponding to biological replicate 2 of DMSO control. **5,6:** Lanes corresponding to biological replicate 3 of DMSO control. **7,8:** Lanes corresponding to biological replicate 1 of 400nM TSA treatment. **9,10:** Lanes corresponding to biological replicate 2 of 400nM TSA treatment. **11,12:** Lanes corresponding to biological replicate 3 of 400nM TSA treatment. **13.** Lane corresponding to HOXB7 negative control. **14.** Lane corresponding to GAPDH negative control. GAPDH was used as gene expression control. Primers used in PCR amplification are represented on Table 4. 46

Figure 22- HOXB7 relative expression upon TSA treatment across cell lines MDA-MB-231 (A), MCF-7 (B), and MDA-MB-468 (C) analyzed by RT-qPCR. **A-C.** HOXB7 relative expression in MDA-MB-231, MCF-7 and MDA-MB-468 upon treatment with 400nM of TSA (p=0.0307 MDA-MB-231; p=0.0283 MCF-7; p=0.0155 MDA-MB-468). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, p ≤ 0.05). All experiments were performed in triplicate, using independent biological replicates. 47

Figure 23- Structure and molecular formula of the five N-1,2,3-Triazole-Isatin derivatives used in this assay: isatin-triazoles (1a), (1b) and (1c) and 3-aryl-2-hydroxyoxindole-/acetal-protected-oxindole triazoles (S)-(2b) and (R)-(2j). 48

Figure 24- Docked compounds (1a), (S)-(2b) and (R)-(2j). A. Docked compound (1a) (cyan) in the binding site of acetyltransferase p-300 (in gray). **B.** Docked compound (S)-(2b) (in salmon) in the binding site of acyl-protein thioesterase 2 (in gray). **C.** Docked compound (R)-(2j) (in pink) in the binding site of acyl-protein thioesterase 2 (in gray). Oxygen atoms in red, nitrogen atoms in blue, sulfur in yellow, bromine in carmine. 51

Figure 25- EP300 relative expression upon treatment with compounds (1a), (1b) and (1c), across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analyzed by RT-qPCR. A-C. EP300 relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon treatment with compounds (1a) (p=0.0802 MCF-10A; p=0.0113 MDA-MB-231; p=0.0048 MCF-7), compounds

(1b) (p=0.0955 MCF-10A; p=0.0086 MDA-MB-231; p=0.0020 MCF-7), and compound (1c) (p=0.0876 MCF-10A; p=0.0128 MDA-MB-231; p=0.0064 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Brown-Forsythe and Welch ANOVA test with Dunnett T3 correction (*, p ≤ 0.05, **, p ≤ 0.01). All experiments were performed in triplicate, using independent biological replicates. 51

Figure 26- LYPLA2 relative expression upon treatment with compounds (R)-(2j) and (S)-(2b) across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analyzed by RT-qPCR. A-C. LYPLA2 relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon treatment with compounds (R)-(2j) (p=0.9603 MCF-10A; p=0.0032 MDA-MB-231; p=0.0127 MCF-7), and compound (S)-(2b) (p=0.1524 MCF-10A; p=0.0054 MDA-MB-231; p=0.0047 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Brown-Forsythe and Welch ANOVA test with Dunnett T3 correction (*, p ≤ 0.05, **, p ≤ 0.01). All experiments were performed in triplicate, using independent biological replicates. 52

Figure 27- HOXB7 relative expression upon treatment with compound (1a), (1b) and (1c), across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analyzed by RT-qPCR. A-C. HOXB7 relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon treatment with compounds (1a) (p=0.0133 MCF-10A; p=0.0048 MDA-MB-231; p=0.0001 MCF-7), compound (1b) (p=0.0012 MCF-10A; p=0.0733 MDA-MB-231; p=0.0002 MCF-7), and compound (1c) (p=0.0004 MCF-10A; p=0.0145 MDA-MB-231; p=0.0006 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Brown-Forsythe and Welch ANOVA test with Dunnett T3 correction (*, p ≤ 0.05, **, p ≤ 0.01, ***, p ≤ 0.001). All experiments were performed in triplicate, using independent biological replicates. ..53

Figure 28- HOXB7 relative expression upon treatment with compound (R)-(2j) and (S)-(2b) across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analyzed by RT-qPCR. A-C. HOXB7 relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon treatment with compounds (R)-(2j) (p=0.1426 MCF-10A; p=0.0005 MDA-MB-231; p=0.0001 MCF-7), and compound (S)-(2b) (p=0.1313 MCF-10A; p=0.0348 MDA-MB-231; p=0.0033 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Brown-Forsythe and Welch ANOVA test with Dunnett T3 correction (*, p ≤ 0.05, **, p ≤ 0.01, ***, p ≤ 0.001). All experiments were performed in triplicate, using independent biological replicates.....54

LIST OF ABBREVIATIONS

%ee	Percentage of Enantiomeric Excess
A	Adenine
Acetyl-CoA	Acetyl Coenzyme A
AJCC	American Joint Committee on Cancer
AKT	Protein Kinase B
Ang2	Angiopoietin-2
ANOVA	Analysis of Variance
APT-2	Acyl-Protein Thioesterase
ASR	Age-Adjusted Standardized Rate
ASXL	Additional Sex Combs-Like 1
ATM	ATM Serine/Threonine Kinase
ATP	Adenosine 5'-Triphosphate
BARD1	BRCA1 Associated RING Domain 1
BCL-XL	B-cell lymphoma-extra large
BISMA	Bisulphite Sequencing DNA Methylation Analysis
BLBC	Basal-Like Breast Cancer
BMI	Body Mass Index
bp	Base Pairs
BrCa	Breast Cancer
BRCA	BReast CAncer gene
BSE	Breast Self-Exam
C	Cytosine
CBE	Clinical Breast Exam
CBP	CREB Binding Protein
CDH1	E-Cadherin
CDK	Cyclin-Dependent Kinases
CDKN	Cyclin Dependent Kinase Inhibitor
cDNA	Complementary DNA
CEACAM	Carcinoembryonic Antigen-Related Adhesion Molecules
CHEK	Checkpoint Kinase

CI	Confidence Level Intervals
CK	Cytokeratin
CREB	cAMP Response Element-Binding Protein
CT	Threshold cycle
CTNNB1	ß-catenin
DCIS	Ductal Carcinoma *In Situ*
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNMT	DNA methyltransferase
E	Primer Efficiency
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial To Mesenchymal Transition
EP300	E1A Binding Protein P300
ER	Oestrogen Receptor
ESR1	Oestrogen Receptor 1
EU	European Union
EZH	Enhancer of Zeste Homolog
FBS	Foetal Bovine Serum
FGF2	Basic Fibroblast Growth Factor
FW	Forward
G	Guanine
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GRO	Growth-regulated oncogene
H3	Histone H3
H3K27	Histone H3 lysine 27
H3K27me3	Histone H3 lysine 27 tri-methylation
H3K36me3	Histone H3 lysine 36 tri-methylation
H3K4me3	Histone H3 lysine 4 tri-methylation
H3K9me3	Histone H3 lysine 9 tri-methylation
H4	Histone H4

HAT	Histone Acetyl Transferase
HD	Homeodomain
HDAC	Histone Deacetylase
HDM	Histone Demethylases
HER2	Human Epidermal Growth Factor Receptor 2
HMT	Histone Methyltransferases
HPLC	High Performance Liquid Chromatography
HRT	Hormonal Replacement Therapy
HX	Hexapeptide Motif
IC50	Half Maximal Inhibitory Concentration
IDC	Invasive Ductal Carcinoma
IL	Interleukin
ILC	Invasive Lobular Carcinoma
IPTG	Isopropyl β- d-1-thiogalactopyranoside
KAT3B	Histone Lysine Acetyltransferase
kb	1000 bp
LB	Luria Broth
LCIS	Lobular Carcinoma *In Situ*
LPCC	Liga Portuguesa Contra o Cancro
LYPLA2	Acyl-Protein Thioesterase 2
MAPK	Mitogen-Activated Protein Kinase
MCS	Multiple Cloning Site
MG	Mammary Gland
MTT	3-(4,5-Dimethylthiazol-2-Yl)-2,5-Diphenyltetrazolium Bromide
NHEJ	Non-Homologous End Joining
PALB2	Partner And Localizer of Brca2
PARP1	Poly [ADP-Ribose] Polymerase 1
PC	Polycomb
PCR	Polymerase Chain Reaction
PRC2	Polycomb Repressive Complex 2
PR	Progesterone Receptor
PREP	Prolyl Endopeptidase

Ras	RAt Sarcoma virus
RhoA	Ras Homolog Family Member A
RNA	Ribonucleic Acid
rpm	Revolutions Per Minute
RT	Room Temperature
RT-qPCR	Real-Time Quantitative PCR
RV	Reverse
SC	Stem Cells
SD	Standard Deviation
SDS	Sodium Dodecyl Sulphate
SEA	*Similarity Ensemble Approach*
STK11	Serine/Threonine Kinase 11
SUMO	Small Ubiquitin-Like Modifiers
SWI/SNF	SWItch/Sucrose Non-Fermentable
T	Thymine
TAE	Tris-Acetate-EDTA
TALE	Transcription Activator-Like Effectors
TDLU	Terminal Ductal Lobular Unit
T-DMR	Tissue-Specific Differentially Methylated DNA Regions
TF	Transcription Factors
TNBC	Triple Negative Breast Cancer
TNM	Tumour, Node, And Metastasis
TOP2A	DNA Topoisomerase II Alpha
TSA	Trichostatin A
VEGF	Vascular Endothelial Growth Factor
X-Gal	5-bromo-4-cloro-3-indoxil-β-D-galactopiranosídeo
ZdCyd	5-aza-2′-deoxycytidine

1. INTRODUCTION

1.1. Breast Cancer

1.1.1. Epidemiology

Breast Cancer (BrCa) is the most incident cancer in women and the leading cause of cancer-related death amongst this group [1-3], both in high-income countries and middle-income ones [3] (Figure 1). In contrast, it has a relatively low incidence in men, comprising only 1% of the total diagnosis [4].

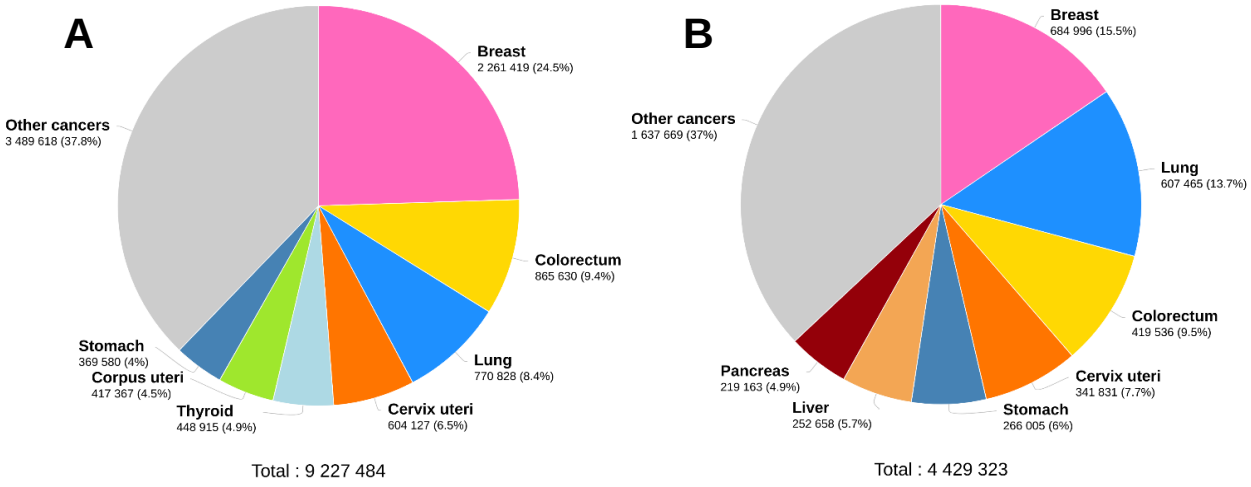

Figure 1- Worldwide cancer incidence (A) and mortality (B) rates among women. Source: GLOBOCAN 2020. Graph Production: Global Cancer Observatory – International Agency for Research on Cancer (http://gco.iarc.fr). Extracted from the World Health Organization Website (accessed on 31/12/2020).

In 2018, the World Health Organization estimated approximately 2 088 849 new cases of female BrCa [5], with 404,920 of those in the European Union alone, corresponding to an age-adjusted standardized rate (ASR) of 144.9 per 100,000 habitants [6] (Figure 2A). Moreover, BrCa accounted for 138,000 deaths in women, a total of 16,2% of all cancer-related deaths in forty European countries in 2018 [6], approximately 13,2 deaths per 100,000 habitants, doubling when only considering the female population [7]. This ASR has been decreasing over the past years [8], declining from 17,9/100,000 in 2002, 15,2 in 2012, and 14,4 in 2017 [6, 7, 9]. Nonetheless,

survival rates are higher in northern and western Europe when compared to the rest of the continent (Figure 2B). Overall, it is estimated that 1 in every 8 women in the EU-28 will develop BrCa before the age of 85 [10], being that the likelihood of developing it after the age of 65 is nine times higher than the chances of cancer development at younger ages [7]. Amongst EU members, the highest standardised death rates for BrCa among women were recorded in Slovakia and Luxembourg, corresponding to a 40,7 and 40,3 death rate per 100 000 inhabitants, respectively, while Spain scores the lowest (23.7/100 000) [7], hereby suggesting that BrCa's mortality rate is dependent on national health system policies.

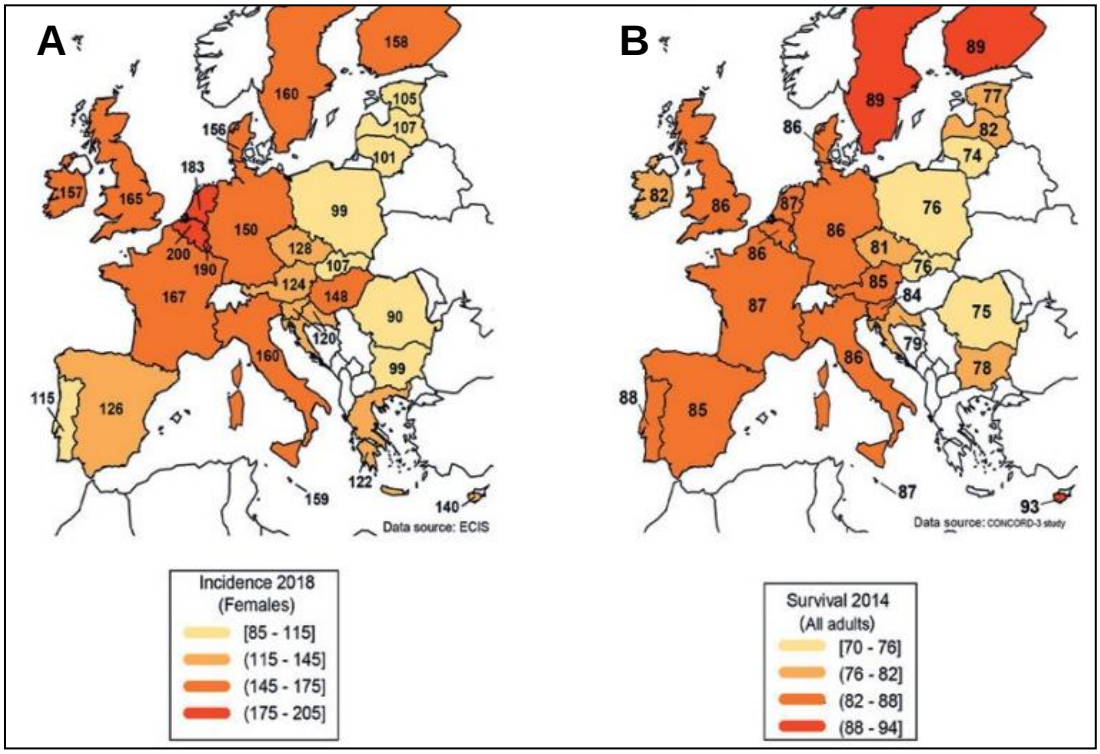

Figure 2- Breast cancer incidence (A) and survival (B) rates across EU-28 countries. A: Age-standardized incidence rates in European women, in 2018. B: Five-year survival rate in EU-28 citizens, in 2014. Adapted from [6].

The financial, physical, and emotional burden caused by cancer tends to increase over the next decades, due to a demographic tendency towards an aging population [5, 6]. However, in spite of its growing incidence rates, the morbidity caused by BrCa tends to decrease, due to the implementation and widespread of early screening and adjuvant therapies [3, 5, 6, 11]. Additionally, these temporal increases in BrCa occurrence can be pinpointed to a number of factors, including delayed childbearing, lower parity, increased usage of contraceptive methods, and a higher likelihood of obesity [5, 12].

An identical trend can be observed amongst the Portuguese population, where overall incidence has been increasing, particularly in the Northern region [12]. These results might be influenced by the increasing number of screenings performed in the north, compared to the centre region, where the Portuguese League Against Cancer (LPCC)'s screening program has been longer and more widely established, and in which the rollout has already been completed [12]. Nonetheless, Portugal stands out as the country with the second-highest screening rate in the EU-28, with more than 80% of women aged between 50 and 69 reporting to have been screened in 2013 [7, 12].

1.1.2. Risk Factors

With the increased burden caused by BrCa, especially among women, it is crucial to pinpoint the risk factors associated with the onset of the disease with the purpose of preventing it and adjusting screening strategies. Up to date, some categories have been established for BrCa risk factors, including demographic, reproductive, and hormonal factors, life habits, family history, and hereditary factors [1, 2, 5, 6, 13].

Some demographic risk factors for BrCa include gender, age, and race. As previously mentioned, BrCa is a malignancy that affects women almost exclusively (around 99%) [4]. Nevertheless, male BrCa patients are often older males, with a history of hormonal imbalances, exposure to radiation, or family history of BrCa, especially if carriers of mutations in the BRCA2 gene [1, 2, 4]. Age has been strongly associated with BrCa diagnosis, progressively increasing with age and greatly affecting women over 65 years [6, 14], possibly due to the accumulation of mutational events throughout women's lifetime. In addition, some studies have found an association between the age of menopause onset and the risk of developing BrCa, possibly related to a more extended exposure to endogenous hormones [6]. Two out of every three cases of invasive BrCa occur in post-menopausal women over 55 years, although in younger women, it tends to be in more advanced, lymph node-positive stages, resulting in higher chances of relapse and death [2, 6, 14, 15]. Therefore, diagnosis at younger age poses an adverse risk factor, suggestive of a worse prognosis [2]. Upon these results, several studies point to 40 to 45 years as the ideal age to start BrCa screening, except in cases of patients with a family history of BrCa, whose screening should begin at the age of 25 [15]. In the United States, "African-Americans" have been considered as having an increased risk to develop BrCa, especially among young women (<45 years), who are more likely to be diagnosed with grade III-IV tumours and triple negative BrCa (TNBC) than any other ethnic group [16]. However, incidence rates among women considered "Caucasian", "Asian"

or “Pacific Islanders”, between 20 to 34 years old, have been increasing in that country, while women considered “African-American”, “American Indian”, “Alaska Native” or “Hispanic” tend to be diagnosed at later stages [1, 16, 17].

The BrCa incidence also relates to family history, especially if the disease has been diagnosed in first-degree relatives before the age of 50 [18]. The genetic background of each individual might be the key to explain this increased incidence, especially if alterations in oncogenes or tumour suppressors are present [19]. Most cases of hereditary BrCa are associated with BRCA1 and BRCA2 mutations, which deficit proper homologous recombination, accumulating damage to the DNA structure, that is unable to be properly repaired [20]. Women with a family history of BrCa are often screened for this mutation and undergo closer surveillance. Following this example, and with the discovery of new BrCa associated mutations, multigene panel testing has been slowly implemented and considered to, simultaneously and cost-effectively, detect mutation carriers [19]. Amongst common candidates for these screenings are mutated variations of *PALB2*, *ATM*, *CHEK2*, *BARD1*, *RAD51*, *STK11*, *CDH1*, *NBN NF1*, and p53 [18-22]. Carriers of these mutations are more likely to develop BrCa at an early age than non-carriers, often displaying a more aggressive phenotype [2, 19]. Therefore, its early screening is considered beneficial to reduce BrCa risk and improve overall survival.

The correlation between reproductive factors and BrCa is associated with the exposure time to high levels of endogenous oestrogens. Several studies have shown that BrCa risk was positively related to early age at menarche (younger than 12 years) [23], while those whose menarche happened after the age of 15 were less susceptible to this disease [13, 24]. On the other hand, later menopause (>50 years) was also a susceptibility factor for breast malignancy [1, 25]. Moreover, nulliparity seems to decrease the risk of BrCa from the triple negative molecular sub-type, while increasing the risk of developing ER+ sub-types [26] and the overall risk of BrCa in older women [27, 28]. The protective effects of parity on ER+ sub-types can be explained due to the diligent changes that occur in the mammary gland during pregnancy, making it terminally differentiated and therefore less prone to accumulate potential carcinogenic factors [27]. These phenomena were also associated with alterations in stem cell activity, hormonal sensitivity, and environment and epithelial-stromal interactions that occur during and after pregnancy, that might justify the protective effect against ER+ sub-types of parity and first full-term pregnancy at a younger age [27-30].

The effects of oestrogen and progesterone exposure go beyond their endogenous exposure, given that women nowadays are often voluntarily exposed to additional doses of these hormones,

through the administration of oral contraceptives and hormonal replacement therapy (HRT). In spite of its numerous advantages, the use of oral contraceptives has been shown to increase BrCa risk amongst its current users, but also amongst women that have recently stopped using hormonal contraception, being that this risk tends to increase with longer usage [26, 28, 31-33]. However, the additional hormonal intake was proposed to have no effect on total cancer risk [32, 33]. The same ambiguity is present in the meta-analysis concerning HRT. BrCa survivors commonly request HRT in order to relieve menopausal symptoms, although its usage is not recommended for women in this group, as it can increase the risk of recurrence [34]. Conversely, carriers of BRCA mutation might be eligible for HRT if they have no history of BrCa, and it is in fact recommended in carriers who have performed a prophylactic mastectomy. Nonetheless, administrated dosage should be low, and done under supervision, with a preference for oestrogen-only HRT [35]. Furthermore, while many studies have reported that HRT increases BrCa risk in low to moderate risk women [30, 35], recent data contradicts it, stating that it might not play a role in the increased risk of BrCa, or that its effects are still inconclusive [36, 37]. Whether this is due to the current advances and hormonal adjustments in both methods, or to the effective stagnation of its associated risks, more studies should be performed to achieve a definite conclusion.

In addition to the features presented above, lifestyle factors have been shown to have an important role in BrCa development, such as diet habits, alcohol intake [38], smoking [13], diabetes [13], high exposure to radiation [39], breast lesions [40], high BMI index [41] and central adiposity [42]. Nonetheless, the risk evaluation needs to take into account the fact that multiple and perhaps simultaneous factors play a role in cancer development.

1.1.3. Diagnostic and Screening

With the rising incidence of BrCa and women's exposure to risk factors, early screening and diagnosis become increasingly important. Many governments have implemented screening programs that aim to find breast tumours at an early stage, and therefore increase its treatment and survival rates [43]. Tumour size is highly suggestive of prognosis, and thus women with larger tumours have lower chances of survival [44]. Currently, three reliable and cost-effective procedures are being implemented for BrCa screening: mammography, clinical breast exam (CBE) and breast self-exam (BSE) [44].

Amongst these methods, mammography is the most effective and sensible means of BrCa detection, as it can detect small pre-palpable tumours (≤15 mm) before any clinical manifestations, with an accuracy rate of 85-90%. This screening reduces mortality by 30-50%, since tumours at this stage have a better prognosis, and is routinely recommended for age-appropriate, asymptomatic women and high-risk patients [44]. High-risk patients are also recommended to perform a CBE, which might include breast magnetic resonance (MRI) or ultrasound exams, especially in the case of women with dense breast tissue, breast implants, or who are not eligible for mammography [43-46]. Nonetheless, mass mammography and CBE screenings are still unattainable in low to middle-income countries, where late diagnosis results in 62% of all BrCa deaths. Therefore, BSE and breast awareness programs, which consists of the manual palpation of breast tissue and surrounding nodal region as well as the education towards its importance and practice, are being offered as a temporary alternative in these countries, since this method does not entail any costs [44, 47].

Nonetheless, BrCa can only be confirmed upon tissue sampling and cytology/histology of possibly involved axillary nodes, through core needle biopsy, and analysed by a pathologist. The pathologist will classify the tumour based on the WHO and the eighth edition of the American Joint Committee on Cancer (AJCC) tumour, node, and metastasis (TNM) staging system [48].

1.1.4. Histopathology

BrCa is extraordinarily heterogeneous from a clinical, morphological, and behavioural perspective, which accentuates the struggle of finding proper distinct classifications for these tumours, both from a molecular and histopathologic standpoint. Given that BrCa can arise from any type of cell within the mammary gland, the classification into each subtype, with the same characteristics and that might, consequently, have similar clinical courses and outcomes, is crucial [49].

Current histopathologic classification divides tumours into two large groups based on their invasive status (*In situ* or Invasive), and their origin (Ductal or Lobular) (Figure 3). This variation in cancer growth is mostly determined by the status of E-cadherin expression.

The ductal carcinoma *in situ* (DCIS) is a non-invasive neoplastic proliferation of ductal epithelial cells, with malignant potential, that can later develop into invasive BrCa [40, 49, 50]. These neoplastic cells are usually comprised within the outer ductal layer, although the latter can be weakened. Tumours from this subtype can precede invasive lesions and are unswervingly

positive for E-cadherin and ß-catenin. Current classification divides DCIS into five subtypes: comedo, solid, cribriform, papillary, and micropapillary [49].

The lobular carcinoma *in situ* (LCIS) is characterized by the abnormal proliferation of small and roughly cohesive neoplastic cells inside the lobule, that originated in the terminal duct lobular unit (TDLU) [49, 51]. Likewise, DCIS might progress into an invasive lesion, although not compulsorily. However, E-cadherin and β-catenin expressions are absent in this subtype [49].

Invasive or infiltrative carcinoma is a malignant neoplastic proliferation that has penetrated through the duct or lobular wall and invaded the stroma. This classification comprehends the Invasive Ductal Carcinoma (IDC) and Invasive Lobular Carcinoma (ILC) subtypes [49, 52].

IDC is the most commonly diagnosed invasive BrCa, accounting for 55% of BrCa cases upon diagnosis. While some IDC can be classified within the wide range of IDC subtypes, its highly heterogenous natured impairs the classification of 75% of IDC, solely based on cytoarchitectural features [52, 53].

Lastly, a small portion of breast carcinomas (5-15%) can be classified into the ILC category. This type of carcinoma is often found in postmenopausal women, possibly associated with HRT [40, 49, 52, 54, 55]. Histological grading divides ILC into five variants: classic type, pleomorphic, histiocytoid, signet ring and tubulolobular [49, 54, 55]. E-cadherin is often inactivated through mutational processes, loss of heterozygosity or methylation [49].

1.1.5. Molecular subtypes and predictive factors

As demonstrated in the previous subtopic, histological classification cannot be exclusively used to classify tumours, as it alone is insufficient to give a clear input into a patient's clinical course and outcome [49]. Over the last decades, molecular markers have been associated with phenotypic characteristics, response to therapies and prognosis, and have therefore been implemented as a supplement in BrCa classification [49, 56]. This molecular distinction was first classified by Perou et al., based on the present or absent expression of oestrogen receptor (ER), and has been refined throughout the years [20, 56-60]. Hereupon, BrCa was classified into four molecular subtypes: Luminal A, Luminal B, HER-2 enriched and Basal-Like Breast Cancers (BLBC) (Figure 3) [49, 56-60].

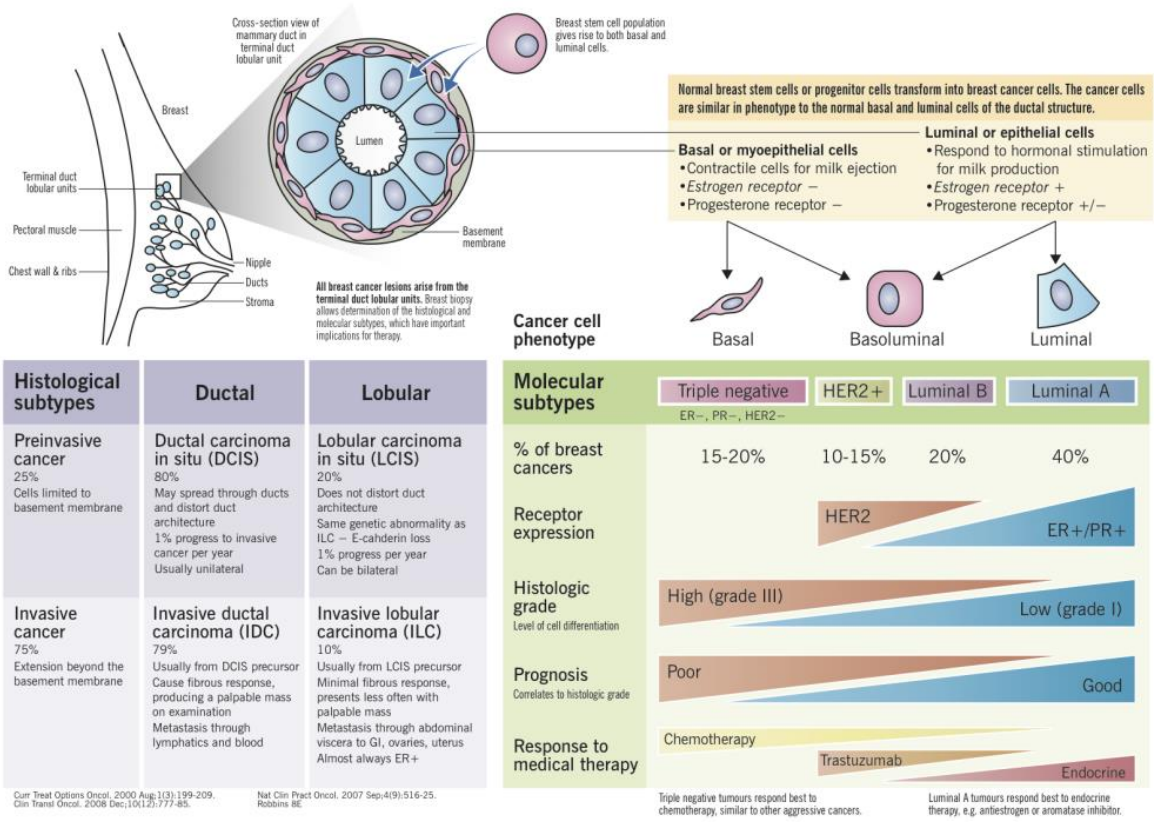

Figure 3- BC pathogenesis and histologic versus molecular subtypes, by Eric Wong and Joana Rebelo.

Luminal A and Luminal B tumours are positive for ER and progesterone receptor (PR) and were named after their genetic resemblance to normal luminal epithelial cells of the breast [49, 56, 58]. Luminal A subtype represents 40-50% of all invasive BrCa, making it the most common subtype [56] and is often HER2-negative. Nonetheless, these tumours are usually low grade and are more likely to benefit from hormonal therapy, without the need for additional chemotherapy, which makes them the subtype with the best prognosis [49, 56-60]. Opposingly, Luminal B tumours usually require supplementary therapy and present a worse prognosis, as they are frequently higher grade, and are not as responsive to hormonal therapy. Even though they are ER+ and PR+, their HER2+ expression is variable, and they present a higher expression of proliferation-related genes, compared to Luminal A [49, 56-61].

HER2 overexpressing tumours comprise 15% of all invasive BrCa and are characterized by the upregulation of HER2 and other genes present in the HER2 amplicon. In addition, they are ER and PR-negative, and often present a mutated p53. As a consequence of their genetic profiles and higher-grade stages at diagnostic, these tumours have a more aggressive phenotype than

the Luminal subtypes. However, they are highly sensitive to anti-HER2 targeted therapy trastuzumab (Herceptin), which can greatly improve their prognostic [49, 56-60].

The BLBC subtype accounts for 15% of all invasive BrCa and most high-grade invasive cancers. BLBC presents an expression profile that is similar to basal epithelial cells and normal myoepithelial cells of the mammary gland. They display an increase in CK5, CK6 and EGFR expression, as well as proliferation associate genes, but most cases are negative for ER, PR and HER2, which is why it is often referred to as Triple Negative subtype. Due to the absence of these markers, they are not responsive to endocrine therapy or trastuzumab, therefore being often associated with worse prognosis and relapse within 5 years after diagnosis [49, 56-60]. Nonetheless, recent studies have shown sensitivity to platinum-based therapies [62-64] and poly (adenosine diphosphate-ribose) polymerase inhibitors [65-68].

Ultimately, molecular classification and biomarker identification provides an adjuvant method for BrCa classification, which will aid health professionals in making therapeutic choices and predict whether or not a patient might benefit from a certain therapeutic approach [49, 56, 59, 60, 69].

1.2. HOXOME

1.2.1. The *HOX* family

The *HOX* genes are a group of highly conserved homeotic genes, comprising a subset of 39 genes in humans, organized into 4 clusters: *HOXA* (7p15), *HOXB* (17q21.2), *HOXC* (12q13), and *HOXD* (2q31). Each cluster comprehends 9 to 11 genes that are aligned into 13 paralogous groups in consonance with sequence homology (

Figure 4) [70, 71].

These genes were first discovered in *Drosophila melanogaster* and encode transcription factors (TF) that work as master transcriptional regulators and determine cellular identity during embryonic morphogenesis and postnatal life [70, 72, 73]. *HOX* genes contain two exons and one intron and are characterized by the presence of a 180-nucleotide sequence, located on exon 2, which encodes for a 60 amino acid homeodomain (HD) [72]. The homeodomain is a helix-turn-helix motif that is capable of recognizing specific DNA sequences containing core motifs

(TAAT/ATTA/TTAT/ATAA), which allow the transcriptional regulation of their target genes [70, 72].

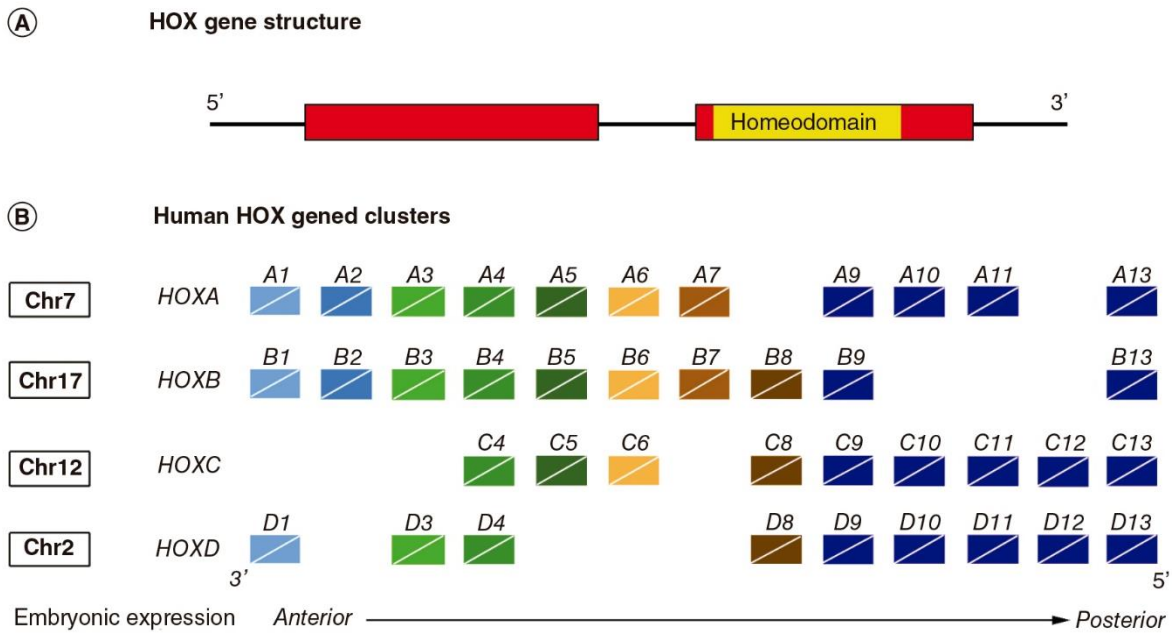

Figure 4- Arrangement of the HOX clusters in the human genome and their connection to cancer. A. *HOX* gene structure highlighting the homeodomain. **B.** Thirty-nine genes dispersed by four clusters (HOXA, HOXB, HOXC, HOXD) located on four distinct chromosomes. In each cluster, genes are arranged in tandem sequence from 3' to 5', which is relevant for the temporal and spatial activation of each *HOX* gene during embryonic development. The white bar emphasizes that all *HOX* genes have been found related to cancer progression. *HOX* genes with the same number are referred to as 'paralogs', in [70].

Although the recognition of these motifs suggests low functional specificity, the recruitment of different cofactors and other transcription factors conduces to the formation of multi-protein complexes, the "Hoxasomes", which result in high specificity and stability of DNA-binding, characteristic of these complexes [72]. These complexes are formed due to the presence of the Hexapeptide motif (HX), located N-terminal to the HOX homeodomain, encoded by exon 1 [72]. Besides aiding the regulation of HOX putative targets, these cofactors participate in post-translational regulation of HOX proteins, aiding its entry into the nucleus [74], and mediate the recruitment of RNA polymerase II and III or transcriptional inhibitors such as histone deacetylases [74]. Among HOX cofactors, the ones that are best characterized are comprised in the TALE (three amino acid loop extension) family, which includes proteins such as PBX (PBX1, 2, 3, and 4), MEIS (MEIS1, 2, and 3) and PREP (PREP1 and 2). Even though they differentially regulate HOX proteins, all of these proteins share highly conserved HD that enable DNA binding, increasing the specificity of these complexes and HOX-target gene interaction, being that PBX

proteins interact with HOX proteins 1-11 [75], MEIS family members bind to HOX proteins 9-13 [75], while PREP tends to bind to promoters and sequences with promoter-like epigenetic marks, frequently interacting with HOX binding sequences in the presence of MEIS [76]. Furthermore, PBX proteins have been shown to form trimers with PREP/MEIS and HOX proteins in a manner that modifies the DNA-binding specificity of HOX proteins [76]. The protein dynamics between both families point towards the presence of widespread autoregulatory phenomena and crosstalk between HOX/TALE expression networks [74, 76, 77].

The role of *HOX* genes in embryonic development and organogenesis has been well established and characterized, being known that their expression patterns follow three elementary principles: spatial and temporal collinearity, and posterior prevalence [73, 78, 79]. Following temporal and spatial collinearity, a unidirectional chromatin opening compels *HOX* genes to be expressed in a manner where 3' genes express afore 5' genes, and to follow a pattern of expression from 3' to 5' along the embryo's anterior-posterior axis, enabling the allocation of segmental identity in the embryo's body [73, 78, 79]. Furthermore, the posterior prevalence principle determines that 5' *HOX* genes exhibit a dominant phenotype over genes located in the 3' of the same *HOX* cluster [73, 78, 79]. The manifestation of these principles results in nested and overlapping expression domains, determinants for proper embryogenesis [71, 73, 78].

Despite its renowned role in embryogenesis and organogenesis, most *HOX* genes are important for the homeostasis of adult cells [71]. In fact, *HOX* genes act as determinants of cell and tissue identity in processes occurring in adulthood, such as mammary development or differentiation of adult stem cells (SC), as they regulate processes such as cell division, adhesion, migration, differentiation, and apoptosis. Furthermore, they have been shown to produce non-coding RNAs that play a regulatory role in development, cell homeostasis, and cancer [71, 72], to contribute to DNA replication and repair, mRNA translation, protein degradation and to act as epigenetic regulators [70]. Hence, the involvement of HOX proteins in oncological contexts comes as no surprise, as many of the above-mentioned genetic and cellular processes are altered in cancer. In fact, several HOX proteins have been linked to a variety of cancer types, although their impact on the molecular networks behind cancer progression remains largely unknown [70-72, 74, 79-83].

1.2.2. *HOX* genes in cancer

The neoplastic process is intimately associated with alterations in cellular proliferation, differentiation, and tissue organization. Although these *de novo* processes are often present during embryogenic processes, there appear to occur in an aberrant manner during oncogenesis. As previously demonstrated, many of these cellular processes have been shown to be regulated by *HOX* genes, hence, the dysregulation of these cellular events hinted towards the involvement of these genes in carcinogenesis. In fact, the “Oncogerminative Theory of Tumorigenesis” states that the dysregulation of developmental genes, such as those from the Homeobox family, is implicated in tumour development, as this neoplastic process mimics the self-organizing dynamics observed in embryogenesis [84].

The potential role of *HOX* genes was first described in haematological cancers, as its dysregulation or chimeric association in fusion genes resulted in aberrant hematopoietic processes [74, 85]. Upon this evidence, these genes started to be more broadly studied in both liquid and solid malignancies [74]. In an attempt to explain the mechanisms underlying the involvement of *HOX* genes in the oncogenic process, Abate-Shen and colleagues have proposed three mechanisms that could justify HOX’s modulating capacity in carcinogenesis: temporospatial dysregulation, gene dominance, and epigenetic dysregulation [86, 87]. The temporospatial dysregulation foresees an altered pattern of *HOX* expression in tumorigenic cells, that can be caused by its re-expression in cells that once expressed these genes during development; by its *de novo* expression in cells whose lineage did not express it at any point of embryogenesis; or by its downregulation in cells derived from fully differentiated tissues that usually express *HOX* genes [86, 87]. Gene dominance is inferred on the premises that the “acquisition” or increased expression of *HOX* genes often confers cells carcinogenic proprieties, different from those displayed during development, although this phenomenon might be differently manifested depending on the cell type and particular gene expression profiles [86]. Lastly, it is proposed that epigenetic regulation, such as loss or gain of methylation, or histone modification, can result in aberrant expression profiles, especially at key timepoints, ultimately leading to the gain or loss of tissue-specific functions, particularly when they play a pro-oncogenic or oncosuppressor role in those tissues [86]. Hereby, given its tissue-specific action and variable role on the onset of carcinogenesis, *HOX* genes are proposed to be positive or negative tumour modulators, instead of classic oncogenes or tumour-suppressors (Figure 5) [70-72, 74, 85, 86].

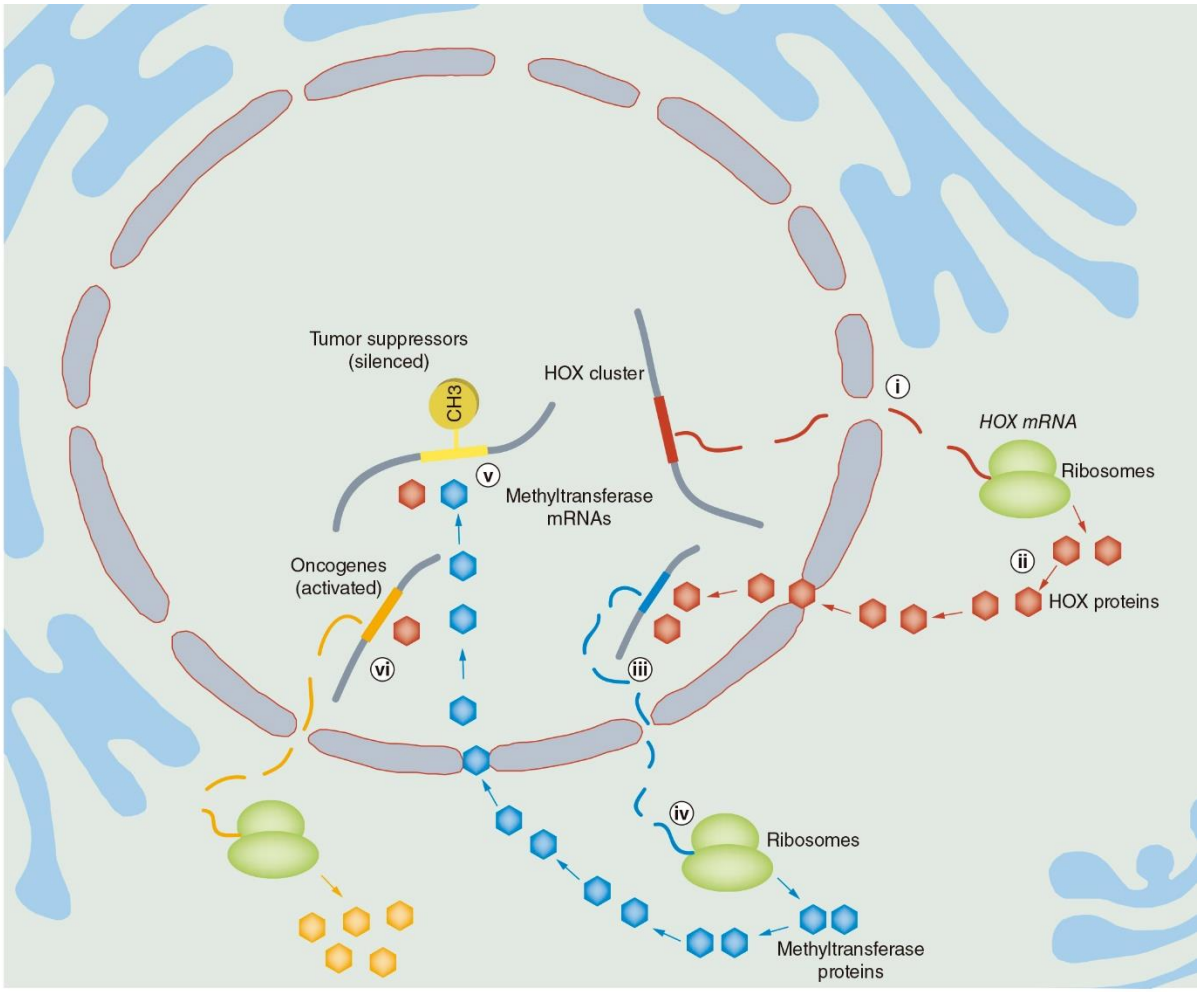

Figure 5 - Representation of the direct and indirect roles that HOX proteins potentially have in tumorigenesis. (i) After HOX gene transcription, mRNA is exported from the nucleus into the cytoplasm. (ii) *HOX* mRNA is then translated into HOX proteins. (iii) These proteins are translocated into the nucleus, affecting the expression of a set of genes that may include methyltransferases such as DNMT3B. (iv) mRNA from these genes are then exported to the cytoplasm and translated. (v) Methyltransferases are translocated into the nucleus, where they inhibit the expression of other genes by adding methyl group to their sequence. (vi) *HOX* genes may also regulate the expression of oncogenes or tumour-suppressor genes, in [70].

The involvement of *HOX* genes has been reported in different cancers, as it influences mechanisms that trigger cancer onset and progression, such as cell proliferation, apoptosis, migration, invasion, metastization, cell autophagy and motility, differentiation, angiogenesis, metabolism, metalloproteinases regulation, and inflammation [71, 72, 74, 79, 85]. These alterations are often related to an increased cancer progression and resistance to therapy, resulting in a poorer prognosis [70-72, 74, 79, 85]. In fact, a total of 39 *HOX* genes have been found dysregulated in several cancers, such as breast, lung, colon, and prostate cancers (Figure 6) [88]. Furthermore, studies have presented evidence suggesting that several molecules related

to cancer progression interfere with *HOX* gene expression, such as vitamin D, testosterone and progesterone (*HOXA10, HOXA11*), retinoic acid (*HOXA1, HOXB1, HOXD4, HOXA5*), and oestrogen (*HOXA1, HOXA10, HOXB7, HOXB13, HOXC10*) [89]. This aspect is particularly noteworthy as specific treatments can target these molecules [90-94].

Ergo, as research data supports *HOX* genes' involvement in the alteration of most cancer hallmarks, these genes show potential as therapeutic targets, both singularly, as a gene set, or by disrupting its binding to specific co-factors [74, 80, 82, 83].

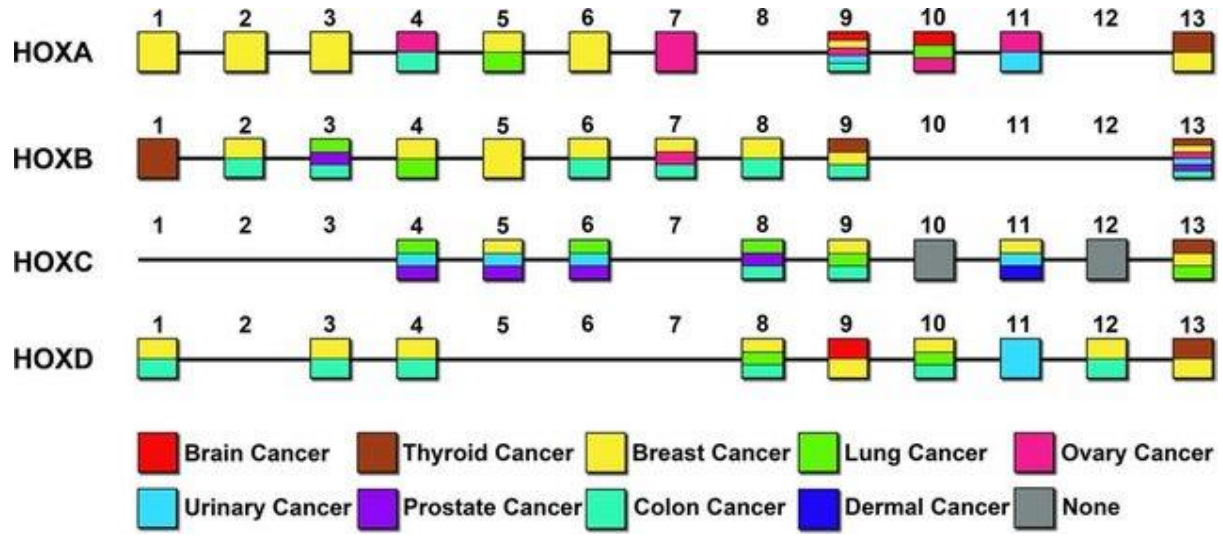

Figure 6 - Expression of HOX genes in the nine most common solid tumours, by Duan et al. 2015 [88]**.**

1.2.3. *HOX* genes in breast cancer

The mammary gland (MG) is an exocrine structure responsible for the production and excretion of milk, which is often fed to mammalian offspring, being crucial in their development [72]. MG development can be divided into 6 main phases: pre-puberty, puberty, pre-pregnancy, pregnancy, lactation, and involution (

Figure 7) [72, 95, 96]. During these phases, the MG undergoes structural changes due to modifications in cell proliferation and differentiation, that require coordinated developmental processes that, if abnormal, can lead to impairments in lactation or promote the onset and progression of breast carcinogenesis [72, 96, 97]. As master transcriptional regulators, several homeobox proteins are crucial for proper MG development, functionality, and remodelling, as they encode TFs that will regulate cell-specific processes [72, 97], and can be modulated by changes

in oestrogen and progesterone levels, thus interfering with a patient's response to hormonal therapy [89].

The MG starts its formation during embryogenesis, with the formation of the mammary line, in which the placodes differentiate, presenting a rudimentary duct tree by the time of birth. The regulation of these processes is performed with the participation of *HOXC8/D8, HOXA9/B9/C9, HOXB6,* and *HOXB3* [72, 96, 97]. Upon birth, *HOXC6* appears to partake in ductal tree branching and epithelium formation, possibly in response to ovarian hormones, and this expression is maintained throughout puberty and adulthood, although decreasing during pregnancy [72, 96, 98].

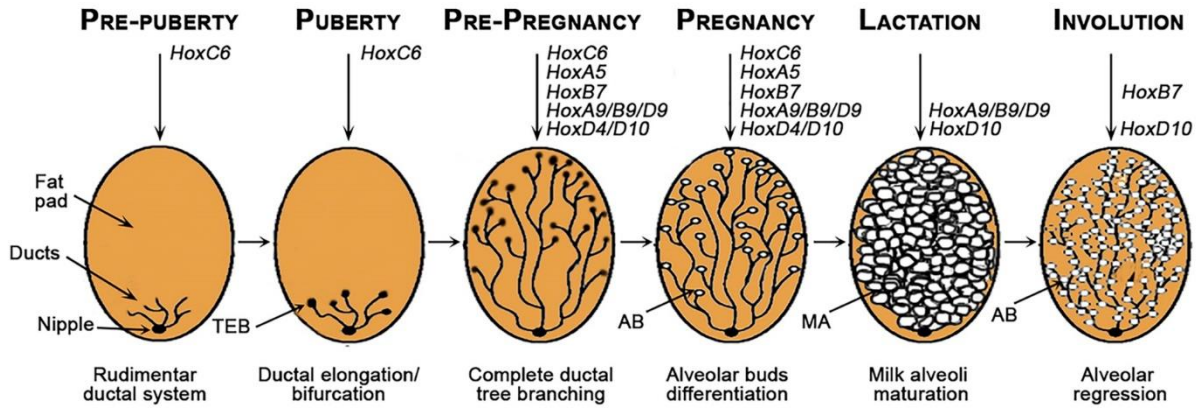

Figure 7- Hox gene expression and function in MG development. MG development throughout its six developmental stages: pre-puberty, puberty, pre-pregnancy, pregnancy, lactation, and involution. TEB: Terminal embryonic buds: AB: Alveolar buds; MA: mature milk alveoli. Adapted from [72].

During pre-pregnancy stages, additional genes such as *HOXA5, HOXA9, HOXB7, HOXB9 HOXD4, HOXD9,* and *HOXD10* are also expressed, concomitantly appearing in the early stages of pregnancy to regulate alveoli expansion, but appearing downregulated later on. It is during pregnancy that the MG is more sensible to progesterone and prolactin, which will enable the formation of alveoli capable of producing and secreting milk during lactation [72, 96, 97, 99, 100]. During the transition between pregnancy and lactation, *HOXD10* expression is particularly noteworthy, as it has been described that its disruption results in lactation impairment due to less expanded alveoli, suggesting that it participates in lobulo-alveolar differentiation during this period. The persistent expression of *HOXA9, HOXB9,* and *HOXD9* also appears critical to ensure breast functionality and lactation [72, 96, 97, 99].

Following lactation, the MG undergoes an involutory process that drastically remodels the gland to a pre-pregnancy state, due to the lack of prolactin and the onset of specific signalling pathways. During this stage, the MG stops expressing most of the previously expressed *HOX* genes, with the exception of *HOXD10* and *HOXB7 [72, 96, 97]*.

Given the role of *HOX* genes during MG development, is it expected that alterations in their expression profiles might result in the disruption of cellular processes that can give rise to breast cancer [72, 96, 97, 99, 101]. In fact, 35 out of the 39 human *HOX* genes have been described to be dysregulated in breast cancer [72, 102]. Nonetheless, given their complex regulation and the high heterogeneity nature of BrCa, some discrepant results concerning *HOX* genes putative functions in BrCa have been reported. This still hampers the interpretation of the precise role of most *HOX* genes in breast carcinogenesis. In fact, the *HOX* gene expression profiles in BrCa seem to be highly dependent on the tumour's molecular subtype [72, 102]. Luminal cell lines present a higher expression of *HOXD12/D13*, and *HOXA* and *HOXB* genes, although these clusters' expression is even higher in ER+/HER2+ cells [72]. On the other hand, tumours that are ER- but HER2+ show the lowest expression of *HOXA* and *HOXB* genes. Curiously, *HOXC* genes are consistently overexpressed in most cell lines [102]. Furthermore, the overexpression of *HOXA1, HOXA7, HOXB3, HOXB5, HOXB7, HOXB9, HOXB13*, and *HOXC6* seem to promote breast carcinogenesis by increasing cell proliferation, survival, migration, and invasion, similarly to the effects of *HOXA2, HOXA3, HOXA5, HOXA9, HOXC11, HOXD3, HOXD4, HOXD8, HOXD9,* and *HOXD10* downregulation [72, 102]. In fact, *HOXA5* appears to be an apoptosis regulator that inhibits epithelial to mesenchymal transition (EMT) by upregulating *CDH1* [103].

Amongst the four *HOX* clusters, the *HOXB* cluster is the best characterized and several of its genes appear upregulated in malignancies [72, 98, 104]. This cluster is situated on chromosome 17, often altered in malignant tissues due to alterations in its harbouring genes: *ERBB2, HER2, p53, BRCA1* and *TOP2A [104]*. Due to its close proximity to these important markers, some *HOXB* genes (such as *HOXB7, HOXB8* and *HOXB9*) are often concomitantly dysregulated in BrCa, associated with higher-grade tumours and resistance to hormonal therapy [72].

HOXB7 function is implicated in MG development, but its overexpression has been associated with breast malignancies, in cell lines and tumours, both at the primary tumour and metastatic sites [72, 105]. Furthermore, this gene has been associated with cancer progression and metastization, through EMT promotion (Figure 8) [106], and to tamoxifen resistance in Luminal subtypes, by activation of the EGFR pathway, or through the activation of receptor tyrosine kinase

pathways [107]. Additionally, the coexpression of *HOXB7/HER2/MYC* presents a significant prognostic factor to determine a patient's survival in a subset of endocrine therapy–treated ER-positive BrCa patients (Figure 8) [105].

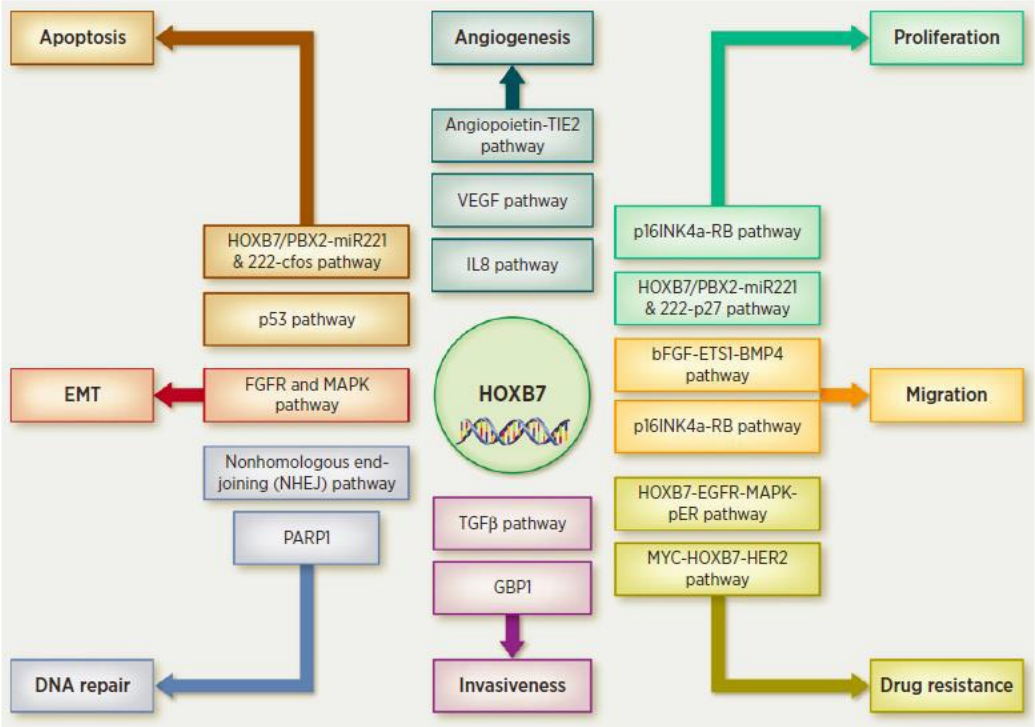

Figure 8- HOXB7 role in pathways critical to initiation and progression of solid cancers, namely, apoptosis, angiogenesis, proliferation, migration, drug resistance, invasiveness, DNA repair, and EMT. Adapted from [105].

The HOXB7 protein is a master regulator of several cellular processes that lead to carcinogenesis, either by overexpression or amplification, and is capable of regulating a wide range of target molecules in the oncogenic hierarchy [72, 105]. Studies have shown that this TF promotes EMT by partaking in the upregulation of basic fibroblast growth factor (FGF2) genes, and the initiation of the Mitogen-Activated Protein Kinase (*MAPK*) pathway (Figure 8) [71, 72, 105]. Additionally, this TF is also able to stimulate DNA double-strand breaks repair, through the mediation of factors involved in the non-homologous end-joining (*NHEJ*) pathway, or through the intervention of poly(ADP)ribose polymerase-1 (*PARP1*), which reinforces irradiation resistance, and allows the accumulation of mutations that can be determinant to trigger carcinogenesis (Figure 8) [105, 108]. Furthermore, it has been described to affect proliferation through p16NK4a-RB, 222-p27, and PBX2-miR221 pathways, being that the latter is also involved in apoptosis regulation (Figure 8) [105]. Regarding angiogenesis, live xenografts have shown that *HOXB7* induces angiogenesis

by upregulating proangiogenic factors VEGF, Ang2, GROα, and IL8 (Figure 8) [105, 109]. Moreover, the overexpression of *HOXB7* seems to induce transforming growth factor-beta (*TGFb2*)'s expression, enhancing cell motility and invasiveness, as well as macrophage recruitment and activation (Figure 8) [110].

Unlike *HOXB7*, the involvement of *HOXB8* in BrCa is still poorly characterized. Even though a study reported no expression of this gene on normal or neoplastic breast tissue [111], in vitro studies reported its expression on BrCa cells such as BT-474 and MDA-MB-231, but not MCF-10A, a cell line representative of the normal breast cells [81, 102]. The scarcity of concise information, makes this gene a potential candidate for further studies in BrCa, as its involvement has been implicated in other malignancies such as gastric, ovarian and colorectal cancers, with effects on cell migration, invasion, and metastasis [81, 112-117].

HOXB9 is another *HOXB* gene that has been found to be dysregulated in BrCa [81, 118]. In fact, its overexpression is associated with increased cell motility, angiogenesis and EMT induction, resulting in highly vascularized higher-grade tumours [81, 118, 119]. This gene's expression is often higher in ER+ tumours, suggesting that oestrogen might also play a role in this gene's regulation, likewise *HOXB7*. Up to date, five main molecular mechanisms have been described regarding *HOXB9*'s contribution toward BrCa progression: 1) induction of cell differentiation, apoptosis and EMT, through the activation of the *TGF-β* signalling pathway; 2) EMT induction and increased invasion through the increased production of ErbB ligands, resulting in higher phosphorylation and posterior activation of *HER2/3*; 3) increased differentiation by upregulation of *bFGF*; 4) higher expression of *VEGF*, thus promoting angiogenesis; 5) enhanced growth, differentiation and survival via *NRG2* [118-121]. Its importance in cancer progression makes it a key prognostic factor in a variety of cancer types.

1.2.4. *HOX* gene epigenetic regulation in cancer

Even though the study of the *HOX* gene function in cancer has been receiving more attention over the last decades, the mechanisms underlying their dysregulation are still poorly characterized. Recent research has looked into epigenetic changes as a critical pathway for cancer genesis, progression, and metastasis by influencing gene expression [122]. Moreover, epigenetic events can be also involved in the maintenance of the genomic structure, without altering the DNA sequence, being therefore reversible and a promising target for therapy [122].

DNA methylation, posttranslational modification of histones, chromatin remodelling, and RNA-based processes, are the four main epigenetic events connected to gene expression alterations [122].

One of the best studied abnormalities in human malignancies is DNA hypermethylation [123]. This process consists of the addition of a methyl group to a cytosine residue usually located in CpG dinucleotides within gene promoters, which is performed by DNA methyltransferases (DNMTs 1, 2, 3 and their variants) [124]. This addition allows the modification of the chromatin structure, which leads to transcriptional activation or repression, on a context-dependent matter [124]. DNMT1 is responsible for the maintenance of the methylation status of DNA upon replication; DNMT2 is a tRNA methyltransferase, while DNMT3 is involved in *de novo* DNA methylation, particularly important in the silencing of oncosuppressors and the initiation of carcinogenesis [122, 123]. Generically, the total degree of DNA methylation declines as a tumour progresses, but the frequency of hypermethylated CpG islands on promoters increases, resulting in the transcriptional suppression of tumour-suppressive genes, such as *BRCA1, CDKN2A, CEACAM6, E-cadherin (CDH1),* and *ESR1* [87, 123]. Additional data has revealed that CpG island methylation changed dramatically in pre-invasive lesions, compared to normal tissues, but was similar to the methylation profiles observed in invasive lesions, implying that CpG island methylation of tumour-related genes occurs early in breast cancer progression [125]. Moreover, DNA methylation status of hormonal receptors can be used to identify cohorts of patients that are likely to respond to hormonal therapy [126, 127]. Thus, the characterization of the methylation profiles of certain genes can be a useful asset in personalized treatment [128, 129].

In fact, several studies have correlated *HOX* gene dysregulation to alterations in the methylation profiles of their promoters, being pointed out as promising cancer biomarkers and therapeutic targets [70, 122, 123, 125, 130-135]. Genes such as *HOXA1* [125, 132], *HOXA5* [131, 136], *HOXA9* [123, 132], *HOXA10* [123, 125, 132], *HOXA11* [133, 134], *HOXB13* [132], *HOXC6* [130], *HOXC9* [131], *HOXC10* [137], and *HOXD13* [133, 135] have been shown to be hypermethylated in breast cancer tissues, and associated with cancer progression and worst prognosis. The methylation of numerous CpGs on the *HOXA* locus has been linked to the progression of breast cancer, suggesting that these genes can be robust prognostic markers [123]. *HOXA5* hypermethylation has also been associated with loss of *p53* expression [136]. Further findings reveal a link between specific methylation of *HOX* genes and *BRCA* mutations, as well as high expression of oestrogen and progesterone receptors and histological and clinical aggressiveness characteristics, implying the possibility of identifying patients at higher risk of breast cancer [132].

Despite being lesser studied, DNA hypomethylation has also been associated with cancer progression, being a phenomenon that affects a wide range of genomic regions and that is a trait present in many cancers [138]. Hypomethylation seems to affect chromatin stability, making it more prone to chromosomal rearrangements, being also able to promote the onset and development of cancers allowing the transcriptional activation of oncogenes [138]. In fact, hypomethylation of well-known oncogenes was identified in malignancies, as well as genes that are activated in cancer due to promoter hypomethylation but with a still unknown role in carcinogenesis [138]. The majority of these genes are members of the cancer-germline (CG) gene family, as their expression in healthy individuals is often limited to testicular germline cells, such as *MAGEA1* [138]. Although *HOX* genes are not included in the CG category, this subset of genes is often implicated in the oncogenic transformations predicted by the OTT theory, which stated that alterations in gene expression profiles can reprogram cells to mimic a germline state [84]. As such, the reactivation of *HOX* genes in tissues that no longer expressed them, or their overexpression, might be related to the hypomethylation of their promotors and influence the development of cancers, as shown in the case of *HOXC10* in gastric cancer [139].

In addition to DNA methylation, aberrant *HOX* gene expression observed in many cancers may be caused by histone changes [122]. These modifications have an impact on the nucleosome's assembly and restructure, which consists of a complex formed by the twice wrapping of a DNA strand around an octamer containing four core histone proteins (H2A, H2B, H3, and H4). These histones can be modified by acetylation and methylation of lysins (K) and arginines (R), respectively, through phosphorylation of serines (S) and threonines (T), ubiquitylation (using an isopeptide bond between lysine residues) and sumoylation (using small ubiquitin-like modifiers) (SUMOs) [122]. Acetyltransferases (HAT), deacetylases (HDAC), methyltransferases (HMT), demethylases (HDM), and kinases are among the enzymes involved in these reactions and interact with ATP-dependent chromatin-remodelling complexes that detect particular histone modifications, influencing nucleosome disassembly and assembly as well as histone octamer movement along the DNA [122].

Chromatin remodelling may improve the accessibility of DNA, allowing transcription factors to interact more easily, or may promote the packing of chromatin into densely folded structures, limiting interactions [122]. As a result, non-methylated CpG island promoters have looser chromatin, whereas methylation promoters have tighter chromatin. Some specific methylation patterns are considered hallmarks of active chromatin such as trimethylation of lysine 4 or 36 on histone H3 (H3K4me3 and H3K36me3), and hyperacetylation of histones H3 and H4, while others

partake in gene silencing, such as the trimethylation of lysins 9 and 27 on histone H3 (H3K9me3 and H3K27me3), concomitantly with the hypoacetylation of histone H3 [122].

The PRC2 complex (polycomb repressive complex 2) promotes H3K27me3 [70, 122], which subsequently facilitates binding of Polycomb (PC), a component of the PRC1 complex, to histone H3 amino-terminal tail [140]. Patients with a mutation in the sex combs-like 1 gene (*ASXL1*) frequently exhibit genome-wide H3K27me3 depletion, caused by the interaction of this gene with Enhancer of Zeste Homolog 2 (*EZH2*), a histone-lysine N-methyltransferase, and member of the PRC2 complex. This interaction will cause overexpression of *HOX* genes, which has been demonstrated for *HOXA9* and *HOXA10*. In these cases, it increases cell proliferation promoting leukaemia. In non-small cell lung cancer, the *HOXB7* promoter containing H3K27m3 hallmark, interacts with EZH2, regulating the *AKT* and *MAPK* pathways [141]. Nonetheless, this pattern is not the only potential histone biomarker related to *HOXB7*'s promoter, as H3K4me3 and H3K36me3 of the promoter regions of the *HOXB7*, *HOXC10,* and *HOXD8* may also present key targets for diagnosis and therapy in oral squamous cell carcinoma [142].

Another curious finding related to histone modification is that coactivator CREB-binding protein (CBP), a protein containing a HAT domain, is able to interact with *HOXB7*, and modulate its expression through histone acetylation [143]. This interaction seems to occur between the CBP HAT domain and the *HOXB7* N-terminal domain, a domain that, as previously mentioned, is less conserved and responsible for *HOX* genes' binding to their cofactors [143]. Additionally, the treatment of MDA-MB-231 cells, a breast cancer line representative of the "triple negative" subtype, with Trichostatin A, a deacetylase inhibitor, resulted in enhanced *HOXB7* transcription. Both results suggest that acetylation might partake in *HOXB7* regulation [143].

Interestingly, recent studies have also shown that *HOX* genes can indirectly regulate the expression of oncogenes or tumour-suppressor genes through alterations in DNA methylation and chromatin post-translational modifications. In fact, *DNMT3B* seems to be a putative target for both HOXB3 and HOXB7 in lung adenocarcinoma, and breast cancer, respectively [144, 145]. Thus, given the role that *HOX* genes exert as master transcriptional regulators, not only through the direct modulation of gene expression but also due to its role in epigenetic modulation, it is crucial to understand the implications of their dysregulation and assess their potential as prognostic and therapeutic targets. Thus, increasing efforts are being made to characterize the methylation profiles of relevant genes for the oncogenic process, including *HOX* genes, together

with the development of new compounds to selectively inhibit methylating agents, and potentially be used as a coadjutant therapy (Figure 9) [128, 129].

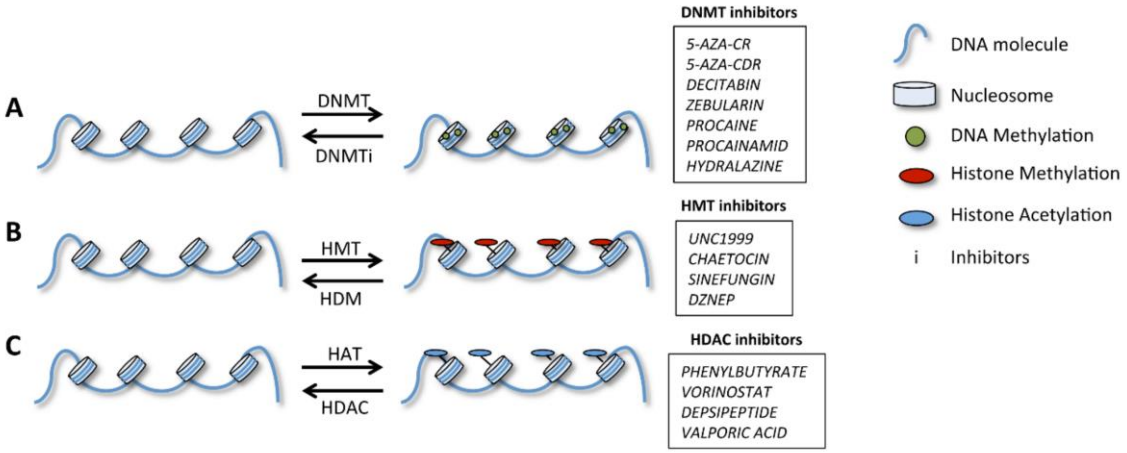

Figure 9- The influence of epigenetic modifications on gene expression and associated therapeutic agents. (A) DNA methylation is accompanied by transcriptional repression and is increased by DNA-methyltransferases (DNMT). (B) Histone methylation is facilitated by histone methyltransferases (HMT) and results in chromatin condensation, which may either inhibit or stimulate transcription**.** (C) Histone acetylation, which is promoted by histone acetyltransferases (HAT) enzymes, is linked with the opening of the chromatin mass and the initiation of transcription, whereas deacetylation, which is promoted by histone deacetylases (HDAC), is the inverse. Adapted from [122].

Two novel compounds, BIX-01294 [146] and GSKJ4 [147] have already shown promising results in pre-clinical trials, by increasing cell death and inhibiting BrCa tumour progression [15-17]. Nonetheless, *HOX* genes' potential as therapeutic targets goes beyond their methylation status. Through the knowledge of its function during development, scientists were able to identify PBX as a HOX cofactor, and design a peptide, HXR9 [148], that impairs their dimerization and consequent binding to DNA, leading to a decrease in tumour size [74, 80, 82, 83]. Hereupon, it is inferable that deeper insights into these genes' function and regulation might potentiate the discovery of new therapeutic targets for breast cancer treatment.

2. AIMS

This project aims to analyse and further understand the molecular machinery behind *HOX* regulation in cancer, and the potential role of novel chemical compounds as therapeutic agents against breast cancer, ultimately contributing to the discovery of new therapeutic targets for BrCa treatment, identifying molecular and/or epigenetic variations that are the source or consequence of *HOXB* dysregulation.
The specific aims of this project were:

AIM 1: Characterize potential variations in *HOXB7-9* methylation profiles, firstly by subjecting cell lines representative of BrCa subtypes to the DNA methylation inhibitor 5-aza-2'-deoxycytidine (ZdCyd) and analysing differences in these genes' expression profiles, and subsequently by analysing *HOXB7* promoter methylation.

AIM 2: Explore variances in gene expression of *HOXB7* downstream targets *CDH1* and *CTNNB1* upon treatment with ZdCyd. This is a part of an ongoing project at the host lab that aims to gain further inside on the epigenetic regulation of *HOXB7* putative targets in BrCa, and infer the role of this gene in these mechanisms.

AIM 3: Explore the role of acetylation in the regulation of *HOXB7* in BrCa, with inhibitor of histone deacetylase Trichostatin A. This will allow to further understand what epigenetic mechanisms might be causing *HOXB7* dysregulation in BrCa.

AIM 4: Study the therapeutic potential of N-1,2,3-Triazole-Isatin Derivatives as acetylation regulators. This aim will allow to understand whether these compounds are effective against breast cancer cells, particularly through the targeting of epigenetic and post-translational regulators.

AIM 5: Explore the effects of p-300 and APT-2 inhibition in their respective gene's expression. We aim to infer whether the expression of EP300 and LYPLA2 changes upon inhibition of their proteins, and if any feedback mechanisms will be triggered upon treatment.

AIM 6: Analyse variations in *HOXB7* expression profiles upon treatment with these compounds. Variations on *HOXB7* expression profiles when treated with N-1,2,3-Triazole-Isatin Derivatives might support the evidence gathered on AIM 3, and further infer the role of acetylation in *HOXB7* dysregulation in BrCa.

3. MATERIALS AND METHODS

3.1. Cell lines and culture

A total of five cell lines were used in the experiments, as representatives of the BrCa molecular subtypes: MCF-7, MDA-MB-231, MDA-MB-468, BT-474, and SKBR3 (Table 1). The non-malignant epithelial cell line MCF-10A was used as non-malignant control (Table 1). The BrCa cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Gibco™) supplemented with 10% foetal bovine serum (FBS, Biowest, South American Origin) and 1% antibiotic solution penicillin-streptomycin (pen-strep, Gibco™), while MCF-10A was grown in DMEM supplemented with 5% of horse serum, human insulin (40µg/mL), human EGF (20ng/mL), cholera toxin (100ng/mL), hydrocortisone (500ng/mL), and 1% antibiotic solution penicillin-streptomycin (pen-strep, Gibco™). All cell lines were incubated at 37 °C under a 5% CO_2 atmosphere, and used in experiments once they reached a confluence of 70-80%.

Table 1- Molecular Classification and Immunoprofile of the cell lines used in the course of this dissertation. Positive (+) and Negative (-) symbols stand for amplified and non-amplified receptors, respectively. Based on [149-152].

Cell line	Immunoprofile	Classification
MCF-10A	ER-, PR-	Normal Epithelial
MCF-7	ER+, PR+, HER2-	Luminal A
BT-474	ER+, PR+, HER2+	Luminal B
MDA-MB-231	ER-, PR-, HER2-	Basal-Like (Claudin Low)
MDA-MB-468	ER-, PR-, HER2-	Basal-Like
SKBR3	ER-, PR-, HER2+	HER2+

3.2. Treatment with DNA methylation inhibitor 5-aza-2′-deoxycytidine (ZdCyd)

Inhibition of DNA methylation was performed using 5-aza-2′-deoxycytidine (ZdCyd). Stock solutions were made in DMSO and diluted, prior to use, in the concentrations 1µM and 5µM, in culture medium. DMSO was used as vehicle control in the concentration of 5µM. Cells were plated into a T25 flasks until reaching a confluence of 70%. Then, they were trypsinized, and cultured into 12-well plates containing 700µL of culture medium. Posteriorly, the cells were grown for 24h

at 37 °C under a 5% CO_2 atmosphere, before being exposed to their respective treatment for 48h, following four conditions: Medium only, Vehicle Control, ZdCyd 1μM, and ZdCyd 5μM. Fresh treatments were added every day. This experiment was conducted for all the cell lines mentioned in Table 1, and all experiments were performed in triplicate.

3.3. Total RNA extraction from cells

To extract the total RNA from cells, two methods were used: **1)** NZY Total RNA Isolation kit (NZYtech) or **2)** TRIzol® Reagent (Invitrogen), following the manufacturer's instructions.

To conduct the TRIzol® Reagent extraction, the treatment was removed and 400μL of Trizol was added to each well, to dissociate nucleoprotein complexes. Samples were incubated at room temperature (RT) for 5 minutes, before adding 80μL of chloroform to each sample. Following that step, the tubes were agitated, incubated for 3 minutes at RT, and centrifuged at 12000 x g for 15 minutes at 4ºC. Afterward, the aqueous phase (containing the eluted RNA) was collected and the phenolic phase (containing DNA and other cellular compounds) was discarded. Isopropanol was added in equal portions to the aqueous phase and incubated for 10 minutes at RT. Samples were then centrifuged at 12000 x g for 15 minutes, at 4ºC. Posteriorly, the supernatant was discarded, and the pellet was resuspended in 400μL of 75% ethanol. The samples were centrifuged at 10000x g for 5 minutes at 4ºC. Posteriorly, the supernatant was discarded, and the pellet resuspended in 400μL of 75% ethanol, before repeating the centrifugation step. The supernatant was discarded, and the pellet resuspended in 150μL of 75% ethanol and followed by another centrifugation. Afterward, the ethanol was removed, and the tubes were left to dry, at RT, for 15 minutes with their lids open. The pellet was resuspended in 25μL of H_2O DNAse/RNAse free, and incubated at 60ºC for 10 minutes, being homogenized every 5 minutes.

To remove contaminant DNA, the extracted RNA was treated with 1μL of Recombinant DNase I (Roche), 4μL of 10x DNase I Recombinant RNA-free Incubation Buffer 10x (Roche) and 10μL of H2O DNAse/RNAse free. Samples were incubated at 37ºC for 1h and then placed on ice for two minutes.

Following the above-mentioned proceeding, RNA precipitation was conducted. A total of 4μL of sodium acetate 2M pH4.6 were added to the sample to reach a 0.2M concentration, followed by 100μL of absolute ethanol, and the samples were incubated overnight at -20ºC to precipitate. On

the following day, samples were centrifuged at 10000x g for 30 minutes at 4ºC, the supernatant was discarded, and 150µL of 75% ethanol was added. The centrifugation step was repeated and another wash with 150µL of 75% ethanol was performed. Lastly, the supernatant was discarded and the pellets were air-dried for 15 minutes, before being resuspended in 25µL of H2O DNAse/RNAse free.

3.4. cDNA synthesis

As the appropriate template for Real-Time quantitative PCR is cDNA, the extracted RNA was subjected to reverse transcription, using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems), according to the protocol displayed in Table 2. Afterward, the samples were placed in a Thermocycler (T100, Bio-Rad) with the following program: 10 minutes at 25ºC, 120 minutes at 37ºC, 5 minutes at 85ºC, and 4ºC until the samples were removed. A total of 800ng of RNA from ZdCyd treated samples, and 500ng of RNA from the N-1,2,3-Triazole-Isatin Derivatives treated samples were converted for the expression analysis.

Table 2-Reagents used for cDNA synthesis using the High-Capacity cDNA Reverse Transcription Kit.

Components	Volume (µL)
RNA + H2O	10
Buffer 10x	2
25x dNTPs	0,8
10x Random Primers	2
MultiScribe 20x	1
RiboLock	0,5
H2O	3,7
Total	20

PCRs were performed routinely to validate cDNA synthesis, using primers for the *GAPDH* region (Table 3). DNAse/RNAse free H2O was used as a negative control. The amplification protocol was carried out on a Thermocycler (T100, Bio-Rad), as follows: 95ºC for 15 minutes, followed by 28 cycles of 94ºC for 30 seconds (s), 60ºC for 90s and 72ºC for 90s, and a final elongation step of 72ºC for 10 minutes. The samples were stored at 4ºC until usage. Amplification was confirmed by loading the previously amplified samples on a 2% agarose gel.

The 2% agarose gel was prepared by dissolving 1g of electrophoresis grade agarose (NZYtech) in 50 mL of Tris-Acetate-EDTA (TAE) 1x. Once dissolved, the mix was cooled down and 2µL of Green Safe Premium (NZYtech) was added, and the mixture was gently poured into a mold to solidify. As the master mix of the validatory PCR already contained loading dye, the samples were ready to be loaded into the gel. Once the samples and a 1 kilobase (Kb) ladder (Solis Biodyne) were loaded, the gel was run in 1X TAE buffer at 100 V. Electrophoresed products were revealed using ultraviolet spectroscopy (Molecular Imager ® Gel Doc ™ XR+ Imaging System, Bio-Rad). A band at around 150bp is expected to confirm that cDNA synthesis was successful.

Table 3- Reagents used in validatory PCR to validate cDNA synthesis. FW- Forward. RV- Reverse (*NZYTaq II 2x Green Master Mix, NZYtech).

Components	Volume (µL)
cDNA	0.5
Primer FW	0.5
Primer RV	0.5
Master Mix*	5
H2O	3.5

3.5. Reverse Transcription Quantitative PCR

Gene expression analyses were performed using the reverse transcription quantitative PCR (RT-qPCR) technique. Samples were subjected to RT-qPCR amplification, using the iTaq™ Universal SYBR® Green Supermix (Bio-Rad), performed on the real-time cycler CFX96™ (Bio-Rad), according to the following amplification protocol: 95ºC for 3 minutes, and 40 cycles of 95ºC for 10s and 60ºC for 30s. Each reaction was carried out using 2µL of cDNA diluted 1:4, 0.5µL of the respective primer sequence (Forward and Reverse) (

Table 4); 2µL of DNAse/RNAse free treated H_2O and 5µL of 2x SYBR Green, comprising a total of 10µL per reaction, in each well of a 96-well PCR plate. All experiments were performed in triplicate, and *GAPDH* gene amplification was used as a reference for normalization.

Table 4- Sequence and product length of the primers used in RT-qPCR analyses. [S]- self-made using Primer3 Online Software. Bp- base pair. FW - forward primer. RV- reverse primer.

		Sequence (5'-3')	Product bp	Origin
HOXB7	FW	TACCCCTGGATGCGAAGCTC	171	[153]
	RV	AATCTTGATCTGTCTTTCCGTGA		
HOXB8	FW	GACAGGTCAAAATCTGGTTCC	111	[S]
	RV	GCTTCTGTTTCTCCAGCTCCT		
HOXB9	FW	CTACGGTCCCTGGTGAGGTA	198	[121]
	RV	TAATCAAAGACCCGGCTACG		
EP300	FW	GACCCTCAGCTTTTAGGAATCC	353	[154]
	RV	TGCCGTAGCAACACAGTGT		
LYPLA2	FW	GTAACACCATGTCTGTGCCC	222	[S]
	RV	GTCAAACCAGGAGGGCAT		
GAPDH	FW	ACTGGCGTCTTCACCACCAT	142	[81]
	RV	TCTTGAGGCTGTTGTCATACTTC		
CDH1	FW	CGAGAGCTACACGTTCACGG	119	[155]
	RV	GGGTGTCGAGGGAAAAATAGG		
CTNNB1	FW	GGGTCCTCTGTGAACTTGCT	165	[S]
	RV	TGTAATCTTGTGGCTTGTCCTC		

The CFX Manager 3.1 software (Bio-Rad) was employed to visualize amplification and melting curves, define the cycle amplification threshold (CT), and assign each sample and negative control to its plate location. This data was exported as an excel file and analysed using the formula "Ratio= E target $^{-\text{(CT sample target gene)}}$ / E *GAPDH*$^{-\text{(CT sample } GAPDH\text{)}}$", where E corresponds to primer efficiency.

3.6. Bisulphite conversion of DNA

DNA extracted from cell lines MCF-10A, MDA-MB-231 and SKBR3 was subjected to bisulphite conversion, using the EZ DNA Methylation-Direct™ Kit (Zymo Research). Each sample contained 350ng of DNA, diluted in up to a total of 20µL of DNAse/RNAse free treated H_2O. Samples were mixed with 130µL of CT conversion reagent, and placed in a thermocycler (T100, Bio-Rad), according to the following steps: 98ºC for 8 minutes, 64ºC for 3.5 hours, 4ºC storage for up to 20 hours. To this reaction were added 600µL of M-binding Buffer, and then loaded onto a Zymo-Spin™ IC Column. The column was centrifuged at 10000x g for 30s, and the flowthrough was discarded. Afterward, 100µL of M-Wash Buffer was added to the column, followed by another centrifuging. Subsequently, 200µL of M-Desulphonation Buffer was loaded onto the column, incubated for 20 minutes at RT, and then centrifuged at full speed for 30s. The following step, consisting of the addition of 200µL of M-Wash buffer followed by a 30s centrifuging at full speed, was repeated twice. Lastly, the DNA was eluted in 10µL of M-Elution Buffer, added directly onto the column, and then centrifuged at full speed for 30s. This DNA was then stored at -20ºC for short-term storage.

3.7. Gradient PCR

In order to amplify the selected region of *HOXB7*'s promoter, a gradient PCR was performed using the Multiplex PCR Kit (Qiagen), following the protocol displayed in Table 5 and Table 6.

Table 5- Components used in gradient PCR reaction. FW - forward primer. RV- reverse primer. (*QIAGEN Multiplex PCR Master Mix).

Components	Volume (µL)
Master Mix*	10
Primer F*W*	2
Primer R*V*	2
Q-Solution	2
Bisulphite treated DNA	1
H_2O	3

Table 6- Thermocycler Protocol for Gradient PCR.

Step	Temperature	Time	
1	95ºC	15 min	
2	94ºC	30 sec	3 cycles
	64ºC	90 sec	
	72ºC	90 sec	
3	94ºC	30 sec	3 cycles
	62ºC	90 sec	
	72ºC	90 sec	
4	94ºC	30 sec	3 cycles
	60ºC	90 sec	
	72ºC	90 sec	
5	94ºC	30 sec	3 cycles
	58ºC	90 sec	
	72ºC	90 sec	
6	94ºC	30 sec	35 cycles
	56ºC	90 sec	
	72ºC	90 sec	
7	72ºC	10 min	
8	4ºC	Pause	

The primers used for promotor amplification were designed to flank a region of *HOXB7*'s promoter that contained a CpG island, and assuming that the DNA had been broadly converted (all the cytosines on the dominant strand had been converted to thymine). Therefore, the following primers were generated, to flank a region of 452 bp of *HOXB7*'s promoter: FW-TGTTTTGGGAGAATTGGATTAA; RV- CCCTTTTTCCTCTCCAAAC.

Afterwards, amplification was verified using a 2% agarose gel. If more than one band was present, the band would be cut and purified using the PureLink™ Quick Gel Extraction and PCR Purification Combo Kit (Invitrogen). Otherwise, the PCR product would be directly purified using the DNA Clean & Concentrator™-5 (Zymo Research), following the manufacturer's instructions.

3.8. Cloning into pGEM®-T Easy Vector

To clone the amplified portion of the *HOXB7* promoter into the pGEM®-T Easy Vector (Promega), a ligation reaction was set up as displayed in Table 7. All the reagents were mixed and the reaction was incubated overnight at 4ºC, for a maximum number of transformants.

Table 7- Components used in the ligation reaction for vector cloning.

Reagents	Volume (µl)	Positive Control (µl)	Background Control (µl)
2x Rapid Ligation Buffer	5	5	5
pGEM®-T Easy Vector	1	1	1
PCR Product	0.5	-	-
Control Insert DNA	-	2	-
T4 DNA Ligase	1	1	1
Deionized Water	2.5	1	3
Total	10	10	10

3.9. Competent Bacteria DH5α Transformation

In order to transform the bacteria with the pGEM®-T Easy Vector, DH5α Competent Cells were thawed in an ice bath for 5 minutes, and gently mixed. Then, 50µl of cells were added to the ligation reaction tubes containing the cloned pGEM®-T Easy Vector, while 100µl of cells were added to a tube containing 2.5µl of control uncut plasmid (pUC19). Both conditions were incubated on ice for 20 minutes. Following incubation, a heat-shock reaction was carried on for 45–50s in a water bath at exactly 42°C and immediately returned to ice for 2 minutes. This step is crucial to ensure the penetration of the plasmid inside the bacteria. Subsequently, the bacteria recovered on 950µl of LB medium, for 1.5 hours at 37°C with shaking (~150rpm). Upon this step, 100µl of each transformation culture was plated onto duplicate plates containing LB agar/ampicillin/IPTG/ X-Gal, and incubated overnight at 37ºC. Bacterial manipulation was performed under a flame, to ensure a sterile environment.

3.10. Blue-white screening and bacterial selection

Blue-white screening is a technique that allows the identification of successfully recombinant bacteria. The presence of IPTG, a non-metabolizable galactose analogue, in the medium will induce the lacZ operon, and lead the bacteria to produce β-galactosidase, an enzyme that metabolizes lactose, in a process called α-complementation. In the presence of this enzyme, X-gal hydrolyses to form 5-bromo-4-chloro-indoxyl, which dimerizes into an insoluble blue pigment called 5,5'-dibromo-4,4'-dichloro-indigo, thus giving the non-recombinant colonies a blue colour. The cloning vector contains a multiple cloning site (MCS) within the lacZ sequence that becomes interrupted once foreign DNA is cloned into that site, therefore not producing a functional β-galactosidase, and resulting in the recombinant colonies remaining white.

Given that, in order to select only recombinant colonies, the colonies that presented the colour white were picked using a pipet tip that was disposed into a 15mL falcon tube containing 3mL of LB medium, and incubated overnight at 37ºC with agitation.

3.11. Miniprep and colony PCR

Upon incubation, 2mL of bacteria in LB-culture were harvested and sedimented for 5 minutes at 14000 rpm, to proceed to the NZYMiniprep Kit (NZYtech). The supernatant was discarded and the pellet resuspended in 250µL of Buffer A1 by vigorous vortexing. Then, 250µL of Buffer A2 was added and the mixture was incubated at RT for less than 4 minutes, before adding 300µL of Buffer A3. The reaction was mixed and centrifuged for 10 minutes at RT, then loaded into an NZYTech spin column. The column was then centrifuged for 1 minute at 11000x g and the flow-through was discarded. The following step required the addition of 500µL of Buffer AY, followed by a 1-minute centrifugation at 11000x g, the addition of 600µL of Buffer A4, and the iteration of the centrifugation step. The flow-through was discarded and the empty spin column was subjected to a 2-minute centrifugation, in order to dry the silica membrane before DNA elution. Purified DNA was eluted in 30µL of Buffer AE, incubated for 1 minute at RT, and centrifuged for 1 minute at 11.000x g, before being stored at -20ºC.

As to confirm the presence of the desired DNA inside the cloning vector, purified DNA was subjected to a PCR using M13 primers that bind to the M13 primer-binding site of the plasmid and flank the MCS, where our interest sequence is inserted. This step ensures that our target sequence is present and that it was cloned into the MCS. The following sequences correspond to the sequences of primers M13 forward and reverse: FW-CGCCACCCTTTTCCCAGTCACGA; RV- TCACACAGGAAACAGCTATGAC.

PCR amplification was conducted using the same reagents displayed in Table 3, and the protocol was carried out on a Thermocycler (T100, Bio-Rad), as follows: 95ºC for 15 minutes, followed by 28 cycles of 94ºC for 30s, 60ºC for 90s, 72ºC for 90s, and a final elongation step of 72ºC for 10 minutes. DNAse/RNAse free H2O was used as negative control. Amplification was confirmed by loading the previously amplified samples on a 2% agarose gel, where a band of 763bp was expected, if the insert was present.

3.12. DNA sequencing and Methylation Status Analysis

DNA samples containing the desired insert were quantified and then diluted to a concentration of 100ng/µL. Afterward, 5µL of each diluted sample was separated, properly labelled, and sent to

the Genomics Group at i3S to be sequenced, using the Sanger Sequencing technique. The samples were sequenced using M13 primers.

Once the samples were sequenced, methylation status was evaluated using the BISMA-Bisulphite sequencing DNA Methylation Analysis program.

3.13. Treatment with inhibitor of histone deacetylase Trichostatin A (TSA)

Inhibition of histone deacetylation was performed using InSolution™ Trichostatin A (TSA, 4,6-Dimethyl-7-[p-dimethylaminophenyl]-7-oxahepta-2,4-dienohydroxamic Acid) from Streptomyces, sp. (Merk Millipore). Stock solutions were already diluted DMSO, to a concentration of 10mM, and diluted prior to use in the concentrations 400nM, in culture medium. DMSO was used as vehicle control in the concentration of 400nM. Cell lines MCF-10A, MDA-MB-231, MCF-7 and MDA-MB-468 were plated into a T25 flasks, until reaching a confluence of 70%. Then, they were trypsinized, and cultured into 12-well plates containing 700µL of culture medium. Posteriorly, the cells were grown for 24h at 37 °C under a 5% CO_2 atmosphere, before being exposed to 400nM of TSA or DMSO for 20h. All the experiments were performed in triplicate.

3.14. MTT Anti-proliferative Assay upon treatment with N-1,2,3-Triazole-Isatin Derivatives

Around 1×10^4 cells from each cell line were cultured on a 96-well plate, containing 200 µl per well of their respective culture medium and incubated at 37 °C under a 5% CO_2 atmosphere, for 24h. Stock solutions of N-1,2,3-Triazole-Isatin Derivatives (1a), (1b), (1c), (S)-(2b) and (R)-(2j) were made in DMSO. Then, the cells were treated with different concentrations of the tested drugs (ranging from 20 µM to 0 µM), dissolved in culture medium, for 72h. Thereupon, the drug treatments were removed and cells were incubated for 3h with 100 µl of 3-(4,5-dimethyltiazol-2-yl)-2,5-diphenyltetrazoliumbromide (MTT, Sigma Aldrich) dissolved in culture medium (500 µg/ml). Subsequently, 100 µl of solubilizing solution (10% (w/v) SDS, 0.01 M HCl) were added to each well and incubated for 18 h. Absorbance reading was performed at 590 nm using a microplate reader (Synergy™ Mx Monochromator-Based Multi-Mode Microplate Reader, BioTek Instruments, Inc.). All experiments were performed in triplicate. DMSO was used as vehicle

control. The IC_{50} values were calculated from MTT data using the GraphPad Prism Software Inc. (version 6.01).

3.15. Treatment for gene expression analysis on MCF-7, MDA-MB-231 and MCF-10A

To assess whether gene expression of the target genes was altered upon treatment with N-1,2,3-Triazole-Isatin Derivatives, cell lines MCF7, MDA-MB-231 and MCF-10A were grown on a 12-well plate containing their respective culture medium, and incubated at 37 °C under a 5% CO_2 atmosphere for 24h. Afterward, each cell line was treated with each compound dissolved in medium, to the concentration of their respective IC_{50} value. Hereupon, cells were incubated for 72h, before initiating the RNA extraction protocol, using the NZY Total RNA Isolation Kit, followed by cDNA synthesis and qPCR, as previously described. All experiments were performed in triplicate. DMSO was used as vehicle control.

3.16. Statistical Analysis

In the experiments using cell lines treated with ZdCyd or TSA, statistical analysis was performed using the unpaired T-Test with Welch's correction, to compare the expression of treated conditions to DMSO control. For the experiments considering N-1,2,3-Triazole-Isatin Derivatives, statistical analysis was carried out through One-Way ANOVA, using Brown-Forsythe and Welch ANOVA test with Dunnett T3 correction. Both analyses were achieved using GraphPad Prism Software Inc. (version 6.01). Differences were considered statistically significant when $p<0.05$.

4. RESULTS

4.1. Regulation of *HOXB7* expression through methylation

4.1.1. Variations on *HOXB7* expression upon 5-aza-2′-deoxycytidine treatment

To investigate whether *HOXB7* dysregulation in BrCa is associated with aberrant methylation profiles, BrCa cells representing different molecular subtypes were subjected to DNA methylation inhibitor 5-aza-2′-deoxycytidine (ZdCyd), and subsequently, the *HOXB7* expression levels were quantified (Figure 10). At low concentrations, ZdCyd is converted into deoxyribonucleoside triphosphates that are incorporated into the cytosine residues of freshly synthesized DNA (Cyt) [156, 157]. DNMT, or (cytosine-C5) methyltransferase, irreversibly binds to the Cyt residues, which results in the rapid loss of DNA methylation, and global hypomethylation of newly synthesized DNA [156, 157].

The results showed a significant increase of *HOXB7* expression in all cell lines treated with ZdCyd, compared to the control (DMSO), except for MCF-7 (Figure 10). Thus, in all molecular subtypes the *HOXB7* endogenous overexpression, reported by de Bessa Garcia and colleagues [81] (Figure 11) seems to be affected by DNA methylation processes (Figure 10). In MCF-10A cells, which have low to inexistent expression of *HOXB7*, a very significant increase was detected after treatments with ZdCyd concentrations of 1µM ($p<0.0001$) and 5µM ($p=0.0030$) (Figure 10A).

For SKBR3 cells, which are among the BrCa cells with lower levels of *HOXB7* overexpression [81], the 1µM ZdCyd treatment was sufficient to cause a significant increase in *HOXB7* expression ($p=0.0072$) (Figure 10B). Then, in cells representative of the triple-negative BrCa molecular subtype, MDA-MB-231 (Figure 10C) and MDA-MB-468 (Figure 10D), which are considered to have moderate to high levels of *HOXB7* overexpression [81], both treatments with ZdCyd were able to cause a significant increase of *HOXB7* expression. However, the increase of *HOXB7* expression in MDA-MB-231 was much more accentuated than the one observed in MDA-MB-468, which expresses higher endogenous levels of this gene than MDA-MB-231. These results suggest a higher grade of methylation influence in MDA-MB-231 cells than MDA-MB-468, in spite of the fact that both belong to the triple-negative subtype, that ultimately results in the lower levels of endogenous *HOXB7* expression in this cell line.

In BT-474 cells (Figure 10F), which were considered to have the highest levels of *HOXB7* overexpression [81], a subtle increase of *HOXB7* expression was only detectable after treatment with 5µM ZdCyd (p=0.0233).

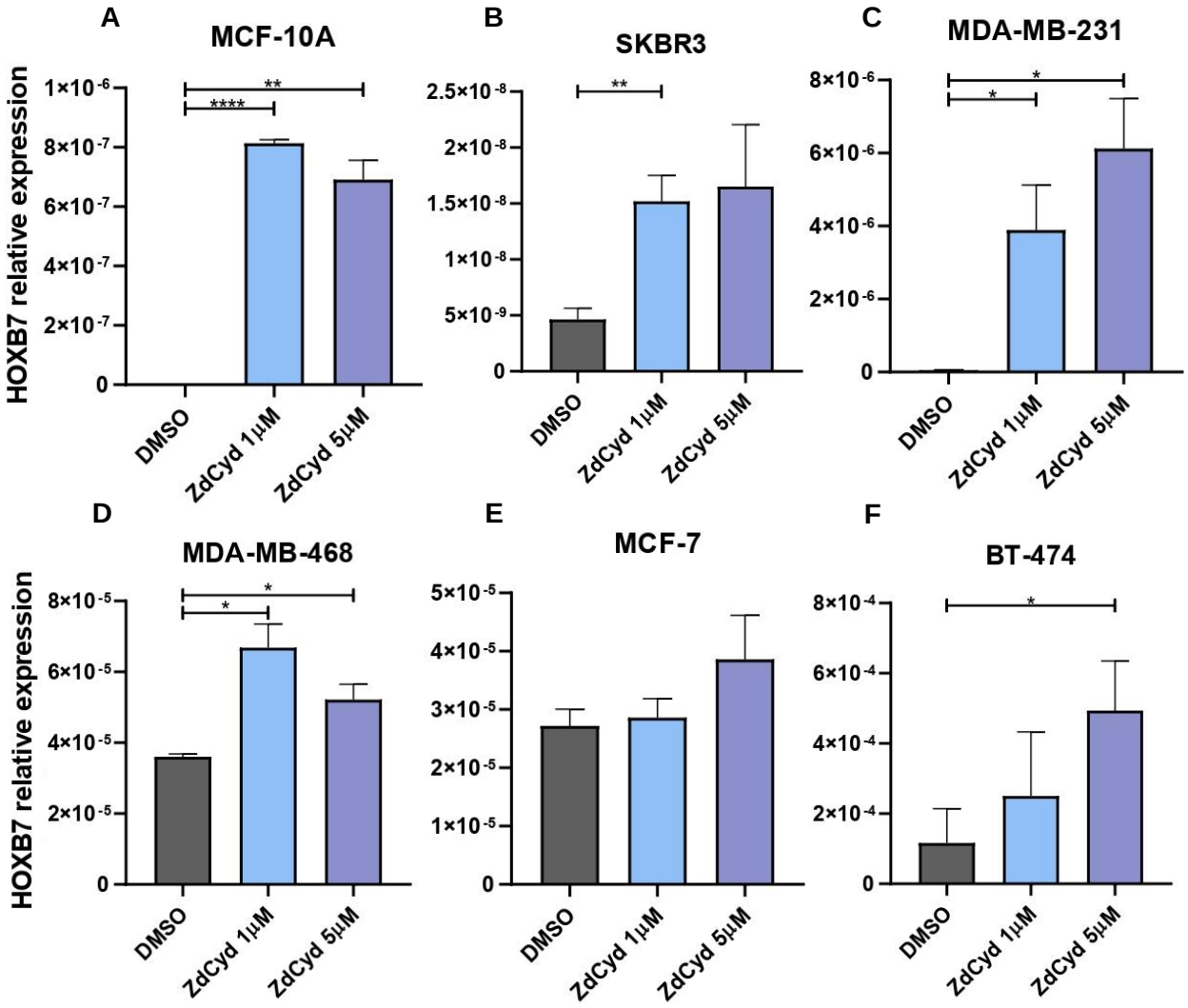

Figure 10- *HOXB7* relative expression upon ZdCyd treatment across different cell lines, analysed by RT-qPCR. **A.** *HOXB7* relative expression in MCF-10A upon ZdCyd treatment at 1µM (p<0.0001) and 5µM (p=0.0030). **B.** *HOXB7* relative expression in SKBR3 upon ZdCyd treatment at 1µM (p=0.0072) and 5µM (p=0.0615). **C.** *HOXB7* relative expression in MDA-MB-231 upon ZdCyd treatment at 1µM (p=0.0325) and 5µM (p=0.0166). **D.** *HOXB7* relative expression in MDA-MB-468 upon ZdCyd treatment at 1µM (p=0.0145) and 5µM (p=0.0210). **E.** *HOXB7* relative expression in MCF-7 upon ZdCyd treatment at 1µM (p=0.6052) and 5µM (p=0.1067). **F.** *HOXB7* relative expression in BT-474 upon ZdCyd treatment at 1µM (p=0.3439) and 5µM (p=0.0233). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, p ≤ 0.05, **, p ≤ 0.01, ****, p ≤ 0.0001). All experiments were performed in triplicate, using independent biological replicates.

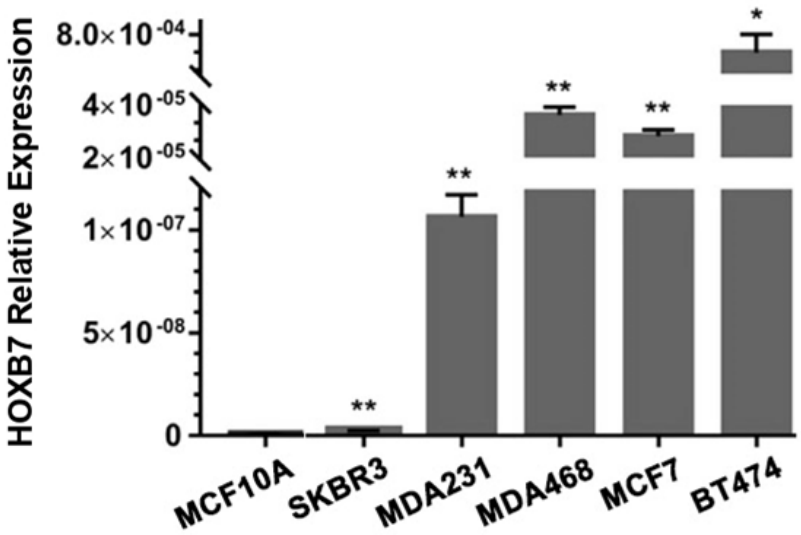

Figure 11- Endogenous *HOXB7* mRNA expression levels analysed by qPCR in breast cancer cell lines SKBR3, MDA-MB-231, MDA-MB-468, MCF-7 and BT-474, compared with normal breast cells MCF10-A. MCF-10A represented as MCF10A, MDA-MB-231 represented as MDA231, MDA-MB-468 represented as MDA468, MCF-7 represented as MCF7, and BT-474 represented as BT474. Blank spaces in each column represent a break in the y scale, to facilitate reading. Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, $p \leq 0.05$, **, $p \leq 0.01$). All experiments were performed in triplicate, using independent biological replicates. Adapted from [81].

Thus, in general, it seems that increment of *HOXB7* expression after ZdCyd treatments is lower in cells that typically express higher levels of *HOXB7* (e.g., BT-474) than in cells with lower *HOXB7* endogenous expression levels (MCF-10A).

4.1.2. Methylation profile of the *HOXB7* promoter region

To further assess the role of methylation in *HOXB7*'s dysregulation in BrCa, and validate the results obtained with the ZdCyd treatments, the characterization of the *HOXB7* promoter methylation status was performed in BrCa cells (MDA-MB-231 and SKBR3) and control (MCF-10A) using bisulphite conversion followed by sequencing analyses.

The *HOXB7* promoter region has been established and corresponds to a region of 1901bp (GenBank accession number AJ414528). From the described region, a sequence of 451bp was selected between positions 1346 and 1796, which contained a total of 29 CpG sites. To confirm the possible existence of a CpG island within this sequence, the whole promoter sequence was analysed using EMBOSS CpGPlot program (ratio > 0.6, percent (C+G) > 50% and length > 200 bp). The results suggested the existence of a potential 386bp CpG island between nucleotides

1425 and 1810 of the promoter sequence, which was mostly comprised inside the selected sequence (Figure 12).

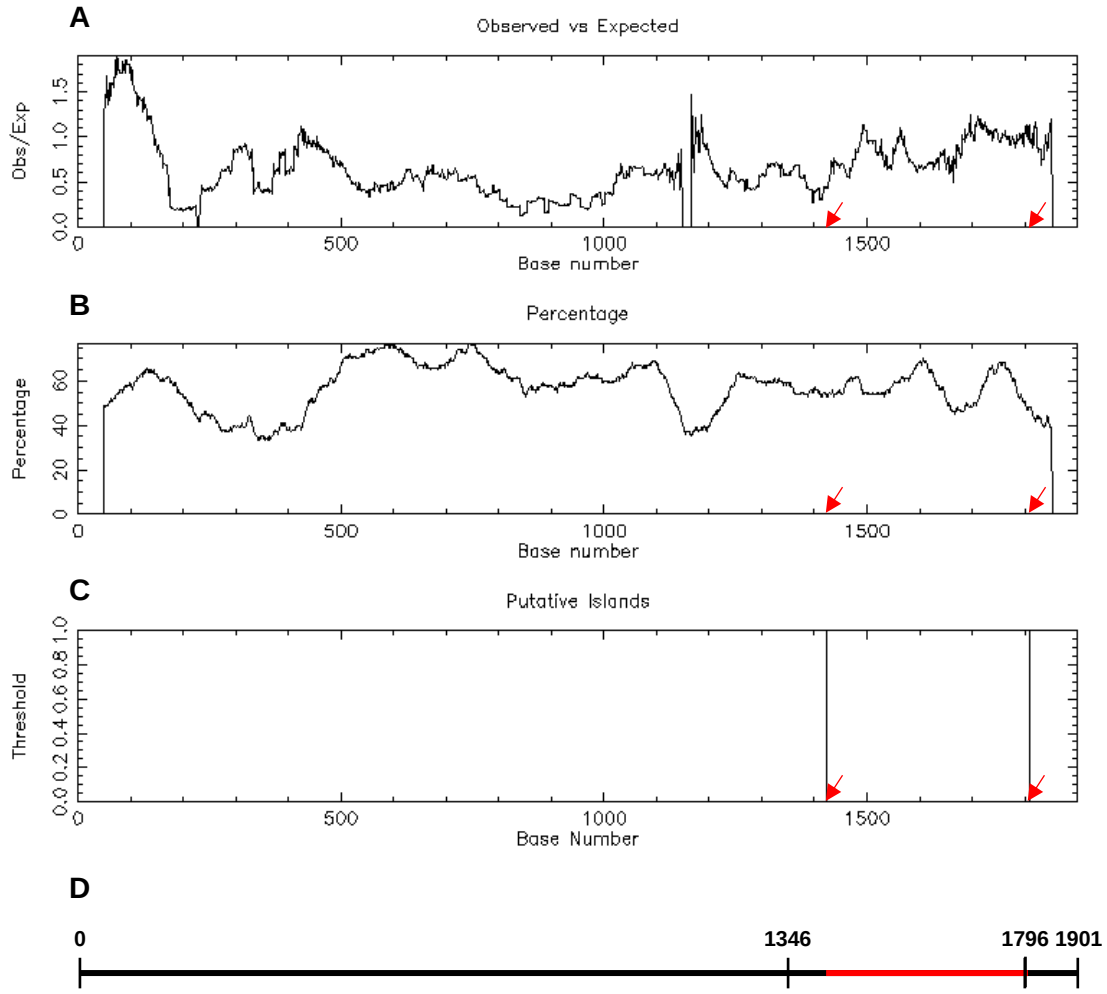

Figure 12- Predicted CpG islands within the promoter sequence of *HOXB7*. NCBI Reference Sequence AJ414528 for *HOXB7* promoter was submitted fully to the EMBOSS CpGPlot website http://www.ebi.ac.uk/emboss/cpgplot website. The selected parameters to define a CpG island were set as follow: sequence of ≥ 200 bp with a percentage of C + G content ≥ 50% (**B**) and threshold of CpG observed/CpG expected ≥ 0.6 (**A**). Following these parameters, the program was able to identify one CpG island between nucleotides 1425 and 1810 (**C**), which are identified by the red arrows. **D.** Schematic representation of HOXB7 promoter sequence. Predicted CpG island in red. Selected region for sequence amplification between positions 1346 and 1796.

For the cell line MCF-10A, 10 clones were analysed and, out of the 289 CpG sites evaluated, 50,9% were methylated, ranging from a methylation status of 62.1% to 41.4% (Figure 13). The analysis revealed 14 CpG residues preferentially methylated, being that CpG sites identified as 4-8, 10, 13, 14, 16, 19, and 20 appear methylated in all of the analysed sequences. The sites 22-24 appeared methylated in 9 of the 10 sequences analysed.

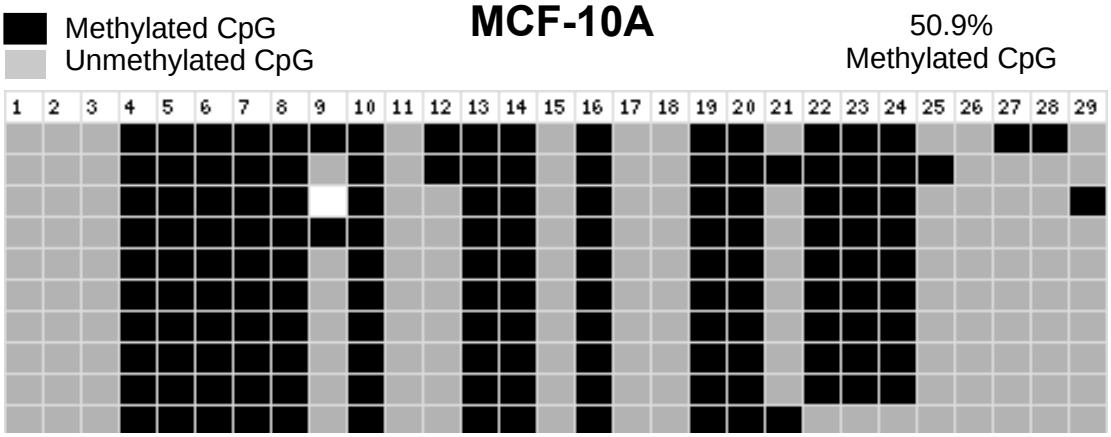

Figure 13- Methylation profile of the selected *HOXB7* promoter region in MCF-10A cell line. Bisulphite sequencing analysis revealed a methylation status of 50.9%, out of 289 CpG sites analysed from 10 independently sequenced clones. Each square represents one CpG site. Methylated CpG sites are represented by a black square, while unmethylated ones are represented by a grey square, and non-analysed CpG islands are identified by a white square. Each row represents a single clone.

In the case of MDA-MB-231 cells, a total of 8 sequences were analysed. From a total of 232 CpG sites evaluated, 51.6% appeared methylated, ranging from a methylation status of 62.1% to 48.3% (Figure 14). Similarly to MCF-10A cells, it is possible to pinpoint 14 CpG residues preferentially methylated, being that CpG sites identified as 4-8, 10, 13, 14, 16, 19, and 22-24 appear methylated in all of the analysed sequences. Furthermore, site 20 appeared methylated on six out of the eight analysed sequences, being that the two sequences that did not present methylation at this site showed a constant methylation pattern at the sites between 21 to 28 that did not appear methylated at the samples that presented methylation at CpG site 20. Thus, we were able to identify two distinct patterns of methylation: the first pattern is characterized by methylation in the CpG residues 21-28 (first two clones) and the second pattern is characterized by methylation in the CpG residues 20 and then 22-24.

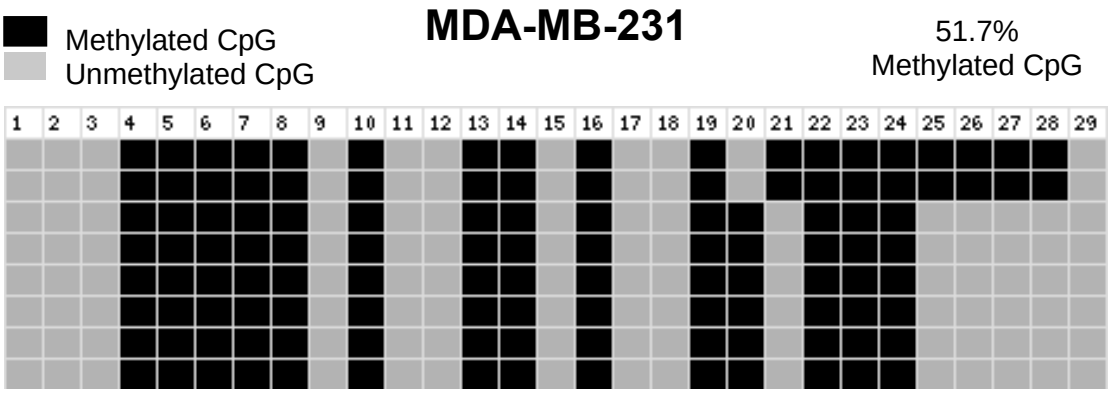

Figure 14- Methylation profile of the selected *HOXB7* promoter region in MDA-MB-231 cell line. Bisulphite sequencing analysis revealed a methylation status of 51.7%, out of 232 CpG sites analysed. Each square represents one CpG site. Methylated CpG sites are represented by a black square, while unmethylated ones are represented by a grey square, and non-analysed CpG islands are identified by a white square. Each row represents a single clone.

To assess the methylation pattern of SKBR3 cells, a total of 5 sequences were analysed. From a total of 170 CpG sites evaluated, 50.0% appeared methylated, in a very homogenous manner across all clones (Figure 15). Similarly to what we observed in the previously analysed cell lines, it is possible to pinpoint 14 CpG residues preferential methylated, being that CpG sites identified as 4-7, 10, 13, 14, 16, 19, and 22-24 appear methylated in all of the analysed sequences. However, the software did not detect a CpG at site 9. Additional analyses revealed a point mutation at position 173 bp, where a thymine (T) in the reference sequence of the promoter was replaced by an adenine (A), and therefore, being no longer considered a CpG site (Figure 16).

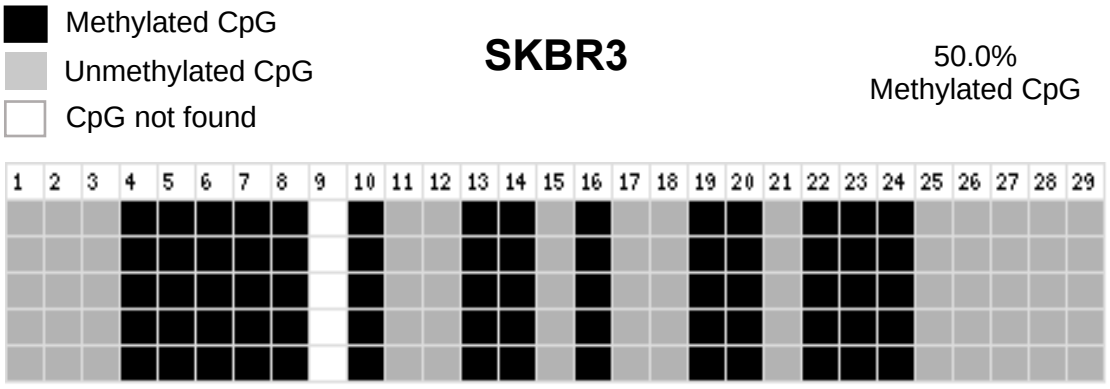

Figure 15- Methylation profile of the selected *HOXB7* promoter region in SKBR3 cell line. Bisulphite sequencing analysis revealed a methylation status of 50.0%, out of 140 CpG site analysed. Each square represents one CpG site. Methylated CpG sites are represented by a black square, while unmethylated ones are represented by a grey square, and non-analysed CpG sites are identified by a white square. Each row represents a single clone.

Species/Abbrv	*		*	*	*	*		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
1. Reference	A	T	G	G	A	T	T	G	T	T	T	T	G	A	G	G	T	T	A	T	T	T
2. SKBR3 clone 1	A	A	G	G	A	T	C	G	T	T	T	T	G	A	G	G	T	T	A	T	T	T
3. SKBR3 clone 2	A	A	G	G	A	T	C	G	T	T	T	T	G	A	G	G	T	T	A	T	T	T
4. SKBR3 clone 3	A	A	G	G	A	T	C	G	T	T	T	T	G	A	G	G	T	T	A	T	T	T
5. SKBR3 clone 4	A	A	G	G	A	T	C	G	T	T	T	T	G	A	G	G	T	T	A	T	T	T
6. SKBR3 clone 5	A	A	G	G	A	T	C	G	T	T	T	T	G	A	G	G	T	T	A	T	T	T

Figure 16- Aligned sequences of the analysed CpG island in the *HOXB7* promoter region. The five analysed clones were aligned and compared to a reference sequence of the promoter region, where all the cytosines had already been converted through bisulphite treatment. List of sequences are presented on the left, while nucleotides are presented on the right. To facilitate analysis, the represented sequence starts at the position corresponding to 172bp of the original sequence. Point mutation highlighted by red boundary. Adenine (A) is presented in green, Thymine (T) by dark pink, Guanine (G) by purple and Cytosine (C) by blue. (*) correspond to conserved nucleotides.

4.2. *HOXB8* and *HOXB9* expression upon ZdCyd treatment

To further explore the role of methylation on the HOXB cluster, BrCa cells MDA-MB-231 and MCF-7, together with normal breast epithelial cell line MCF-10A were treated with ZdCyd. Then, after 48h, the expression levels of *HOXB8* and *HOXB9* were quantified using RT-qPCR.

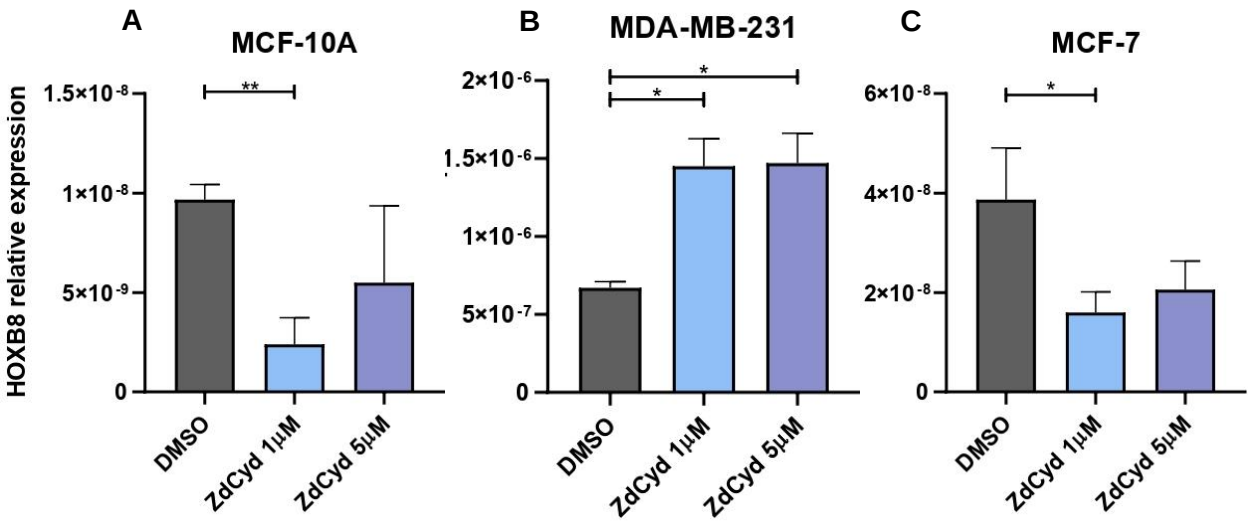

Figure 17- *HOXB8* relative expression upon ZdCyd treatment across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analysed by RT-qPCR. **A-C.** *HOXB8* relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon ZdCyd treatment at 1µM (p=0.0031 MCF-10A; p=0.0134 MDA-MB-231; p=0.0485 MCF-7) and 5µM (p=0.1987 MCF-10A; p=0.0158 MDA-MB-231; p=0.0747 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on analyses using Unpaired T-Test with Welch correction (*, $p \leq 0.05$, **, $p \leq 0.01$). All experiments were performed in triplicate, using independent biological replicates.

Unlike what was observed for *HOXB7*, the expression of *HOXB8* significantly decreased upon ZdCyd 1µM treatment, in cell lines MCF-10A (p=0.0031) and MCF-7 (p=0.0485), when compared to the DMSO control condition (Figure 17). Similarly, both cell lines showed an even more significant decrease in *HOXB9* expression, across both conditions, particularly at the concentration of 1µM (Figure 18). Although these results do not indicate that methylation directly regulates *HOXB8* and *HOXB9* expression in these cell lines, it suggests that it might regulate an upstream modulator, which then leads to their downregulation.

Furthermore, opposingly to what was observed for MCF-10A and MCF-7, the MDA-MB-231 cell line revealed an upregulation of *HOXB8* in both conditions (p=0.0134; p=0.0158) (Figure 17B). The same was observed for *HOXB9* after 5µM ZdCyd treatment (p=0.0198) (Figure 18B). The increase of *HOXB8* and *HOXB9* after inhibition of DNA methylation is suggestive that methylation directly partakes in the regulation of these genes in this particular cell line, similarly to what was reported for *HOXB7*. Thus, the role of methylation in the control of *HOXB7-HOXB9* expression might be relevant in particular cases of Triple-Negative BrCa.

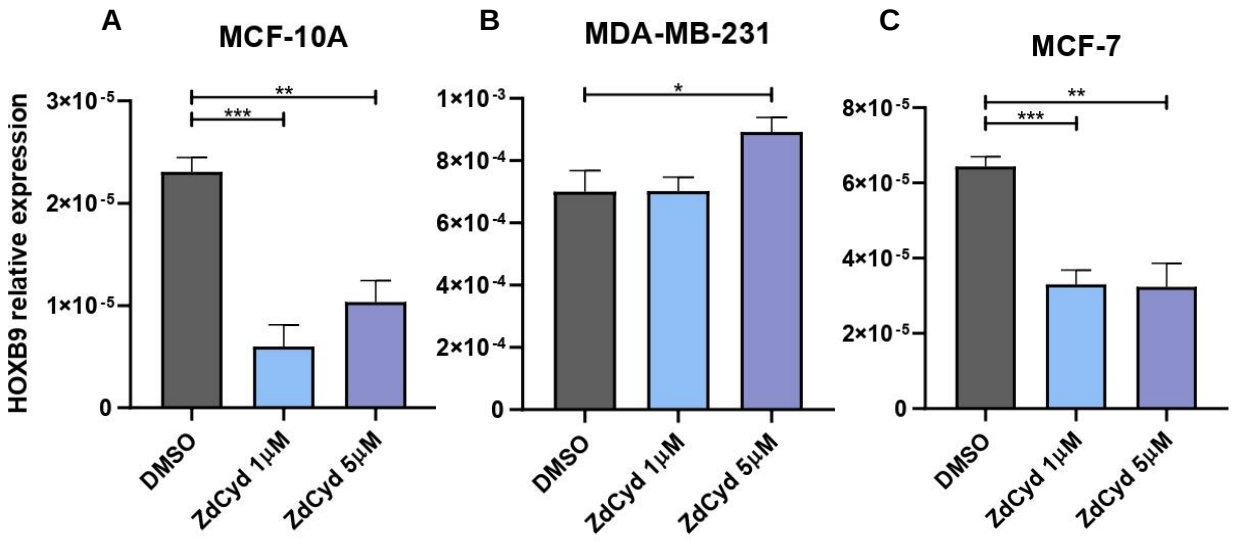

Figure 18- *HOXB9* relative expression upon ZdCyd treatment across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analysed by RT-qPCR. **A-C.** *HOXB9* relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon ZdCyd treatment at 1µM (p=0.0006 MCF-10A; p=0.9685 MDA-MB-231; p=0.0006 MCF-7) and 5µM (p=0.0017 MCF-10A; p=0.0198 MDA-MB-231; p=0.0060 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, p ≤ 0.05, **, p ≤ 0.01, ***, p ≤ 0.001). All experiments were performed in triplicate, using independent biological replicates.

4.3. Expression of *HOXB7* putative downstream ZdCyd treatment: *CDH1* and *CTNNB1*

Our laboratory recently proposed that *CDH1* and *CTNNB1* might be putative targets of *HOXB7* in Triple-Negative BrCa cells [145]. To gain further insight into this idea, the expression levels of these genes in specific cells lines (MCF-10A, MDA-MB-231, and MDA-MB-468) were evaluated upon treatment with ZdCyd. As in previous assays, the cells were subjected to treatment with ZdCyd at the concentrations of 1µM and 5µM, and expression variations quantified using RT-qPCR (Figure 19 and Figure 20).

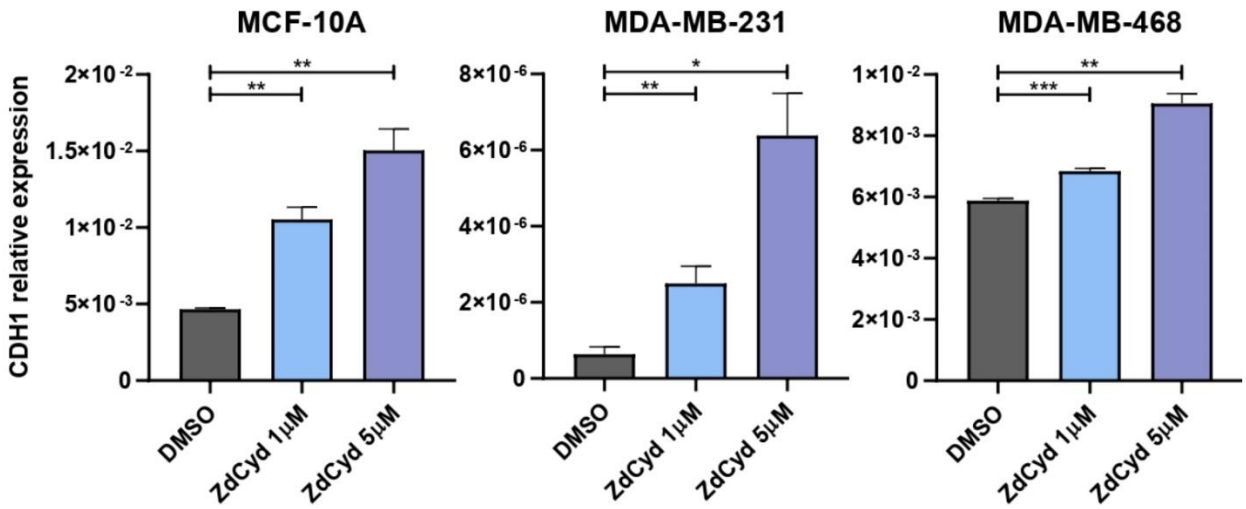

Figure 19- *CDH1* relative expression upon ZdCyd treatment across cell lines MCF-10A (A), MDA-MB-231 (B) and MDA-MB-468 (C), analysed by RT-qPCR. A-C. *CDH1* relative expression in MCF-10A, MDA-MB-231 and MDA-MB-468 upon ZdCyd treatment at 1µM (p=0.0057 MCF-10A; p=0.0097 MDA-MB-231; p=0.0001 MDA-MB-468) and 5µM (p=0.0059 MCF-10A; p=0.0104 MDA-MB-231; p=0.0022 MDA-MB-468). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, p ≤ 0.05, **, p ≤ 0.01, ***, p ≤ 0.001). All experiments were performed in triplicate, using independent biological replicates.

The results obtained suggest a significant increase in *CDH1* expression upon ZdCyd treatments across all cell lines analysed, when compared to the DMSO control (Figure 19). This result suggests that *CDH1* is controlled by methylation in these cell lines, which is concordant with previous findings [158, 159].

Furthermore, gene expression analyses revealed a significant decrease in *CTNNB1* expression in MCF-10A cells, in both treatment conditions, when compared to the control (Figure 20A). Similar to what was verified regarding *HOXB8* and *HOXB9* expression in these cells, ZdCyd

treatment might interfere in the regulatory network of *CTNNB1*, by increasing an upstream regulator gene that consequently downregulates *CTNNB1*. This is highly suggestive that, even though methylation might not directly regulate *CTNNB1* expression, it might interfere with its regulatory network, indirectly modulating its expression.

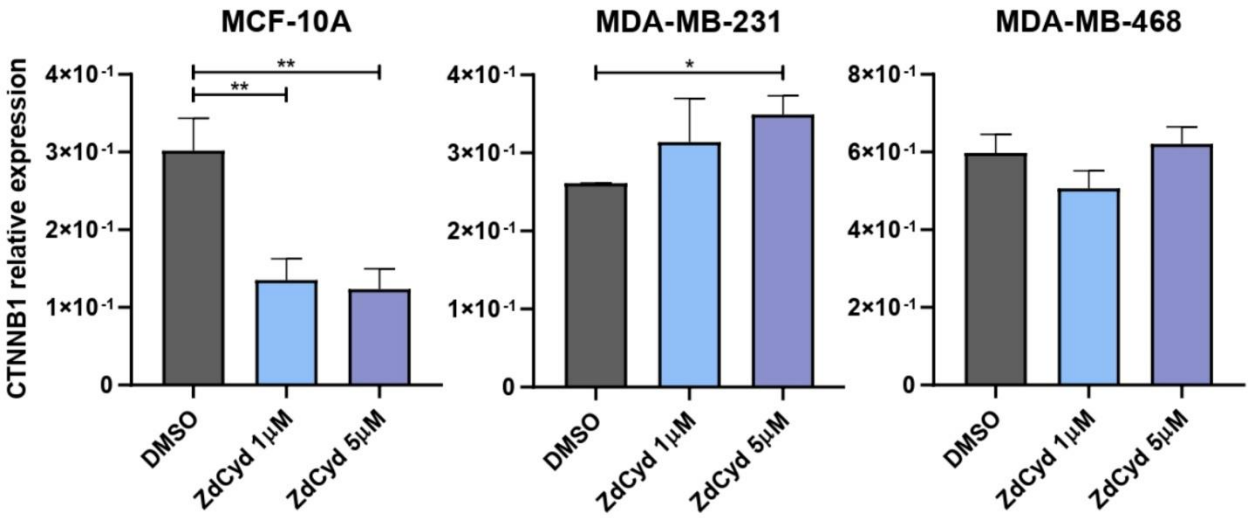

Figure 20 - *CTNNB1* relative expression upon ZdCyd treatment across cell lines MCF-10A (A), MDA-MB-231 (B) and MDA-MB-468 (C), analysed by RT-qPCR. A-C. *CTNNB1* relative expression in MCF-10A, MDA-MB-231 and MDA-MB-468 upon ZdCyd treatment at 1µM (p=0.0071 MCF-10A; p=0.2472 MDA-MB-231; p=0.0740 MDA-MB-468) and 5µM (p=0.0058 MCF-10A; p=0.0231 MDA-MB-231; p=0.5767 MDA-MB-468). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, $p \leq 0.05$, **, $p \leq 0.01$). All experiments were performed in triplicate, using independent biological replicates.

In contrast, *CTNNB1* expression increased upon treatment with 5µM ZdCyd in cell line MDA-MB-231 (p=0.0231) (Figure 20B), suggesting that methylation directly modulates *CTNNB1* expression in these cells, likewise what was found for *HOXB7* and *CDH1*. However, no significant differences were found in the expression of *CTNNB1* for cell line MDA-MB-468 (Figure 20C), which indicates that this regulatory process does not modulate the expression of this gene in a similar manner for

all Triple-Negative BrCa. Thus, interference of methylation in HOXB7 targets seems to be highly dependent on the BrCa cell type, and not exclusively to molecular subtype.

4.4. *HOXB7* expression regulation through histone deacetylation

To explore the role of histone deacetylation in *HOXB7* expression, cell lines MCF-10A, MDA-MB-231, MCF-7, and MDA-MB-468 were subjected to treatments with 400nM of Trichostatin A (TSA). This compound is a streptomyces product that inhibits all class I and II histone deacetylases (HDACs) in a non-competitive and reversible manner, and can induce apoptosis in cancer cells [160-162].

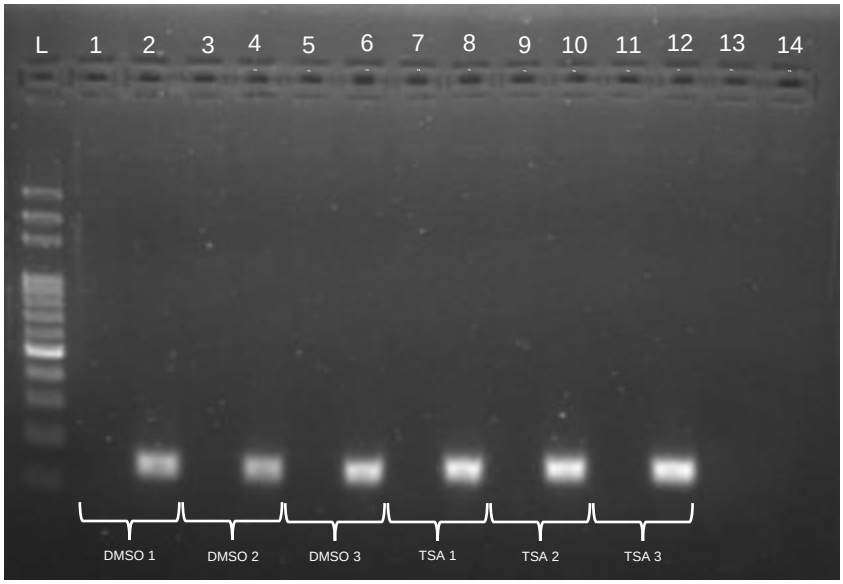

Figure 21- 2% Agarose gel representative of *HOXB7* and *GAPDH* expression on MCF-10A cells upon treatment with 400nM TSA. (L) 100bp Ladder. **1,3,5,7,9,11:** Lanes corresponding to *HOXB7* expression. **2,4,6,8,10,12:** Lanes corresponding to GAPDH amplification. **1,2:** Lanes corresponding to biological replicate 1 of DMSO control. **3,4:** Lanes corresponding to biological replicate 2 of DMSO control. **5,6:** Lanes corresponding to biological replicate 3 of DMSO control. **7,8:** Lanes corresponding to biological replicate 1 of 400nM TSA treatment. **9,10:** Lanes corresponding to biological replicate 2 of 400nM TSA treatment. **11,12:** Lanes corresponding to biological replicate 3 of 400nM TSA treatment. **13.** Lane corresponding to *HOXB7* negative control. **14.** Lane corresponding to *GAPDH* negative control. *GAPDH* was used as gene expression control. Primers used in PCR amplification are represented on

Table *4*.

Given that *HOXB7* expression is virtually inexistent under normal conditions, a PCR followed by agarose gel was used to assess whether TSA treatment altered this gene's expression. As observed in Figure 21, MCF-10A cells do not express *HOXB7*, neither in the control conditions or upon treatment. This suggests that the TSA treatment does not affect *HOXB7* expression in

normal conditions and, therefore, histone deacetylation does not seem to be the mechanism responsible for the repression of *HOXB7* expression in non-malignant cells.

However, treatment with 400nM TSA revealed a significant downregulation of *HOXB7* across all the malignant cell lines studied: MDA-MB-231, MCF-7, and MDA-MB-468 (Figure 22). This suggests that histone deacetylation may interfere with upstream repressors of *HOXB7* expression in malignant conditions.

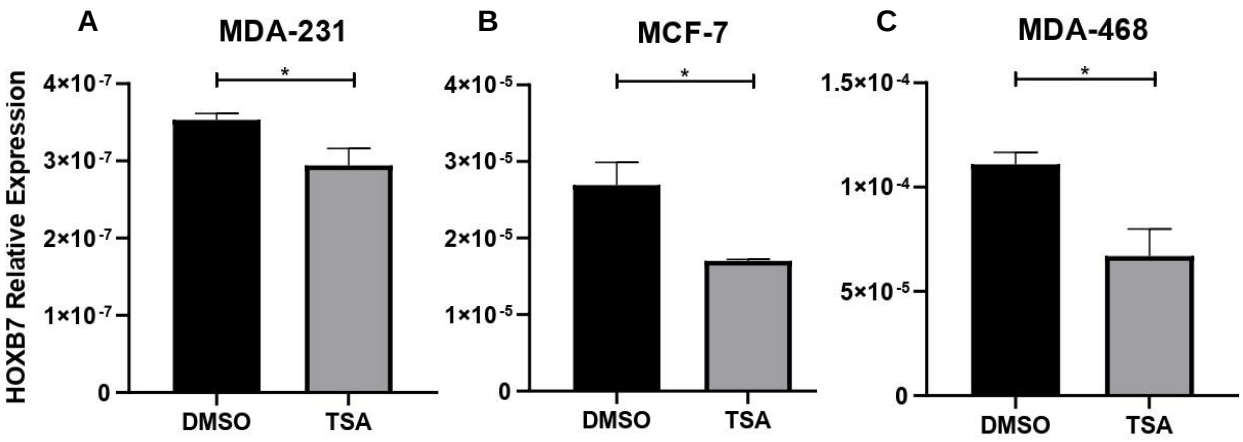

Figure 22- *HOXB7* relative expression upon TSA treatment across cell lines MDA-MB-231 (A), MCF-7 (B), and MDA-MB-468 (C) analysed by RT-qPCR. **A-C.** *HOXB7* relative expression in MDA-MB-231, MCF-7 and MDA-MB-468 upon treatment with 400nM of TSA (p=0.0307 MDA-MB-231; p=0.0283 MCF-7; p=0.0155 MDA-MB-468). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Unpaired T-Test with Welch correction (*, $p \leq 0.05$). All experiments were performed in triplicate, using independent biological replicates.

4.5. Exploring targeted therapies to overcome *HOXB* dysregulation in BrCa

Over the last decade, triazoles and oxindoles have been increasingly studied, as they are considered extremely valuable frameworks in pharmaceutical chemistry due to its biological and pharmacological properties, especially as anti-cancer and antibiotic compounds [163-166]. With that in mind, and giving the ongoing effort to discover and screen new anti-tumour drugs, Professor Burke's Lab at University of Évora have developed several N-1,2,3-Triazole-Isatin Derivatives, five of which were tested in our lab to assess its cytotoxic and anti-proliferative effect

on BrCa cells, as well as target identification[1] (Figure 23). The provided compounds can be divided in two categories: isatin-triazoles **(1),** which comprise compounds (1a), (1b) and (1c), and 3-aryl-2-hydroxyoxindole-/acetal-protected-oxindole triazoles **(2)**, which comprise compounds (S)-(2b) and (R)-(2j).

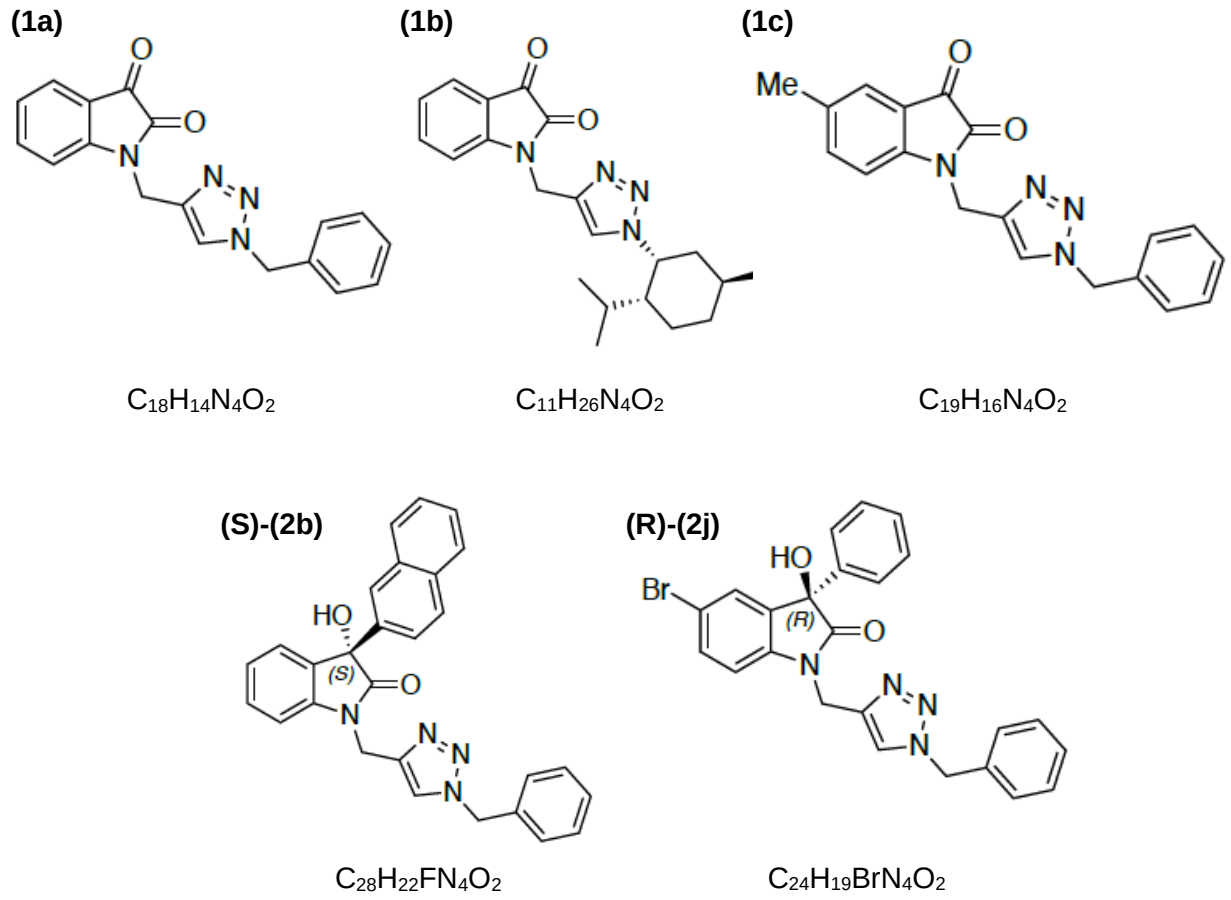

Figure 23- Structure and molecular formula of the five N-1,2,3-Triazole-Isatin derivatives used in this assay: isatin-triazoles (1a), (1b) and (1c) and 3-aryl-2-hydroxyoxindole-/acetal-protected-oxindole triazoles (S)-(2b) and (R)-(2j).

[1] Unpublished results

4.5.1. Anti-Proliferative effect of N-1,2,3-Triazole-Isatin Derivatives

The cytotoxic effects of N-1,2,3-Triazole-Isatin Derivatives (1a), (1b), (1c), (S)-(2b) and (R)-(2j) were evaluated on BrCa cell lines MCF-7 and MDA-MB-231, and in non-tumorigenic MCF-10A cells, using MTT assays. The calculated IC_{50} values are shown in Table 8, next to the results obtained for the human embryonic kidney reference cell line Hek293.

Table 8- Calculated IC_{50} of the target compounds against MCF-7, MDA-MB-231, MCF-10A and Hek293 cell lines using the MTT assay. Enantiomeric purity was Determined using HPLC with a chiral column, and is presented in percentage of enantiomeric excess (%ee). IC_{50} values calculated from MTT data using the *GraphPad Prism* Software Inc. (version 8.0.1), and are presented in µM ± 95% Confidence Level Intervals (CI).

Compound	Enantiomeric purity (%ee)	IC_{50} (µM) MCF-7	IC_{50} (µM) MDA-MB-231	IC_{50} (µM) MCF-10A	IC_{50} (µM) Hek293
(1a)	-	1.39 ± 0.2	2.95 ± 0.4	1.80 ± 0.3	63.7 ± 4.5
(1b)	-	1.34 ± 0.2	1.79 ± 0.2	1.08 ± 0.2	23.1 ± 1.4
(1c)	-	1.50 ± 0.2	2.40 ± 0.7	1.54 ± 0.2	26.9 ± 1.7
(*S*)-(2b)	88	1.75 ± 0.5	3.10 ± 0.4	1.33 ± 0.2	12.4 ± 1.1
(*R*)-(2j)	92	1.49 ± 0.3	1.41 ± 0.2	1.02 ± 0.1	36.1 ± 1.9

Overall, N-1,2,3-triazole-isatin derivatives presented slightly higher cytotoxic activity in the MCF-7 cell line than in MDA-MB-231, for all compounds except (R)-(2j), the only one that presented a lower IC_{50} in MDA-MB-231 than in MCF-7. Compounds (1a) and (1b) showed the strongest anti-proliferative activity in MCF-7 cells. On the other hand, (R)-(2j) seemed to be the most effective against MDA-MB-231, followed by compound (1b). However, compound (S)-(2b) presented the lowest IC_{50} values out of the 3-aryl-2-hydroxyoxindole-/acetal-protected-oxindole triazoles for both MCF-7 and MDA-MB-231.

Cytotoxic activity of the studied compounds was also found in MCF-10A cells, derived from non-malignant fibrocystic breast cells and commonly considered as representatives of the non-tumorigenic condition. These results were particularly evident for compounds (1b) and (R)-(2j), as they presented the lowest IC_{50} value in MCF-10A, but also revealed a lower IC_{50} than the ones obtained for cell lines MCF-7 and MDA-MB-231. Furthermore, all compounds seemed to present lower cytotoxicity for MDA-MB-231 cells than MCF-10A.

Gratifying, all the tested compounds showed very low IC_{50} values in breast cells, compared to the ones obtained in the human embryonic kidney cells Hek293 that correspond to cells derived from

the adrenal gland of an embryonic kidney, and that are commonly used as controls representing cells non-malignant and with normal growth. Due to their origin, these cells are highly vulnerable to *in vitro* environmental factors, making them widely employed as a model for novel drug discovery and toxicity testing [167]. Hereby, these results suggest that these compounds might have minor side effects when cell growth is not altered, and therefore present less severe side effects on other cells of the human body.

4.5.2. Gene expression studies of the predicted biological targets

The compounds used for this study have been previously subjected to *in silico* analyses to determine possible biological targets for each compound. These analyses were performed using the *similarity ensemble approach* (SEA) and the SwissADME tools and concluded that compounds (1a), (1b) and (1c) may target acetyltransferase p-300 (coded by the *EP300* gene, also known as KAT3B). In addition, (S)-(2b) and (R)-(2j) were predicted to target acyl-protein thioesterase 2 (also known as APT-2, coded by the *LYPLA2* gene). Docking assays showed that (1a), (1b) and (1c) bind to Tyr1397, Trp1436, Trp1436, Phe1595, and His1591, while (S)-(2b) binds to Leu33, Trp148, Leu33 and Leu81 of the p-300, and (R)-(2j) to Gly80, Trp148, Leu33 and Leu81 of the APT-2 (Figure 24).

Given the *in silico* predictions, it is possible to infer which are the target proteins of these compounds, but it is important to explore the effects they might have at the gene expression level. Therefore, in order to assess whether the expression of the *EP300* and *LYPLA2* genes changed with treatment, and if feedback mechanisms would be triggered upon protein inhibition, we exposed cell lines MCF-10A, MDA-MB-231 and MCF-7 to the five previously studied compounds, at doses that inhibited cell viability by 50% (IC_{50}, calculated in Table 8).

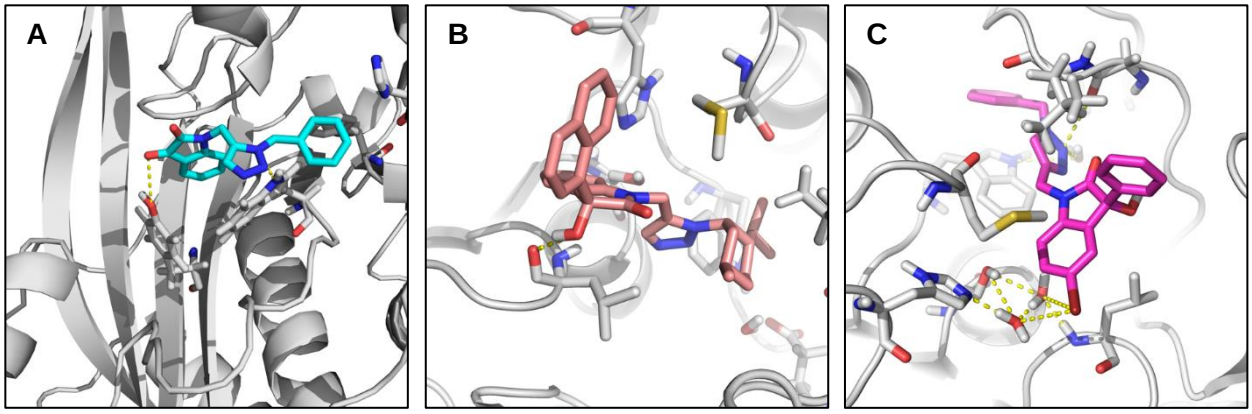

Figure 24- Docked compounds (1a), (S)-(2b) and (R)-(2j). A. Docked compound (1a) (cyan) in the binding site of acetyltransferase p-300 (in grey). **B.** Docked compound (S)-(2b) (in salmon) in the binding site of acyl-protein thioesterase 2 (in grey). **C.** Docked compound (R)-(2j) (in pink) in the binding site of acyl-protein thioesterase 2 (in grey). Oxygen atoms in red, nitrogen atoms in blue, sulphur in yellow, bromine in carmine.

Results showed that *EP300* expression was downregulated upon treatment with compound (1a), (1b) and (1c) in cell lines MDA-MB-231 and MCF-7, but not MCF-10A (Figure 25). This might indicate that all isatin-triazoles efficiently inhibit EP300 expression on malignant cells, but not on benign ones.

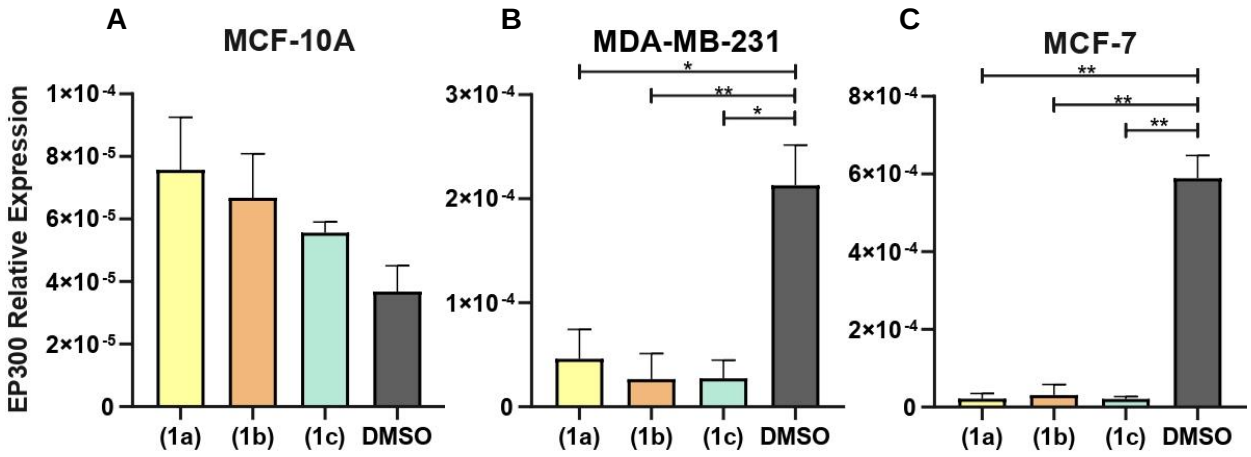

Figure 25- *EP300* relative expression upon treatment with compounds (1a), (1b) and (1c), across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analysed by RT-qPCR. A-C. *EP300* relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon treatment with compounds (1a) (p=0.0802 MCF-10A; p=0.0113 MDA-MB-231; p=0.0048 MCF-7), compounds (1b) (p=0.0955 MCF-10A; p=0.0086 MDA-MB-231; p=0.0020 MCF-7), and compound (1c) (p=0.0876 MCF-10A; p=0.0128 MDA-MB-231; p=0.0064 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Brown-Forsythe and Welch ANOVA test with Dunnett T3 correction (*, $p \leq 0.05$, **, $p \leq 0.01$). All experiments were performed in triplicate, using independent biological replicates.

Furthermore, gene expression analyses of *LYPLA2* showed a significant decrease of *LYPLA2* expression on MDA-MB-231 upon treatment with both compounds (Figure 26B). However, *LYPLA2* expression only decreased upon treatment with (R)-(2j) on cell line MCF-7 (Figure 26C). In fact, when treated with (S)-(2b), *LYPLA2* expression significantly increased (Figure 26C). Furthermore, no significant differences were shown upon treatment in MCF-10A (Figure 26A). These results might illustrate feedback mechanism on a genetic level on cancer cells that does not occur on non-malignant cells.

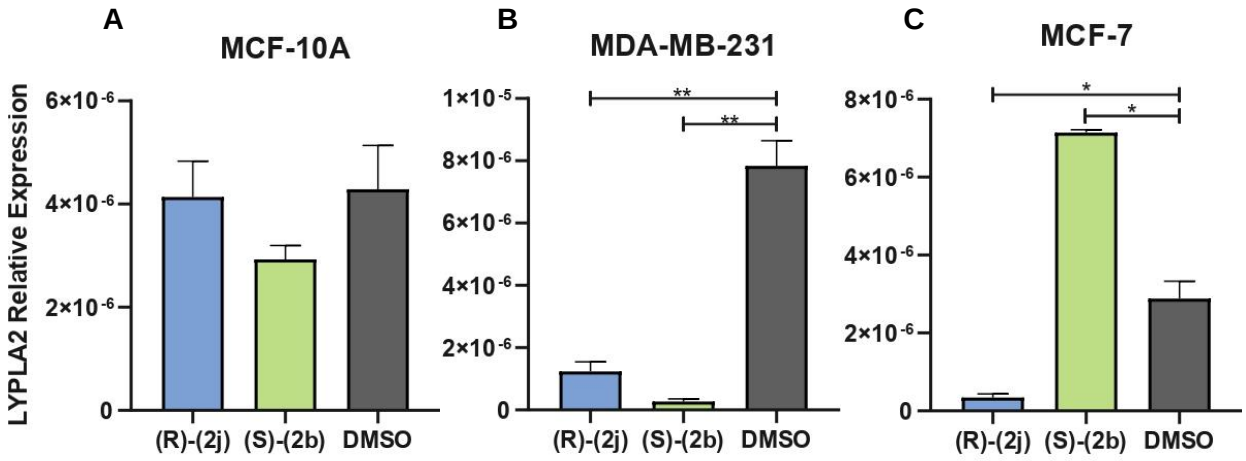

Figure 26- *LYPLA2* relative expression upon treatment with compounds (R)-(2j) and (S)-(2b) across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analysed by RT-qPCR. A-C. *LYPLA2* relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon treatment with compounds (R)-(2j) (p=0.9603 MCF-10A; p=0.0032 MDA-MB-231; p=0.0127 MCF-7), and compound (S)-(2b) (p=0.1524 MCF-10A; p=0.0054 MDA-MB-231; p=0.0047 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Brown-Forsythe and Welch ANOVA test with Dunnett T3 correction (*, $p \leq 0.05$, **, $p \leq 0.01$). All experiments were performed in triplicate, using independent biological replicates.

4.5.3. N-1,2,3-Triazole-Isatin Derivatives as modulators of *HOXB7* expression

Since epigenic regulation through acetylation is a result of an interplay between HATs and HDACs, and given that this dissertation aimed to unravel the regulatory mechanisms of *HOXB7* epigenetic regulation, the next obvious step was to determine the effect of triazole-isatin derivates (1a), (1b) and (1c) had on *HOXB7* expression, as they were shown to be inhibitors of HAT protein p-300. Furthermore, a following analysis on the effect of 3-aryl-2-hydroxyoxindole-/acetal-protected-oxindole triazoles (R)-(2j) and (S)-(2b), would also aid to illustrate the role of post-translational modifiers on the regulation of *HOXB7* expression, as these compounds are predicted

to target acyl-protein thioesterase 2 (APT-2), which is responsible for hydrolysing fatty acids from S-acylated cysteine residues in some proteins. Up to date, no studies have been performed to assess the relationship between APT-2 and *HOXB7* expression. This assay aimed to determine whether the inhibition of these proteins significantly altered *HOXB7* expression, whether by epigenetic regulation or by direct targeting of this gene.

Tests performed using the isatin-triazoles (1a), (1b) and (1c) on cell line MCF-10A, showed a significant upregulation of *HOXB7* expression (Figure 27A), while treatment with the same compounds resulted in a significant decrease of *HOXB7* expression on MCF-7 cell line (Figure 27B). In the case of MDA-MB-231, it was possible to observe a significant increase in *HOXB7* expression, only upon treatment with compounds (1a) and (1c). In contrast, compound (1b) led to an increase of *HOXB7* expression but not in a significant manner (Figure 27C).

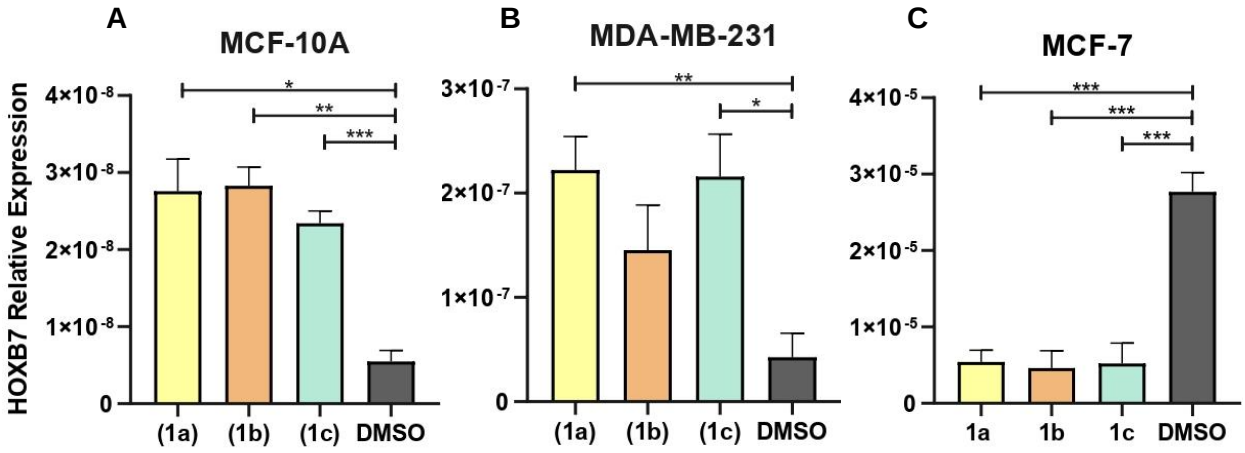

Figure 27- *HOXB7* relative expression upon treatment with compound (1a), (1b) and (1c), across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analysed by RT-qPCR. A-C. *HOXB7* relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon treatment with compounds (1a) (p=0.0133 MCF-10A; p=0.0048 MDA-MB-231; p=0.0001 MCF-7), compound (1b) (p=0.0012 MCF-10A; p=0.0733 MDA-MB-231; p=0.0002 MCF-7), and compound (1c) (p=0.0004 MCF-10A; p=0.0145 MDA-MB-231; p=0.0006 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Brown-Forsythe and Welch ANOVA test with Dunnett T3 correction (*, $p \leq 0.05$, **, $p \leq 0.01$, ***, $p \leq 0.001$). All experiments were performed in triplicate, using independent biological replicates.

For compounds (R)-(2j) and (S)-(2b), on the other hand, no alterations were observed in reference to *HOXB7* expression in MCF-10A cells (Figure 28A), although the results obtained for cell lines MDA-MB-231 and MCF-7 showed a similar tendency to the ones found for isatin-triazoles. In fact, treatment with both compounds resulted in a significant increase of *HOXB7* expression in MDA-MB-231 (Figure 28B), and a significant downregulation of this gene in MCF-7 (Figure 28C).

These results suggest that *HOXB7* expression is regulated by mechanisms involving p300 and APT-2, that ultimately results in expression variations upon their treatment-induced inhibition.

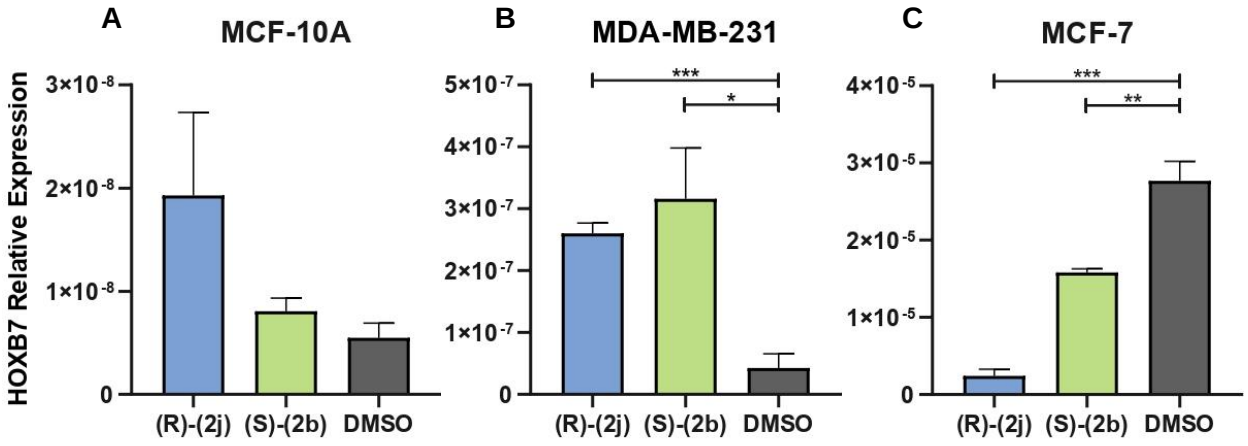

Figure 28- *HOXB7* relative expression upon treatment with compound (R)-(2j) and (S)-(2b) across cell lines MCF-10A (A), MDA-MB-231 (B) and MCF-7 (C), analysed by RT-qPCR. A-C. *HOXB7* relative expression in MCF-10A, MDA-MB-231 and MCF-7 upon treatment with compounds (R)-(2j) (p=0.1426 MCF-10A; p=0.0005 MDA-MB-231; p=0.0001 MCF-7), and compound (S)-(2b) (p=0.1313 MCF-10A; p=0.0348 MDA-MB-231; p=0.0033 MCF-7). Each column represents the mean ± SD. Significant differences indicated with asterisks based on the analyses using Brown-Forsythe and Welch ANOVA test with Dunnett T3 correction (*, $p \leq 0.05$, **, $p \leq 0.01$, ***, $p \leq 0.001$). All experiments were performed in triplicate, using independent biological replicates.

5. DISCUSSION

5.1. Regulation of *HOXB7* through DNA methylation

HOXB7 is a master transcriptional regulator implicated in several developmental processes, including in the formation of the MG [72]. This gene has been shown to regulate a plethora of cellular processes and that, when dysregulated, contribute to carcinogenesis [72, 145]. In fact, its overexpression has been linked to BrCa, in cell lines and tumours, in a molecular subtype dependent manner [72, 145]. However, no studies have yet been performed on the regulatory events leading to its overexpression in BrCa.

For many years, researchers have demonstrated the potential role of DNA methylation in the regulation of gene expression. These modifications can directly alter transcription factor that interact with their regulatory sequences. In addition, they can also modify the recruitment of methylated DNA binding proteins that recruit enzymes, such as histone deacetylases and histone methyltransferases, and ultimately alter chromatin structure [70, 122, 124, 125, 128, 129, 133, 138, 142, 168-171]. DNA methylation is one of the most well-studied anomalies in human cancers, which make it a preferential candidate to study the underlying mechanisms of *HOXB7* dysregulation.

In this work, we found that *HOXB7* expression increased in all cell lines upon being treated with an inhibitor of DNA methylation, ZdCyd, which strongly suggests that methylation is one of the processes that regulates this gene's expression, and that it may play a role in the dysregulation of *HOXB7* expression during breast carcinogenesis. Nonetheless, this increase seems to be dependent on the molecular subtype and inversely proportional to *HOXB7* endogenous expression. Furthermore, given that exposure to a DNA methylation inhibitor appears to activate *HOXB7* expression in MCF-10A cells, and that less drastic increments in *HOXB7* expression are detectable in the malignant cell lines analysed, we hypothesize that the *HOXB7* promoter becomes less methylated during the development of BrCa, particularly in Luminal A, Luminal B and Triple negative molecular subtypes. Thus, promotor hypomethylation might explain the overexpression of *HOXB7* in these cell lines. Processes of global hypomethylation of genomic DNA have been frequently described in cancer [168, 169] and *HOXB7* appears to be another gene affected by this mechanism, which might lead to its overexpression in BrCa.

Tissue-specific (T-DMR) differentially methylated DNA regions, commonly linked to developmental genes, particularly members of the *HOX* family, overlap a tiny fraction of annotated

gene promoters [169, 172]. Some of the genes with T-DMR promoters become activated when demethylation with ZdCyd is induced experimentally [173], which is compatible with the results obtained in our analysis. Following this line of thought, we decided to further analyse the methylation profile of the *HOXB7* promoter. This revealed that cell lines MCF-10A, MDA-MB-231 and SKBR3 show 50.9%, 51.7% and 50% of methylated CpG sites inside the *HOXB7* promoter CpG island, respectively. This allowed us to take away two main conclusions.

Firstly, given that MCF-10A did not present a high level of promoter methylation for *HOXB7*, this led us to conclude that promoter methylation at the studied CpG island was not the reason why this cell line does not express this gene. Secondly, considering the close proximity of methylation percentage between the analysed cell lines, promoter CpG island hypomethylation does not seem to be the determinant factor causing *HOXB7* overexpression in MDA-MB-231 or SKBR3, as these cell lines do not present a less methylated promoter CpG island than MCF-10A. The modest methylation of *HOXB7* promoter region, in conjunction with the significant increase in gene expression upon treatment, might also imply that DNA methylation does not restrict transcription but that treatment-induced gene up-regulation due to a secondary impact or the participation of other mechanisms that are also altered by ZdCyd exposure.

These results were particularly surprising, although they are not unprecedented. Studies had already shown that ZdCyd is capable of inducing gene expression, in ways that were not necessarily dependent on DNA demethylation [174, 175]. In fact, Tachibana and colleagues have observed a reactivation of *CDH1* expression upon ZdCyd treatment in cells that originally did not express this gene, without detectable differences on CpG island methylation or histone modification [175]. This study indicates that ZdCyd treatment might increase gene expression without necessarily causing demethylation of its CpG islands, but instead by affecting other mechanisms that regulate this gene's transcription [175].

A possible explanation would be the existence of CpG sites that are crucial for the activation of a certain gene, and whose demethylation can increase gene transcription, but that are not comprised within our analysis. These sites might not necessarily be within the analysed CpG island of the promoter region but within the gene sequence, or other non-CpG island sites of the promoter that we did not analyse. This can be explained by the fact that the phenomena of DNA hypomethylated implicated in cancer have been shown, not only to occur in sequences around genes (e.g., promoters), but also in the genes themself [169, 175].

Furthermore, ZdCyd exposure could induce the formation of an abnormal nucleosomal conformation that facilitates the transcription of genes that were initially not expressed, or that cause genetic upregulation [174]. ZdCyd treatments could produce changes in the chromatin structure without altering epigenetic modifications, through influencing pathways involving ATP-dependent chromatin remodelling factors, such as SWI/SNF complexes [176]. The SWI/SNF complex regulates transcriptional activity by disrupting chromatin structure, and is essential for *Homeobox* gene activation in *Drosophila melanogaster* [176-178]. In addition, this compound could lead to treatment-induced histone modifications, such as reduction of repressive histone tail marks or gain of active marks in the region around gene's cytosines [174, 175]. An example of that is that ZdCyd exposure is capable of removing methylation from histone hallmark H3K9 (Histone H3 Lysine 9), associated with chromatin inactivation [179]. In fact, this might lead to the binding of upstream transcription factors that can reactivate *HOXB7* gene expression.

Moreover, treatment could also induce the upregulation of genes that code transcription factors that are able to bind to methylated sites [174, 175], and whose binding leads to an increase in *HOXB7* expression. It is noteworthy that the *HOXB7* promoter belongs to the TATAA-less family of gene promoters, which are generally controlled by GC boxes, CCAAT boxes, and other non-specific elements [173]. Inside the CpG islands analysed, a GC box containing a Sp1 consensus sequence (GGC GGG), is located at CpG site 27. The binding of the transcription factor Sp1 to a GC box aids in the assembly of the transcriptional machinery, increasing the transcription of that gene [173, 180]. Studies have found Sp1 to be a regulator of *HOXB7* expression and that this site is necessary for *HOXB7* promoter activation [173]. Hereby, and given that methylation of certain cytosine residues in or near transcription regulatory motifs has been shown to prevent transcription factor accessibility [170], methylation at this particular site (CpG site 27, in our analysis) could prevent Sp1 binding to the *HOXB7* promoter. However, cytosine residues flanking functional Sp1 binding site at CpG site 27 are methylated on only one of the MCF-10A sequence clones. This likely dismisses the possibility that *HOXB7* repression in non-malignant cells is caused by Sp1's inability to bind to the *HOXB7* promoter. Although Sp1 is a ubiquitous transcriptional activator, when regarded as part of a multifactorial complex, it may contribute to specific regulatory activity [173], therefore, a better comprehension of the factors regarding this regulation should be taken into consideration. Promoter binding of *HOXB7* transcription factors should also be considered regarding the point mutation observed in the *HOXB7* promoter of SKBR3. The reported substitution of thymine for adenine could create a new binding site for a

transcription factor that enhances *HOXB7* transcription, thus promoting this genes' overexpression in SKBR3.

During the analysis of the methylation pattern of *HOXB7* promoter in both cell lines, we were able to identify highly conserved methylation patterns (namely sites 4 to 8, 10, 13, 14, 16, and 19). The consistent methylation in all clones of both cell lines strongly suggests that these sites are unlikely to be responsible for the differences in *HOXB7* endogenous expression between both cells.

With all these evidence in mind, it is also important to consider that the specific removal of DNA methylation from transcription regulatory areas may facilitate, but not independently cause, alterations in expression [169]. Hereby, it is difficult to reach definite conclusions on the role of promoter methylation in *HOXB7* dysregulation based only on the provided evidence, as the mechanisms underlying this phenomenon may vary depending on each molecular subtype, and even among cell lines. Additional studies comprising the analysed region of *HOXB7* promoter in other cell lines such as MCF-7, BT-474, and MD-MB-468 should be performed, as well as studies including the analysis of other regions of the *HOXB7* gene and promoter that have not yet been analysed in this study. In addition, studies in human BrCa samples from distinct molecular subtypes should be also considered.

5.2. Regulation of *HOXB8* and *HOXB9* through DNA methylation

The *HOXB8* gene has been implicated in malignancies such as gastric, ovarian, and colorectal cancers, with effects on cell migration, invasion, and metastasis[112-117]. However, the involvement of this gene in BrCa is still poorly characterized [81, 112-117]. On the other hand, *HOXB9* is another *HOXB* gene that partakes in MG development and that appears dysregulated in BrCa, being associated with increased cell motility, angiogenesis, and EMT induction, resulting in highly vascularized higher-grade tumours [81, 118, 119]. Even though both genes were found to be dysregulated in cancer, no explanation for their dysregulation has been found so far. Given that methylation is the biggest cause of epigenetic dysregulation in cancer, we decided to preliminary investigate its effects on *HOXB8* and *HOXB9* expression in BrCa cell lines MDA-MB-231 and MCF-7, and non-malignant cell line MCF-10A, upon treatment with DNA methylation inhibitor ZdCyd. We found that both genes displayed a similar tendency of expression upon treatment within the same cell lines, notably, a decrease in *HOXB8* and *HOXB9* expression in MCF-10A and MCF-7, and a significant increase of both genes in MDA-MB-231.

The increase in these genes' expression in MDA-MB-231 can be explained similarly to what was found for *HOXB7*: treatment with ZdCyd led to a loss of methylation in these genes' promoter region, leading to an increase in *HOXB8* and *HOXB9* expression. Once again, this raises the hypothesis that these genes' expression is regulated by methylation in this cell line, and that their promoters might become less methylated during the development of BrCa in Triple-Negative molecular subtypes. However, promoter methylation analysis through bisulphite sequencing should be performed in the future, along with analysis in other cell lines representative of this subtype, such as MDA-MB-468.

On the other hand, gene expression assays for MCF-10A and MCF-7 revealed that both *HOXB8* and *HOXB9* become downregulated upon exposure to ZdCyd. In theory, treatment with a DNA methylation inhibitor should result in genome demethylation and consequently in gene upregulation [156, 157]. However, exposure to this compound might induce the expression of a strong transcriptional repressor of *HOXB8* and *HOXB9*, which results in the downregulation of these genes' expression. Downregulation of some genes upon ZdCyd treatment had already been reported in previous studies [181-183], which indicates that the effects on gene expression produced by this compound are gene- and cell line-specific.

Both *HOXB8* and *HOXB9* are endogenously upregulated in MCF-7 and MDA-MB-231, being this overexpression much more accentuated on the latter for both genes [81]. The similar expression pattern, both endogenously and upon treatment suggests that these genes are regulated in a similar manner in BrCa cells, perhaps even suggesting a synergic function between both genes. This is not only supported by the observed levels of expression in both conditions (treatment and control) [81], but also by the fact that these genes are often co-expressed in human development, as well as the development of other metazoans [184]. Because HOX proteins have a high degree of structural and functional similarity, synergistic effects involving many of these transcription factors are frequently reported during development [185-187]. In the context of development and illness, compensatory effects between HOX proteins were also discovered [188, 189]. Furthermore, genes from the same cluster were shown to share transcriptional regulatory mechanisms, a phenomenon that has been thoroughly explored in the *HOXD* cluster [190]. These findings raise the hypothesis that these genes share similar function and regulation, that can be dependent on BrCa molecular subtypes. Nonetheless, further studies should be performed to assess the regulatory mechanism underlying *HOXB8* and *HOXB9* regulation in BrCa.

5.3. Regulation of *CDH1* and *CTNNB1* through DNA methylation

Breast cancer is still one of the leading causes of mortality in women, and the epithelial-mesenchymal transition (EMT) is presumably a key process in cancer progression associated with the onset of the metastatic process and higher-grade tumours [159, 191]. The EMT is thought to be intimately related to tumour cell invasion and migration, and it is defined by a cellular phenotypic change that involves the acquisition of mesenchymal features and the loss of epithelial properties [116, 192]. However, this phenotype presents some level of plasticity, which enables the cells to transition between states [116, 192]. E-cadherin/ β-catenin protein complexes seem to play a key role in EMT, as this balance is responsible for the maintenance of cell adherents junction between epithelial cells [145]. Disrupt of this interaction can lead to fibrosis and cancer, which makes it a promising new target for therapy of these diseases [192].

Recent studies in the host laboratory have proposed that *CDH1* and *CTNNB1* might be putative targets of the HOXB7 transcription factor in Triple-Negative BrCa cells [145] and that *HOXB7* expression might mediate their methylation, by regulating *DNMT3B*, as previously described for *HOXB3* [144]. Hereby, the expression levels of these genes in Triple-Negative BrCa cells (MDA-MB-231, and MDA-MB-468) and non-malignant cells (MCF-10A) were examined after treatment with ZdCyd to get further insight into this concept.

Our studies revealed an increase in *CDH1* expression in the studied cell lines upon exposure to the DNA methylation inhibitor ZdCyd. These results suggest that this gene is regulated by methylation, and are concordant with previously reported studies [158, 159, 193]. In fact, loss of *CDH1* expression, which is often caused by promoter hypermethylation, has been extensively observed in sites of EMT, and associated with higher grade tumours and ILC [159, 182, 192, 194]. Hereby, it comes as no surprise that this gene's expression increases in cancer cells upon exposure to a de-methylating agent. However, this increased expression was not only observed in cancer cells, but also in the non-malignant cell MCF-10A. In fact, Lombaerts and colleagues have demonstrated an intermediate level of *CDH1* promoter methylation, that can be explained by their fibroblastic-like phenotype [159], and therefore explains the increased level of *CDH1* expression upon ZdCyd exposure that we have observed in this line.

CTNNB1 encodes for *β*-catenin protein, which is a key regulator in the canonical Wnt signalling, whose abnormal activation can lead to the upregulation of genes involved in the EMT [114, 191, 192]. *CDH1* loss promotes *β*-catenin release, which is accumulated in the cytoplasm, making it

available for nuclear import, and consequently inducing EMT. However, this phenomenon can be reverted by *CDH1* re-expression. Studies have shown that dysregulation in *CTNNB1* transcription promotes the expression of oncogenes, such as *c-myc* and *cyclin D1*, leading to the endorsement of the carcinogenic process.

Given that, we decided to study the effect of ZdCyd on *CTNNB1* expression and found that its effect varies according to BrCa molecular subtypes. Our results have shown a significant decrease in *CTNNB1* expression in MCF-10A upon treatment. Similar to what we hypothesized for *HOXB8* and *HOXB9*, ZdCyd treatment might induce the expression of a strong repressor of *CTNNB1* that consequently results in its downregulation. Previous studies conducted in our lab have pointed that *CTNNB1* might be a putative target of HOXB7 and that *HOXB7* overexpression can lead to the downregulation of this gene in MDA-MB-231 cells [145]. Given the very accentuated increase in *HOXB7* expression that we have observed in MCF-10A upon ZdCyd exposure, we hypothesize that this upregulation leads to repression of *CTNNB1*, in a manner that is similar to what was observed upon *HOXB7* overexpression in MDA-MB-231. These results lead to the conclusion that this gene's downregulation might be an indirect effect of the ZdCyd treatment, mediated by *HOXB7*, and thus, that methylation is not the direct regulatory mechanism behind this gene's modulation in MCF-10A.

However, we made the opposite observation in MDA-MB-231 cells, as treatment with 5µM induced upregulation of *CTNNB1*. This can suggest that methylation partakes in this gene's regulation in this cell line, or that the increased expression is a consequence of an upstream positive regulator of this gene. In spite of the evidence that suggests that *HOXB7* overexpression leads to the downregulation of this gene in MDA-MB-231 [145], one can argue that the increase of *HOXB7* expression in MDA-MB-231 upon treatment may not have been sufficient to trigger this effect, whereas the much more accentuated increase of *HOXB7* in MCF-10A may have been.

Lastly, the exposure to ZdCyd showed no effects on *CTNNB1* expression in MDA-MB-468, indicating that the methylation is not responsible for this gene's regulation in this cell line. Furthermore, this also suggests that *CTNNB1* regulation is not uniform in Triple-Negative BrCa, but instead that it is specific to particular types. Hereby, further analysis should be performed to assess the mechanistic effects behind the regulation of these targets, as their regulation can be a very promising target in cancer therapeutics [191].

5.4. Regulation of *HOXB7* through histone deacetylation (HDAC)

Histone acetylation is a synergetic process finely regulated by two counteracting enzyme families that control the acetylation state of lysine residues at the N-terminal extension of histones, and consequently modulate chromatin structure: the histone acetyl transferases (HATs) and histone deacetylases (HDACs) [195-197]. While HATs are responsible for acetylating lysine residues, which results in the dispersion of chromatin structure and consequent transcriptional activation [198], HDACs instigate chromatin condensation and increasingly block assess of transcriptional machinery to DNA, which progressively promotes gene silencing [195, 196]. The imbalance between these two mechanisms can lead to abnormal gene expression in cancer [173].

Given that the studies regarding the methylation of the *HOXB7* promoter didn't provide enough evidence to fully explain the mechanism behind *HOXB7* dysregulation in BrCa, we decided to investigate whether histone deacetylation was the reason why *HOXB7* was virtually not expressed in non-malignant cells and if the potential loss of histone deacetylation led to the overexpression of *HOXB7* in BrCa. Hereby, we subjected cell lines MCF-10A, MDA-MB-231, MDA-MB-468, and MCF-7 to a treatment with a Class I and II HDAC inhibitor (HDACi), Trichostatin A (TSA). Histone deacetylase inhibitors (HDACi) are an emerging family of anticancer drugs which are capable of restoring chromatin structure to a more open state, allowing suppressed genes to re-express and cell cycle arrest and death to occur through various mechanisms [173].

Our analysis showed that TSA was incapable to induce the expression of *HOXB7* in MCF-10A cells, therefore showing that histone deacetylation is not the reason behind *HOXB7* repression in non-malignant cells. Previous studies have shown that some genes with the T-DMR promoter can become reactivated upon ZdCyd treatment, but not upon TSA exposure [169], which supports the results obtained in this analysis.

Studies regarding malignant cell lines MDA-MB-231, MDA-MB-468, and MCF-7 revealed that TSA exposure led to a decrease in *HOXB7* across all lines. In theory, treatment with TSA would inhibit HDAC activity, leading to the unfolding of chromatin, which would allow access of transcriptional regulators to the promoter region and induce gene expression [196]. This is the opposite of what we found for these cell lines. Although this suggests that histone deacetylation might somehow interfere with *HOXB7* regulation in BrCa, this process's interference is not clear. It is noteworthy that TSA treatment in MCF-7 has previously been reported to inhibit *ESR1* transcription and consequent Erα expression [199], as well as downregulating *YCLIN D1, CDK4, CDK6,* and *BCL-XL* in Basal-Like cells [200], as is the case of MDA-MB-231 and MDA-MB-468,

thus enforcing that genetic downregulation upon TSA treatment is not an isolated event. The same study revealed that TSA disrupts the MAPK pathways, by decreasing ERK phosphorylation [200]. As HOXB7 triggers MAPK/ERK pathways and ERK phosphorylation [201], it is possible to hypothesize that TSA disruption of MAPK pathways might be a side effect of TSA-induced *HOXB7* downregulation. Previous reports have once again demonstrated that gene downregulation upon TSA treatment can be a secondary effect of HDAC inhibition on an upstream regulator of the target gene, as was seen by Song and colleagues [196]. They observed that HDAC inhibition induced *RAS* expression, which consequently promotes repression of *Spry1* via AT1R [196]. Given the intricacy that gene expression networks pose, studies encompassing upstream genes of *HOXB7* in BrCa should be performed, especially since HDAC interference appears to be a phenomenon that partakes in *HOXB7* regulation in BrCa but not in non-malignant cells, making it a possible biomarker for patients that would benefit from HDACi therapies.

5.5. Therapeutic potential of N-1,2,3-Triazole-Isatin derivatives and *HOXB7* modulation

Given the complex nature of BrCa, and heterogenous clinical outcomes that it presents, it is extremely important to discover new therapies that might efficiently target genetic and transcriptional dysregulations, in a way that can help tailor the most appropriate treatment for each patient, by reducing the adverse effects of systemic chemotherapy and increasing overall survival [202, 203]. The biological and pharmacological characteristics of triazoles and oxindoles make them very promising study candidates in pharmaceutical chemistry due to their potential as anti-cancer and antibiotic compounds [62-65].

In this dissertation, we studied the therapeutic potential of five N-1,2,3-Triazole-Isatin Derivatives in BrCa, that can be divided into two categories: isatin-triazoles **(1),** which comprise compounds (1a), (1b) and (1c), and 3-aryl-2-hydroxyoxindole-/acetal-protected-oxindole triazoles **(2)**, which comprise compounds (S)-(2b) and (R)-(2j).

The obtained results revealed that the compounds produced cytotoxic effects in all cell lines, including non-malignant MCF-10A. This may have happened in the line used as control/normal condition (MCF-10A) because it derives from proliferating fibrocystic breast cells, and not directly from "normal" breast tissue, thus presenting altered proliferative rates [204]. Thus, the results

obtained for MCF-10A cells suggest that these compounds may not be specific for malignant cells, but rather for cells with altered growth, a common characteristic shared by all strains studied.

Furthermore, MCF-7 cells appeared to be more sensitive to treatment than MDA-MB-231 cells, which suggests that treatment with these compounds might be more effective in Luminal A tumours than in Triple Negative ones. Nonetheless, the lower IC_{50} values of compounds (1a) and (1c) in MCF-7 compared to MCF-10A place both compounds as great candidates for the treatment of malignant tumours within the Luminal A subtype. Moreover, all of the studied compounds presented low IC_{50} values, compared to the ones obtained for the normal Hek293 cells, a commonly used reference model in drug testing [167]. This suggests that breast cells are far more sensitive to the effects of those compounds than normal kidney cells, potentially indicating that these drugs may have minimal side effects on normal tissues, as a much higher concentration is required to inhibit 50% of cell growth. Notably, cytotoxic effects were higher than the ones observed for any other cell type, including colon cancer cells (SW480), lung cancer cells (A549) and lymphoma models (DOHH2, VAL, OCI-LY-10), except for compound (1a) and (1c), which presented lower IC_{50} values in SU-DHL-2 (Supplementary Table 1). This indicates that the studied compounds might be highly specific for breast tissue with altered proliferation, which can be advantageous in the treatment of benign and malignant breast tumours.

The cytotoxicity of these compounds is probably related to the inhibition of their predicted targets, the histone acetyltransferase protein p-300 encoded by the *EP300* gene and the acyl-protein thioesterase 2, also known as APT-2, encoded by *LYPLA2* gene [205-207]. Together with its homolog, the CREB binding protein (CBP), p-300 acts as a transcriptional activator due to its role as a histone acetyltransferase (HAT) [205, 208]. However, p-300 is often responsible for acetylation of other transcription-related factors, and may suffer autoacetylation, which results in gene transcription activation [205, 208, 209]. They are also responsible for controlling transcription by functioning as scaffolds that connect sequence-specific DNA binding factors with the basal transcriptional machinery [205].

Given its function as HAT, p300 has been implicated in a variety of cellular processes such as proliferation, differentiation, and apoptosis, as well as in DNA repair synthesis upon UV or radiation-induced DNA damage [209, 210]. While *EP300*, the gene coding for p-300, is considered an onco-suppressor in normal epithelial cells, it is overexpressed in various tumours and associated with lower overall survival [210]. This phenomenon is probably because this protein's affinity to substrate is highly dependable on the availability of acetyl-CoA, which is often

dysregulated in cancer cells, due to their abnormal metabolism [209], but also by their frequent exploitation as cofactors of oncoproteins [205]. Given the function that this protein present in carcinogenic, its targeting can be an interesting approach in the development of novel pharmaceuticals [211].

Our gene expression assays, regarding this compound inhibition, revealed downregulation of *EP300* in BrCa cell lines, but not in MCF-10A, suggesting that a negative feedback loop might be activated upon protein inhibition, resulting in the downregulation of this gene in BrCa. As previously mentioned, this protein also partakes in autoacetylation, therefore, its inhibition, might impair this mechanism, and result in lower expression of this gene in malignant cells [205, 208]. It is known that p300 plays an important part in MDA-MB-231 invasion, as well as MCF-7 proliferation, mainly due to acetylation of histone H3 and H4, and particularly H3K27 acetylation [212, 213]. Furthermore, this protein is often recruited by *c-myc* to induce *RhoA* expression in MDA-MB-231, prompting invasion and metastasis [213]. Studies with other p300 inhibitors have demonstrated that its inhibition is capable of significantly revert these epigenetic marks, and therefore, reduce proliferation and invasion of these cells [211-213]. Additional studies in breast cancer cells revealed that p300 is recruited through the intermediate of MAPK, leading to a kinase mediated growth factor signalling to the oestrogen receptor via a complex cross-talk between pathways [214], and that inhibition of p300 impairs MAPK activation [215].

Our results regarding *HOXB7* expression upon treatment with our isatin-triazoles revealed a significant decrease in this gene's expression in MCF-7. As p-300 functions as an acetyltransferase, and given that its inhibition resulted in a decrease of *HOXB7* expression in MFC-7 cells, it is logical to conclude that acetylation partakes in *HOXB7* regulation in this cell line, possibly being the reason behind why this gene is overexpressed in MCF-7 [81].Considering the involvement of p300 in MAPK mediated activation of ER and the role of *HOXB7* in the MAPK pathway, it comes as no surprise that this gene is also downregulated upon p300 targeting in MCF-7 cells, which are ER+. However, this gene's expression increased in MCF-10A and MDA-MB-231, meaning that p300 inactivation might modulate *HOXB7* expression through different mechanisms, possibly through the modulation of an upstream target, and that upregulation might be an important factor for these compounds' toxicity.

APT-2 is a protein of the β-hydrolase family of serine hydrolases, that partakes in a reversible process of protein diacylation called palmitoylation [216]. Through the creation of high-energy thioester bonds, post-translational S-palmitoylation anchors target proteins to membranes [207].

This step is crucial for sorting, distribution, trafficking of key proteins that participate in cell growth and signalling [206, 207]. Evidence relates carcinogenesis, cancer cell proliferation, survival, and treatment resistance to specific palmitoylated proteins and the activity of APT enzymes. As a result, enzymes that regulate protein palmitoylation might be potential anti-cancer targets in pharmaceutical development [217], such as the case of APT-2.

In fact, Ras trafficking is blocked and Ras oncogenic activities are suppressed when APT-2 is pharmacologically inhibited [218-220]. Given the potential of APT-2 inhibition in cancer, the high cytotoxicity that we observed upon treatment with (S)-(2b) and (R)-(2j) in breast cell lines is not surprising. Our results indicate that these compounds are potent inhibitors of cellular proliferation, especially compound (R)-(2j), which presented the lowest IC_{50} value between the two compounds. Gene expression assays revealed similar results for 3-aryl-2-hydroxyoxindole-/acetal-protected-oxindole triazoles (2), as they did for isatin-triazoles. Once again, we were able to observe a downregulation of *LYPLA2* in MDA-MB-231 for both compounds, and in MCF-7 for (R)-(2j), suggesting that treatment triggers a negative feedback loop that inhibits this gene's expression as well, through a yet unknown mechanism. However, this inhibitory effect was only observed in malignant cells, as it exerted no effect on MCF-10A. Nevertheless, treatment with (S)-(2b) resulted in an upregulation of *LYPLA2* in MCF-7. This increased expression might be due to a compensatory mechanism to counteract APT-2 inhibition and might justify why this compound presented lower cytotoxic values than its enantiomeric derivative (R)-(2j).

In concern to their modulation of *HOXB7*, these compounds only sorted significant effects on malignant cells, which strongly suggests that the cause of their cytotoxic effects in non-malignant cells is distinct from malignant cells. When analysing *HOBX7* expression in MDA-MB-231 we noticed an upregulation in this gene's expression, opposingly to MCF-7, in which treatment led to a downregulation of *HOXB7*. No mechanisms regarding APT-2 mediated HOX regulation in breast cancer have been described to date. However, ER has been described as protein that suffers palmitoylation [221]. Given that this receptor is intimately related to HOXB7 pathways in breast cancer, ER palmitoylation might be a key factor to explain the different tendencies of expression upon treatment that we have observed in both cells.

Given all of this evidence, it is possible to conclude that all of these compounds constitute promising therapeutical agents in breast cancer treatment. However, the mechanisms underlying these compounds cytotoxicity need further clearance. Furthermore, it would be interesting to perform functional assays using these compounds regarding cellular motility, apoptosis, invasion

capacity, among others, as well as a western blot assay to confirm that both proteins are effectively inhibited by these compounds.

6. CONCLUSION

The research conducted during this dissertation contributed to a better knowledge of the molecular processes underlying *HOXB7* dysregulation, and the therapeutic potential of five novel compounds in BrCa.

The results suggest that loss of methylation could be one of the reasons behind *HOXB7* overexpression in breast cancer cells, as evidenced by its increase upon inhibition of methylation by ZdCyd treatment. Furthermore, it enabled the description of the methylation status of the CpG island within the *HOXB7* promoter region in several cellular models (MCF-10A, MDA-MB-231 and SKBR3) and to identify a conserved point mutation in SKBR3 cells.

In addition, we also investigated the role of methylation in *HOXB8*, *HOXB9,* and in two putative targets of *HOXB7*: *CDH1* and *CTNNB1*, in three distinct cell lines (MCF-10A, MDA-MB-231, and MCF-7). We found that *HOXB8* and *HOXB9* seem to be regulated in a similar manner, given that treatment with ZdCyd resulted in a decrease in these genes' expression in MCF-7 and MCF-10A and an upregulation in MDA-MB-231. This suggests that both genes are regulated in these cell lines, whether through direct intervention (in MDA-MB-231 cells) or indirect intervention (in MCF-7 and MCF-10A). Regarding *CDH1*, we confirmed previous results indicating that this gene is controlled by methylation in all of the analysed cells. Furthermore, *CTNNB1* regulation through methylation, is not uniform across all cells, even within the same subtype.

Deeper characterization of the epigenetic events in *HOXB7* transcriptional regulation revealed that histone deacetylation is not the mechanism suppressing HOXB7 expression in non-malignant cells, but could be repressing a *HOXB7* inhibitor in breast cancer cells from distinct molecular sub-types (MDA-MB-231, MCF-7, and MDA-MB-468).

Histone acetyltransferase inhibition was also found to cause *HOXB7* downregulation in Luminal-representative cells MCF-7, which led us to suggest that acetylation is one of the epigenetic mechanisms responsible for *HOXB7* overexpression in this molecular sub-type.

Furthermore, the *in vitro* therapeutic potential of five N-1,2,3-Triazole-Isatin derivatives was evaluated in breast cancer treatment and revealed that these compounds are able to efficiently inhibit cell growth in lines MCF-10A, MDA-MB-231, and MCF-7, which suggest that these compounds are effective in cells with altered proliferation rates.

Comparison with normal kidney cells Hek293 unveiled that while these compounds are strong inhibitors in breast cells, they need a much higher dose to induce cytotoxic effects on these cells, therefore proposing that they might cause less therapeutic side effects on normal cells. Furthermore, they seem to be highly specific for breast tissue, as studies with lymphoma, lung, and colorectal cancers revealed overall higher IC_{50} values.

Gene expression assays showed that N-1,2,3-Triazole-Isatin derivatives are also capable of modulating *HOXB7* expression, possibly through the inhibition of their target proteins, p-300 or APT-2.

Ultimately, the intricacy of the cancerous gene expression landscape makes it difficult to fully identify all the mechanisms behind its dysregulation and efficiently target it, but we believe this work might constitute a step forward in this characterization.

7. REFERENCES

1. Momenimovahed, Z. and H. Salehiniya, *Epidemiological characteristics of and risk factors for breast cancer in the world.* Breast Cancer (Dove Med Press), 2019. **11**: p. 151-164.

2. Tao, Z., et al., *Breast Cancer: Epidemiology and Etiology.* Cell Biochem Biophys, 2015. **72**(2): p. 333-8.

3. Torre, L.A., et al., *Global Cancer in Women: Burden and Trends.* Cancer Epidemiol Biomarkers Prev, 2017. **26**(4): p. 444-457.

4. Konduri, S., et al., *Epidemiology of male breast cancer.* The Breast, 2020. **54**: p. 8-14.

5. World Health, O., *WHO report on cancer: setting priorities, investing wisely and providing care for all*. 2020, Geneva: World Health Organization.

6. Dafni, U., Z. Tsourti, and I. Alatsathianos, *Breast Cancer Statistics in the European Union: Incidence and Survival across European Countries.* Breast Care, 2019. **14**(6): p. 344-352.

7. Union, E.C.S.O.o.t.E., et al., *Eurostat Regional Yearbook: 2020 Edition*. 2020: Publications Office of the European Union.

8. Allemani, C., et al., *Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries.* Lancet, 2018. **391**(10125): p. 1023-1075.

9. Wojtyla, C., et al., *European trends in breast cancer mortality, 1980–2017 and predictions to 2025.* European Journal of Cancer, 2021. **152**: p. 4-17.

10. Curado, M.-P., et al., *Cancer incidence in five continents, Volume IX*. 2007: IARC Press, International Agency for Research on Cancer.

11. Ferlay, J., et al., *Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods.* Int J Cancer, 2019. **144**(8): p. 1941-1953.

12. Forjaz de Lacerda, G., et al., *Breast cancer in Portugal: Temporal trends and age-specific incidence by geographic regions.* Cancer Epidemiology, 2018. **54**: p. 12-18.

13. Escala-Garcia, M., et al., *Breast cancer risk factors and their effects on survival: a Mendelian randomisation study.* BMC Med, 2020. **18**(1): p. 327.

14. Jenkins, E.O., et al., *Age-Specific Changes in Intrinsic Breast Cancer Subtypes: A Focus on Older Women.* The Oncologist, 2014. **19**(10): p. 1076-1083.

15. Fu, J., et al., *How Young Is Too Young in Breast Cancer?-Young Breast Cancer Is Not a Unique Biological Subtype.* Clin Breast Cancer, 2018. **18**(1): p. e25-e39.

16. Shoemaker, M.L., et al., *Differences in breast cancer incidence among young women aged 20-49 years by stage and tumor characteristics, age, race, and ethnicity, 2004-2013.* Breast Cancer Res Treat, 2018. **169**(3): p. 595-606.

17. Davis Lynn, B.C., et al., *Black-White Breast Cancer Incidence Trends: Effects of Ethnicity.* J Natl Cancer Inst, 2018. **110**(11): p. 1270-1272.

18. Brewer, H.R., et al., *Family history and risk of breast cancer: an analysis accounting for family structure.* Breast Cancer Res Treat, 2017. **165**(1): p. 193-200.

19. Copur, M.S., *Universal Genetic Testing for All Breast Cancer Patients.* Oncology (Williston Park), 2019. **33**(8).

20. Denkert, C., et al., *Molecular alterations in triple-negative breast cancer-the road to new treatment strategies.* Lancet, 2017. **389**(10087): p. 2430-2442.

21. Duffy, M.J., N.C. Synnott, and J. Crown, *Mutant p53 in breast cancer: potential as a therapeutic target and biomarker.* Breast Cancer Res Treat, 2018. **170**(2): p. 213-219.

22. Manahan, E.R., et al., *Consensus Guidelines on Genetic` Testing for Hereditary Breast Cancer from the American Society of Breast Surgeons.* Ann Surg Oncol, 2019. **26**(10): p. 3025-3031.

23. Goldberg, M., et al., *Pubertal timing and breast cancer risk in the Sister Study cohort.* Breast Cancer Res, 2020. **22**(1): p. 112.

24. Wang, J.M., et al., *Evaluation of the association between mammographic density and the risk of breast cancer using Quantra software and the BI-RADS classification.* Medicine (Baltimore), 2020. **99**(46): p. e23112.

25. Halbony, H., et al., *Breast cancer epidemiology among surgically treated patients in Jordan: A retrospective study.* Med J Islam Repub Iran, 2020. **34**: p. 73.

26. Phipps, A.I., et al., *Reproductive history and oral contraceptive use in relation to risk of triple-negative breast cancer.* J Natl Cancer Inst, 2011. **103**(6): p. 470-7.

27. Coughlin, S.S., *Epidemiology of Breast Cancer in Women.* Adv Exp Med Biol, 2019. **1152**: p. 9-29.

28. Kelsey, J.L., M.D. Gammon, and E.M. John, *Reproductive factors and breast cancer.* Epidemiol Rev, 1993. **15**(1): p. 36-47.

29. Liu, Y., N. Nguyen, and G.A. Colditz, *Links between alcohol consumption and breast cancer: a look at the evidence.* Womens Health (Lond), 2015. **11**(1): p. 65-77.

30. Subramani, R. and R. Lakshmanaswamy, *Pregnancy and Breast Cancer.* Prog Mol Biol Transl Sci, 2017. **151**: p. 81-111.

31. Del Pup, L., G. Codacci-Pisanelli, and F. Peccatori, *Breast cancer risk of hormonal contraception: Counselling considering new evidence.* Crit Rev Oncol Hematol, 2019. **137**: p. 123-130.

32. Mørch, L.S., et al., *Contemporary Hormonal Contraception and the Risk of Breast Cancer.* N Engl J Med, 2017. **377**(23): p. 2228-2239.

33. Westhoff, C.L. and M.C. Pike, *Hormonal contraception and breast cancer.* Contraception, 2018. **98**(3): p. 171-173.

34. Deli, T., M. Orosz, and A. Jakab, *Hormone Replacement Therapy in Cancer Survivors - Review of the Literature.* Pathology oncology research : POR, 2020. **26**(1): p. 63-78.

35. Vinogradova, Y., C. Coupland, and J. Hippisley-Cox, *Use of hormone replacement therapy and risk of breast cancer: nested case-control studies using the QResearch and CPRD databases.* BMJ, 2020. **371**: p. m3873.

36. Hodis, H.N. and P.M. Sarrel, *Menopausal hormone therapy and breast cancer: what is the evidence from randomized trials?* Climacteric, 2018. **21**(6): p. 521-528.

37. Yu, X., et al., *Hormone replacement therapy and breast cancer survival: a systematic review and meta-analysis of observational studies.* Breast Cancer, 2017. **24**(5): p. 643-657.

38. Berstad, P., et al., *Alcohol intake and breast cancer risk among young women.* Breast Cancer Res Treat, 2008. **108**(1): p. 113-20.

39. Ronckers, C.M., C.A. Erdmann, and C.E. Land, *Radiation and breast cancer: a review of current evidence.* Breast Cancer Res, 2005. **7**(1): p. 21-32.

40. Bahl, M., *Management of High-Risk Breast Lesions.* Radiol Clin North Am, 2021. **59**(1): p. 29-40.

41. Druesne-Pecollo, N., et al., *Excess body weight and second primary cancer risk after breast cancer: a systematic review and meta-analysis of prospective studies.* Breast Cancer Res Treat, 2012. **135**(3): p. 647-54.

42. Houghton, S.C., et al., *Central adiposity and subsequent risk of breast cancer by menopause status.* J Natl Cancer Inst, 2020.

43. Niell, B.L., et al., *Screening for Breast Cancer.* Radiol Clin North Am, 2017. **55**(6): p. 1145-1162.

44. Coleman, C., *Early Detection and Screening for Breast Cancer.* Semin Oncol Nurs, 2017. **33**(2): p. 141-155.

45. Narayan, A.K., C.I. Lee, and C.D. Lehman, *Screening for Breast Cancer.* Med Clin North Am, 2020. **104**(6): p. 1007-1021.

46. Nguyen, T.L., et al., *Interval breast cancer risk associations with breast density, family history and breast tissue aging.* Int J Cancer, 2020. **147**(2): p. 375-382.

47. Oezaras, G., et al., *Analysis of breast self-examination training efficiency in women between 20-60 years of age in Turkey.* Asian Pac J Cancer Prev, 2010. **11**(3): p. 799-802.

48. Cardoso, F., et al., *Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†.* Ann Oncol, 2019. **30**(8): p. 1194-1220.

49. Makki, J., *Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance.* Clin Med Insights Pathol, 2015. **8**: p. 23-31.

50. Badve, S.S. and Y. Gökmen-Polar, *Ductal carcinoma in situ of breast: update 2019.* Pathology, 2019. **51**(6): p. 563-569.

51. Wen, H.Y. and E. Brogi, *Lobular Carcinoma In Situ.* Surg Pathol Clin, 2018. **11**(1): p. 123-145.

52. Chen, Z., et al., *Invasive lobular carcinoma of the breast: A special histological type compared with invasive ductal carcinoma.* PLoS One, 2017. **12**(9): p. e0182397.

53. Goh, C.W., et al., *Invasive ductal carcinoma with coexisting ductal carcinoma in situ (IDC/DCIS) versus pure invasive ductal carcinoma (IDC): a comparison of clinicopathological characteristics, molecular subtypes, and clinical outcomes.* J Cancer Res Clin Oncol, 2019. **145**(7): p. 1877-1886.

54. McCart Reed, A.E., et al., *Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics.* Breast Cancer Res, 2015. **17**(1): p. 12.

55. Thomas, M., et al., *Invasive lobular breast cancer: A review of pathogenesis, diagnosis, management, and future directions of early stage disease.* Semin Oncol, 2019. **46**(2): p. 121-132.

56. Tsang, J.Y.S. and G.M. Tse, *Molecular Classification of Breast Cancer.* Adv Anat Pathol, 2020. **27**(1): p. 27-35.

57. *Comprehensive molecular portraits of human breast tumours.* Nature, 2012. **490**(7418): p. 61-70.

58. Perou, C.M., et al., *Molecular portraits of human breast tumours.* Nature, 2000. **406**(6797): p. 747-52.

59. Prat, A. and C.M. Perou, *Deconstructing the molecular portraits of breast cancer.* Mol Oncol, 2011. **5**(1): p. 5-23.

60. Provenzano, E., G.A. Ulaner, and S.F. Chin, *Molecular Classification of Breast Cancer.* PET Clin, 2018. **13**(3): p. 325-338.

61. Solanki, M. and D. Visscher, *Pathology of breast cancer in the last half century.* Hum Pathol, 2020. **95**: p. 137-148.

62. Poggio, F., et al., *Platinum-based neoadjuvant chemotherapy in triple-negative breast cancer: a systematic review and meta-analysis.* Ann Oncol, 2018. **29**(7): p. 1497-1508.

63. Telli, M.L., et al., *Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer.* Clin Cancer Res, 2016. **22**(15): p. 3764-73.

64. Pandy, J.G.P., et al., *Triple negative breast cancer and platinum-based systemic treatment: a meta-analysis and systematic review.* BMC Cancer, 2019. **19**(1): p. 1065.

65. Medina, M.A., et al., *Triple-Negative Breast Cancer: A Review of Conventional and Advanced Therapeutic Strategies.* Int J Environ Res Public Health, 2020. **17**(6).

66. Khosravi-Shahi, P., L. Cabezón-Gutiérrez, and M.I. Aparicio Salcedo, *State of art of advanced triple negative breast cancer.* Breast J, 2019. **25**(5): p. 967-970.

67. McCann, K.E., S.A. Hurvitz, and N. McAndrew, *Advances in Targeted Therapies for Triple-Negative Breast Cancer.* Drugs, 2019. **79**(11): p. 1217-1230.

68. Geenen, J.J.J., et al., *PARP Inhibitors in the Treatment of Triple-Negative Breast Cancer.* Clin Pharmacokinet, 2018. **57**(4): p. 427-437.

69. Rakha, E.A., et al., *Breast cancer prognostic classification in the molecular era: the role of histological grade.* Breast Cancer Res, 2010. **12**(4): p. 207.

70. Paço, A. and R. Freitas, *HOX genes as transcriptional and epigenetic regulators during tumorigenesis and their value as therapeutic targets.* Epigenomics, 2019. **11**(13): p. 1539-1552.

71. Paço, A., et al., *Roles of the HOX Proteins in Cancer Invasion and Metastasis.* Cancers, 2020. **13**(1): p. 10.

72. de Bessa Garcia, S.A., et al., *HOX genes function in Breast Cancer development.* Biochim Biophys Acta Rev Cancer, 2020. **1873**(2): p. 188358.

73. Gaunt, S.J., *Hox cluster genes and collinearities throughout the tree of animal life.* Int J Dev Biol, 2018. **62**(11-12): p. 673-683.

74. Morgan, R., et al., *Targeting HOX/PBX dimers in cancer.* Oncotarget, 2017. **8**(19): p. 32322-32331.

75. Longobardi, E., et al., *Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates.* Dev Dyn, 2014. **243**(1): p. 59-75.

76. Penkov, D., et al., *Analysis of the DNA-Binding Profile and Function of TALE Homeoproteins Reveals Their Specialization and Specific Interactions with Hox Genes/Proteins.* Cell Reports, 2013. **3**(4): p. 1321-1333.

77. Bridoux, L., et al., *HOX paralogs selectively convert binding of ubiquitous transcription factors into tissue-specific patterns of enhancer activation.* PLOS Genetics, 2020. **16**(12): p. e1009162.

78. Krumlauf, R., *Hox genes, clusters and collinearity.* Int J Dev Biol, 2018. **62**(11-12): p. 659-663.

79. Shah, N. and S. Sukumar, *The Hox genes and their roles in oncogenesis.* Nature Reviews Cancer, 2010. **10**(5): p. 361-371.

80. Ando, H., et al., *Peptide-based inhibition of the HOXA9/PBX interaction retards the growth of human meningioma.* Cancer Chemother Pharmacol, 2014. **73**(1): p. 53-60.

81. Garcia, S.A.d.B., M. Araújo, and R. Freitas, *Dataset of HOXB7, HOXB8 and HOXB9 expression profiles in cell lines representative of the breast cancer molecular subtypes Luminal a (MCF7), Luminal b (BT474), HER2+ (SKBR3) and triple-negative (MDA231, MDA468), compared to a model of normal cells (MCF10A).* Data in brief, 2020. **30**: p. 105572-105572.

82. Morgan, R., et al., *Targeting HOX transcription factors in prostate cancer.* BMC Urology, 2014. **14**(1): p. 17.

83. Platais, C., et al., *Targeting HOX-PBX interactions causes death in oral potentially malignant and squamous carcinoma cells but not normal oral keratinocytes.* BMC cancer, 2018. **18**(1): p. 723-723.

84. Vinnitsky, V., *The development of a malignant tumor is due to a desperate asexual self-cloning process in which cancer stem cells develop the ability to mimic the genetic program of germline cells.* Intrinsically disordered proteins, 2014. **2**(1): p. e29997-e29997.

85. Jonkers, J., P. Pai, and S. Sukumar, *Multiple roles of HOX proteins in Metastasis: Let me count the ways.* Cancer Metastasis Rev, 2020. **39**(3): p. 661-679.

86. Abate-Shen, C., *Deregulated homeobox gene expression in cancer: cause or consequence?* Nat Rev Cancer, 2002. **2**(10): p. 777-85.

87. Rodrigues, M.F.S.D., et al., *Methylation status of homeobox genes in common human cancers.* Genomics, 2016. **108**(5): p. 185-193.

88. Duan, R., et al., *HOXA13 is a potential GBM diagnostic marker and promotes glioma invasion by activating the Wnt and TGF-β pathways.* Oncotarget, 2015. **6**(29).

89. Jin, K. and S. Sukumar, *HOX genes: Major actors in resistance to selective endocrine response modifiers.* Biochim Biophys Acta, 2016. **1865**(2): p. 105-10.

90. Siddikuzzaman, C. Guruvayoorappan, and V.M. Berlin Grace, *All trans retinoic acid and cancer.* Immunopharmacol Immunotoxicol, 2011. **33**(2): p. 241-9.

91. Dou, R., et al., *Vitamin D and colorectal cancer: molecular, epidemiological and clinical evidence.* Br J Nutr, 2016. **115**(9): p. 1643-60.

92. Crawford, E.D., et al., *Androgen-targeted therapy in men with prostate cancer: evolving practice and future considerations.* Prostate Cancer Prostatic Dis, 2019. **22**(1): p. 24-38.

93. Waks, A.G. and E.P. Winer, *Breast Cancer Treatment: A Review.* Jama, 2019. **321**(3): p. 288-300.

94. Nagini, S., *Breast Cancer: Current Molecular Therapeutic Targets and New Players.* Anticancer Agents Med Chem, 2017. **17**(2): p. 152-163.

95. Watson, C.J. and W.T. Khaled, *Mammary development in the embryo and adult: a journey of morphogenesis and commitment.* Development, 2008. **135**(6): p. 995-1003.

96. Chen, H. and S. Sukumar, *Role of homeobox genes in normal mammary gland development and breast tumorigenesis.* J Mammary Gland Biol Neoplasia, 2003. **8**(2): p. 159-75.

97. Lewis, M.T., *Homeobox genes in mammary gland development and neoplasia.* Breast Cancer Res, 2000. **2**(3): p. 158-69.

98. Schep, R., et al., *Control of Hoxd gene transcription in the mammary bud by hijacking a preexisting regulatory landscape.* Proceedings of the National Academy of Sciences, 2016. **113**(48): p. E7720-E7729.

99. Chen, F. and M.R. Capecchi, *Paralogous mouse Hox genes, Hoxa9, Hoxb9, and Hoxd9, function together to control development of the mammary gland in response to pregnancy.* Proceedings of the National Academy of Sciences, 1999. **96**(2): p. 541.

100. Garin, É., et al., *Stromal Hoxa5 function controls the growth and differentiation of mammary alveolar epithelium.* Developmental Dynamics, 2006. **235**(7): p. 1858-1871.

101. Schep, R., et al., *Control of Hoxd gene transcription in the mammary bud by hijacking a preexisting regulatory landscape.* Proceedings of the National Academy of Sciences, 2016. **113**(48): p. E7720.

102. Hur, H., et al., *Analysis of HOX Gene Expression Patterns in Human Breast Cancer.* Molecular Biotechnology, 2014. **56**(1): p. 64-71.

103. Teo, W.W., et al., *HOXA5 determines cell fate transition and impedes tumor initiation and progression in breast cancer through regulation of E-cadherin and CD24.* Oncogene, 2016. **35**(42): p. 5539-5551.

104. Reinholz, M.M., et al., *Breast cancer and aneusomy 17: implications for carcinogenesis and therapeutic response.* Lancet Oncol, 2009. **10**(3): p. 267-77.

105. Errico, M.C., et al., *The Widening Sphere of Influence of HOXB7 in Solid Tumors.* Cancer Research, 2016. **76**(10): p. 2857.

106. Wu, X., et al., *HOXB7, a homeodomain protein, is overexpressed in breast cancer and confers epithelial-mesenchymal transition.* Cancer Res, 2006. **66**(19): p. 9527-34.

107. Jin, K., et al., *The HOXB7 protein renders breast cancer cells resistant to tamoxifen through activation of the EGFR pathway.* Proc Natl Acad Sci U S A, 2012. **109**(8): p. 2736-41.

108. Wu, X., et al., *ADP ribosylation by PARP-1 suppresses HOXB7 transcriptional activity.* PloS one, 2012. **7**(7): p. e40644-e40644.

109. Carè, A., et al., *HOXB7: a key factor for tumor-associated angiogenic switch.* Cancer Res, 2001. **61**(17): p. 6532-9.

110. Liu, S., et al., *HOXB7 promotes malignant progression by activating the TGFβ signaling pathway.* Cancer Res, 2015. **75**(4): p. 709-19.

111. Cantile, M., et al., *In vivo expression of the whole HOX gene network in human breast cancer.* Eur J Cancer, 2003. **39**(2): p. 257-64.

112. Ding, W.J., et al., *HOXB8 promotes tumor metastasis and the epithelial-mesenchymal transition via ZEB2 targets in gastric cancer.* J Cancer Res Clin Oncol, 2017. **143**(3): p. 385-397.

113. Li, R., et al., *MicroRNA-128/homeobox B8 axis regulates ovarian cancer cell progression.* Basic Clin Pharmacol Toxicol, 2019. **125**(6): p. 499-507.

114. Li, X., et al., *Knock-Down of HOXB8 Prohibits Proliferation and Migration of Colorectal Cancer Cells via Wnt/β-Catenin Signaling Pathway.* Med Sci Monit, 2019. **25**: p. 711-720.

115. Stavnes, H.T., et al., *HOXB8 expression in ovarian serous carcinoma effusions is associated with shorter survival.* Gynecol Oncol, 2013. **129**(2): p. 358-63.

116. Wang, T., et al., *HOXB8 enhances the proliferation and metastasis of colorectal cancer cells by promoting EMT via STAT3 activation.* Cancer Cell Int, 2019. **19**: p. 3.

117. Ying, Y., et al., *Oncogenic HOXB8 is driven by MYC-regulated super-enhancer and potentiates colorectal cancer invasiveness via BACH1.* Oncogene, 2020. **39**(5): p. 1004-1017.

118. Seki, H., et al., *HOXB9 Expression Promoting Tumor Cell Proliferation and Angiogenesis Is Associated with Clinical Outcomes in Breast Cancer Patients.* Annals of Surgical Oncology, 2012. **19**(6): p. 1831-1840.

119. Hayashida, T., et al., *HOXB9, a gene overexpressed in breast cancer, promotes tumorigenicity and lung metastasis.* Proceedings of the National Academy of Sciences, 2010. **107**(3): p. 1100.

120. Chiba, N., et al., *Homeobox B9 induces epithelial-to-mesenchymal transition-associated radioresistance by accelerating DNA damage responses.* Proc Natl Acad Sci U S A, 2012. **109**(8): p. 2760-5.

121. Shrestha, B., et al., *Homeodomain-containing protein HOXB9 regulates expression of growth and angiogenic factors, facilitates tumor growth in vitro and is overexpressed in breast cancer tissue.* Febs j, 2012. **279**(19): p. 3715-3726.

122. Paço, A., S.A. de Bessa Garcia, and R. Freitas, *Methylation in HOX Clusters and Its Applications in Cancer Therapy.* Cells, 2020. **9**(7).

123. Park, S.-M., et al., *A long-range interactive DNA methylation marker panel for the promoters of HOXA9 and HOXA10 predicts survival in breast cancer patients.* Clinical epigenetics, 2017. **9**: p. 73-73.

124. Mahmood, N. and S.A. Rabbani, *DNA Methylation Readers and Cancer: Mechanistic and Therapeutic Applications.* Frontiers in Oncology, 2019. **9**(489).

125. Park, S.Y., et al., *Promoter CpG island hypermethylation during breast cancer progression.* Virchows Arch, 2011. **458**(1): p. 73-84.

126. Widschwendter, M., et al., *Association of Breast Cancer DNA Methylation Profiles with Hormone Receptor Status and Response to Tamoxifen.* Cancer Research, 2004. **64**(11): p. 3807.

127. Gampenrieder, S.P., et al., *DNA Methylation Signatures Predicting Bevacizumab Efficacy in Metastatic Breast Cancer.* Theranostics, 2018. **8**(8): p. 2278-2288.

128. Chen, Y., et al., *Lung Cancer Therapy Targeting Histone Methylation: Opportunities and Challenges.* Computational and Structural Biotechnology Journal, 2018. **16**: p. 211-223.

129. Ganesan, A., et al., *Epigenetic Therapy: Histone Acetylation, DNA Methylation and Anti-Cancer Drug Discovery.* Current Cancer Drug Targets, 2009. **9**(8): p. 963-981.

130. Ansari, K.I., et al., *HOXC6 Is transcriptionally regulated via coordination of MLL histone methylase and estrogen receptor in an estrogen environment.* J Mol Biol, 2011. **411**(2): p. 334-49.

131. Kikuyama, M., et al., *Development of a novel approach, the epigenome-based outlier approach, to identify tumor-suppressor genes silenced by aberrant DNA methylation.* Cancer Letters, 2012. **322**(2): p. 204-212.

132. Pilato, B., et al., *HOX gene methylation status analysis in patients with hereditary breast cancer.* Journal of Human Genetics, 2013. **58**(1): p. 51-53.

133. Shan, M., et al., *DNA Methylation Profiles and Their Diagnostic Utility in BC.* Disease Markers, 2019. **2019**: p. 6328503.

134. Xia, B., et al., *Homeobox A11 hypermethylation indicates unfavorable prognosis in breast cancer.* Oncotarget, 2017. **8**(6): p. 9794-9805.

135. Zhong, Z., et al., *HOXD13 methylation status is a prognostic indicator in breast cancer.* Int J Clin Exp Pathol, 2015. **8**(9): p. 10716-24.

136. Raman, V., et al., *Compromised HOXA5 function can limit p53 expression in human breast tumours.* Nature, 2000. **405**(6789): p. 974-8.

137. Ansari, K.I., et al., *HOXC10 is overexpressed in breast cancer and transcriptionally regulated by estrogen via involvement of histone methylases MLL3 and MLL4.* J Mol Endocrinol, 2012. **48**(1): p. 61-75.

138. Van Tongelen, A., A. Loriot, and C. De Smet, *Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes.* Cancer Lett, 2017. **396**: p. 130-137.

139. Kim, J., et al., *HOXC10 overexpression promotes cell proliferation and migration in gastric cancer.* Oncol Rep, 2019. **42**(1): p. 202-212.

140. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing.* Science, 2002. **298**(5595): p. 1039-43.

141. Zhang, E.B., et al., *P53-regulated long non-coding RNA TUG1 affects cell proliferation in human non-small cell lung cancer, partly through epigenetically regulating HOXB7 expression.* Cell Death Dis, 2014. **5**(5): p. e1243.

142. Marcinkiewicz, K.M. and L.J. Gudas, *Altered histone mark deposition and DNA methylation at homeobox genes in human oral squamous cell carcinoma.* J Cell Physiol, 2014. **229**(10): p. 1405-16.

143. Chariot, A., et al., *CBP and histone deacetylase inhibition enhance the transactivation potential of the HOXB7 homeodomain-containing protein.* Oncogene, 1999. **18**(27): p. 4007-14.

144. Palakurthy, R.K., et al., *Epigenetic silencing of the RASSF1A tumor suppressor gene through HOXB3-mediated induction of DNMT3B expression.* Molecular cell, 2009. **36**(2): p. 219-230.

145. de Bessa Garcia, S.A., et al., *HOXB7 Overexpression Leads Triple-Negative Breast Cancer Cells to a Less Aggressive Phenotype.* Biomedicines, 2021. **9**(5): p. 515.

146. Ciechomska, I.A., et al., *BIX01294, an inhibitor of histone methyltransferase, induces autophagy-dependent differentiation of glioma stem-like cells.* Scientific Reports, 2016. **6**(1): p. 38723.

147. Yan, N., et al., *GSKJ4, an H3K27me3 demethylase inhibitor, effectively suppresses the breast cancer stem cells.* Exp Cell Res, 2017. **359**(2): p. 405-414.

148. Morgan, R., et al., *Antagonism of HOX/PBX dimer formation blocks the in vivo proliferation of melanoma.* Cancer Res, 2007. **67**(12): p. 5806-13.

149. Holliday, D.L. and V. Speirs, *Choosing the right cell line for breast cancer research.* Breast Cancer Res, 2011. **13**(4): p. 215.

150. Neve, R.M., et al., *A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes.* Cancer cell, 2006. **10**(6): p. 515-527.

151. Comşa, Ş., A.M. Cîmpean, and M. Raica, *The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research.* Anticancer Res, 2015. **35**(6): p. 3147-54.

152. Subik, K., et al., *The Expression Patterns of ER, PR, HER2, CK5/6, EGFR, Ki-67 and AR by Immunohistochemical Analysis in Breast Cancer Cell Lines.* Breast cancer : basic and clinical research, 2010. **4**: p. 35-41.

153. Ma, R., et al., *HOXB7-S3 inhibits the proliferation and invasion of MCF-7 human breast cancer cells.* Molecular Medicine Reports, 2015. **12**(4): p. 4901-4908.

154. Li, T., et al., *Histone acetyltransferase p300 regulates the expression of human pituitary tumor transforming gene (hPTTG).* Journal of Genetics and Genomics, 2009. **36**(6): p. 335-342.

155. Anzai, E., et al., *FOXA1 Induces E-Cadherin Expression at the Protein Level via Suppression of Slug in Epithelial Breast Cancer Cells.* Biol Pharm Bull, 2017. **40**(9): p. 1483-1489.

156. Brank, A.S., et al., *Inhibition of HhaI DNA (Cytosine-C5) Methyltransferase by Oligodeoxyribonucleotides Containing 5-Aza-2'-deoxycytidine: Examination of the Intertwined Roles of Co-factor, Target, Transition State Structure and Enzyme Conformation.* Journal of Molecular Biology, 2002. **323**(1): p. 53-67.

157. Christman, J.K., *5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy.* Oncogene, 2002. **21**(35): p. 5483-5495.

158. Le, A.V., et al., *DNA Methylation Profiling of Breast Cancer Cell Lines along the Epithelial Mesenchymal Spectrum-Implications for the Choice of Circulating Tumour DNA Methylation Markers.* Int J Mol Sci, 2018. **19**(9).

159. Lombaerts, M., et al., *E-cadherin transcriptional downregulation by promoter methylation but not mutation is related to epithelial-to-mesenchymal transition in breast cancer cell lines.* Br J Cancer, 2006. **94**(5): p. 661-71.

160. Cui, S.N., et al., *Trichostatin A modulates the macrophage phenotype by enhancing autophagy to reduce inflammation during polymicrobial sepsis.* Int Immunopharmacol, 2019. **77**: p. 105973.

161. Tang, D., et al., *Trichostatin A reverses the chemoresistance of lung cancer with high IGFBP2 expression through enhancing autophagy.* Sci Rep, 2018. **8**(1): p. 3917.

162. Zhu, S., et al., *TSA-induced JMJD2B downregulation is associated with cyclin B1-dependent survivin degradation and apoptosis in LNCap cells.* J Cell Biochem, 2012. **113**(7): p. 2375-82.

163. Kaur, M., et al., *Oxindole: A chemical prism carrying plethora of therapeutic benefits.* Eur J Med Chem, 2016. **123**: p. 858-894.

164. Marques, C.S. and A.J. Burke, *Enantioselective Rhodium(I)-Catalyzed Additions of Arylboronic Acids to N-1,2,3-Triazole-Isatin Derivatives: Accessing N-(1,2,3-Triazolmethyl)-3-hydroxy-3-aryloxindoles.* ChemCatChem, 2016. **8**(22): p. 3518-3526.

165. Guo, H. and Q.P. Diao, *The Anti-Breast Cancer Potential of Bis-Isatin Scaffolds.* Curr Top Med Chem, 2020. **20**(16): p. 1499-1503.

166. Gupta, A.K., et al., *Systematic Review on Cytotoxic and Anticancer Potential of N-Substituted Isatins as Novel Class of Compounds Useful in Multidrug-Resistant Cancer Therapy: In Silico and In Vitro Analysis.* Top Curr Chem (Cham), 2019. **377**(3): p. 15.

167. Hu, J., et al., *Human Embryonic Kidney 293 Cells: A Vehicle for Biopharmaceutical Manufacturing, Structural Biology, and Electrophysiology.* Cells Tissues Organs, 2018. **205**(1): p. 1-8.

168. Ehrlich, M., *DNA hypomethylation in cancer cells.* Epigenomics, 2009. **1**(2): p. 239-259.

169. Ehrlich, M. and M. Lacey, *DNA hypomethylation and hemimethylation in cancer.* Adv Exp Med Biol, 2013. **754**: p. 31-56.

170. Liu, H., et al., *Promoter methylation inhibits BRD7 expression in human nasopharyngeal carcinoma cells.* BMC Cancer, 2008. **8**: p. 253.

171. Pfeifer, G.P., *Defining Driver DNA Methylation Changes in Human Cancer.* International Journal of Molecular Sciences, 2018. **19**(4).

172. Pinnick, K.E. and F. Karpe, *DNA methylation of genes in adipose tissue.* Proceedings of the Nutrition Society, 2011. **70**(1): p. 57-63.

173. Meccia, E., et al., *HOXB7 expression is regulated by the transcription factors NF-Y, YY1, Sp1 and USF-1.* Biochim Biophys Acta, 2003. **1626**(1-3): p. 1-9.

174. Mossman, D., K.-T. Kim, and R.J. Scott, *Demethylation by 5-aza-2'-deoxycytidine in colorectal cancer cells targets genomic DNA whilst promoter CpG island methylation persists.* BMC cancer, 2010. **10**: p. 366-366.

175. Tachibana, K., K. Takeda, and M. Shiraishi, *5-Aza-2'-deoxycytidine reactivates the CDH1 gene without influencing methylation of the entire CpG island or histone modification in a human cancer cell line.* Proceedings of the Japan Academy, Series B, 2004. **80**(7): p. 342-348.

176. Wilson, B.G. and C.W. Roberts, *SWI/SNF nucleosome remodellers and cancer.* Nat Rev Cancer, 2011. **11**(7): p. 481-92.

177. Kassis, J.A., J.A. Kennison, and J.W. Tamkun, *Polycomb and Trithorax Group Genes in <em>Drosophila</em&gt.* Genetics, 2017. **206**(4): p. 1699.

178. Lusser, A. and J.T. Kadonaga, *Chromatin remodeling by ATP-dependent molecular machines.* Bioessays, 2003. **25**(12): p. 1192-200.

179. Nguyen, C.T., et al., *Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2'-deoxycytidine.* Cancer Res, 2002. **62**(22): p. 6456-61.

180. Godbey, W.T., *Chapter 4 - Genes: The Blueprints for Proteins*, in *An Introduction to Biotechnology*, W.T. Godbey, Editor. 2014, Woodhead Publishing. p. 65-105.

181. Lai, J., et al., *TAp73 and ΔNp73 have opposing roles in 5-aza-2'-deoxycytidine-induced apoptosis in breast cancer cells.* Mol Cells, 2014. **37**(8): p. 605-12.

182. Greville, G., et al., *5-AZA-dC induces epigenetic changes associated with modified glycosylation of secreted glycoproteins and increased EMT and migration in chemo-sensitive cancer cells.* Clinical Epigenetics, 2021. **13**(1): p. 34.

183. Stich, M., et al., *5-aza-2'-deoxycytidine (DAC) treatment downregulates the HPV E6 and E7 oncogene expression and blocks neoplastic growth of HPV-associated cancer cells.* Oncotarget, 2016. **8**(32).

184. Hooiveld, M.H., et al., *Novel interactions between vertebrate Hox genes.* Int J Dev Biol, 1999. **43**(7): p. 665-74.

185. Beccari, L., et al., *A role for HOX13 proteins in the regulatory switch between TADs at the HoxD locus.* Genes & development, 2016. **30**(10): p. 1172-1186.

186. Favier, B., et al., *Functional cooperation between the non-paralogous genes Hoxa-10 and Hoxd-11 in the developing forelimb and axial skeleton.* Development, 1996. **122**(2): p. 449-60.

187. Mann, R.S., K.M. Lelli, and R. Joshi, *Hox specificity unique roles for cofactors and collaborators.* Current topics in developmental biology, 2009. **88**: p. 63-101.

188. Bijl, J., et al., *Analysis of HSC activity and compensatory Hox gene expression profile in Hoxb cluster mutant fetal liver cells.* Blood, 2006. **108**(1): p. 116-22.

189. Sotillos, S., M. Aguilar, and J.C. Hombría, *Forces shaping a Hox morphogenetic gene network.* Proc Natl Acad Sci U S A, 2013. **110**(11): p. 4303-8.

190. Fabre, P.J., et al., *Heterogeneous combinatorial expression of Hoxd genes in single cells during limb development.* BMC Biol, 2018. **16**(1): p. 101.

191. Navas, T., et al., *Clinical Evolution of Epithelial-Mesenchymal Transition in Human Carcinomas.* Cancer Res, 2020. **80**(2): p. 304-318.

192. Tian, X., et al., *E-cadherin/β-catenin complex and the epithelial barrier.* J Biomed Biotechnol, 2011. **2011**: p. 567305.

193. Liu, J., et al., *CDH1 promoter methylation correlates with decreased gene expression and poor prognosis in patients with breast cancer.* Oncol Lett, 2016. **11**(4): p. 2635-2643.

194. Lopes, N., et al., *1Alpha,25-dihydroxyvitamin D3 induces de novo E-cadherin expression in triple-negative breast cancer cells by CDH1-promoter demethylation.* Anticancer Res, 2012. **32**(1): p. 249-57.

195. Gallinari, P., et al., *HDACs, histone deacetylation and gene transcription: from molecular biology to cancer therapeutics.* Cell Research, 2007. **17**(3): p. 195-211.

196. Song, R., T. Van Buren, and I.V. Yosypiv, *Histone deacetylases are critical regulators of the renin-angiotensin system during ureteric bud branching morphogenesis.* Pediatr Res, 2010. **67**(6): p. 573-8.

197. Rousseaux, S. and S. Khochbin, *Histone Acylation beyond Acetylation: Terra Incognita in Chromatin Biology.* Cell journal, 2015. **17**(1): p. 1-6.

198. Sun, X.-J., et al., *The Role of Histone Acetyltransferases in Normal and Malignant Hematopoiesis.* Frontiers in Oncology, 2015. **5**(108).

199. Al-Rayyan, N., et al., *5-Aza-2-deoxycytidine and trichostatin A increase COUP-TFII expression in antiestrogen-resistant breast cancer cell lines.* Cancer Lett, 2014. **347**(1): p. 139-50.

200. Song, X., et al., *Trichostatin A inhibits proliferation of triple negative breast cancer cells by inducing cell cycle arrest and apoptosis.* Neoplasma, 2018. **65**(6): p. 898-906.

201. Wang, W.-M., et al., *HOXB7 promotes tumor progression via bFGF-induced activation of MAPK/ERK pathway and indicated poor prognosis in hepatocellular carcinoma.* Oncotarget, 2017. **8**(29): p. 47121-47135.

202. Aleskandarany, M.A., et al., *Tumour Heterogeneity of Breast Cancer: From Morphology to Personalised Medicine.* Pathobiology, 2018. **85**(1-2): p. 23-34.

203. Taneja, P., et al., *Classical and Novel Prognostic Markers for Breast Cancer and their Clinical Significance.* Clin Med Insights Oncol, 2010. **4**: p. 15-34.

204. Qu, Y., et al., *Evaluation of MCF10A as a Reliable Model for Normal Human Mammary Epithelial Cells.* PLOS ONE, 2015. **10**(7): p. e0131285.

205. Attar, N. and S.K. Kurdistani, *Exploitation of EP300 and CREBBP Lysine Acetyltransferases by Cancer.* Cold Spring Harb Perspect Med, 2017. **7**(3).

206. Adibekian, A., et al., *Characterization of a Selective, Reversible Inhibitor of Lysophospholipase 2 (LYPLA2)*, in *Probe Reports from the NIH Molecular Libraries Program*. 2010, National Center for Biotechnology Information (US): Bethesda (MD).

207. Mohammed, A., et al., *Inhibition of cell proliferation and migration in non-small cell lung cancer cells through the suppression of LYPLA1.* Oncol Rep, 2019. **41**(2): p. 973-980.

208. Iyer, N.G., H. Ozdag, and C. Caldas, *p300/CBP and cancer.* Oncogene, 2004. **23**(24): p. 4225-31.

209. Dutta, R., B. Tiu, and K.M. Sakamoto, *CBP/p300 acetyltransferase activity in hematologic malignancies.* Mol Genet Metab, 2016. **119**(1-2): p. 37-43.

210. Ono, H., M.D. Basson, and H. Ito, *P300 inhibition enhances gemcitabine-induced apoptosis of pancreatic cancer.* Oncotarget, 2016. **7**(32): p. 51301-51310.

211. He, Z.X., et al., *Current development of CBP/p300 inhibitors in the last decade.* Eur J Med Chem, 2021. **209**: p. 112861.

212. Hügle, M., et al., *Beyond the BET Family: Targeting CBP/p300 with 4-Acyl Pyrroles.* Angewandte Chemie International Edition, 2017. **56**(41): p. 12476-12480.

213. Chan, C.-H., et al., *Deciphering the transcriptional complex critical for RhoA gene expression and cancer metastasis.* Nature Cell Biology, 2010. **12**(5): p. 457-467.

214. Font de Mora, J. and M. Brown, *AIB1 Is a Conduit for Kinase-Mediated Growth Factor Signaling to the Estrogen Receptor.* Molecular and cellular biology, 2000. **20**: p. 5041-7.

215. Peng, J., et al., *p300/CBP inhibitor A-485 alleviates acute liver injury by regulating macrophage activation and polarization.* Theranostics, 2019. **9**(26): p. 8344-8361.

216. Tomatis, V.M., et al., *Acyl-protein thioesterase 2 catalyzes the deacylation of peripheral membrane-associated GAP-43.* PLoS One, 2010. **5**(11): p. e15045.

217. Ko, P.-J. and S.J. Dixon, *Protein palmitoylation and cancer.* EMBO reports, 2018. **19**(10): p. e46666.

218. Dekker, F.J., et al., *Small-molecule inhibition of APT1 affects Ras localization and signaling.* Nat Chem Biol, 2010. **6**(6): p. 449-56.

219. Zimmermann, T.J., et al., *Boron-based inhibitors of acyl protein thioesterases 1 and 2.* Chembiochem, 2013. **14**(1): p. 115-22.

220. Bürger, M., B.C. Willige, and J. Chory, *A hydrophobic anchor mechanism defines a deacetylase family that suppresses host response against YopJ effectors.* Nat Commun, 2017. **8**(1): p. 2201.

221. Adlanmerini, M., et al., *Mutation of the palmitoylation site of estrogen receptor α in vivo reveals tissue-specific roles for membrane versus nuclear actions.* Proceedings of the National Academy of Sciences, 2014. **111**(2): p. E283-E290.

8. SUPPLEMENTARY TABLES

Supplementary Table 1- Calculated IC_{50} of the target compounds against DOHH2, VAL, OCI-LY-10, SU-DHL-2, SW480, A549 cell lines using the MTT assay. Enantiomeric purity was Determined using HPLC with a chiral column, and is presented in percentage of enantiomeric excess (%ee). IC_{50} values calculated from MTT data using the *GraphPad Prism* Software Inc. (version 8.0.1), and are presented in µM ± 95% Confidence Level Intervals (CI).

Compound	Enantiomeric purity (%ee)	IC_{50}/µM					
		DOHH2	VAL	OCI-LY-10	SU-DHL-2	SW480	A549
(1a)	-	>20	>20	10	0.75	32.2 ± 2.9	54.7 ± 2.1
(1b)	>98	>20	>20	7	15	8.9 ± 0.8	53.8 ± 1.6
(1c)	-	10	>20	5	0.75	26.6 ± 2.3	41.5 ± 2.3
(S)-(2b)	86	-	-	-	-	9.9 ± 1.6	28.3 ± 1.5
(R)-(2j)	92	>20	>20	>20	>20	28.7 ± 1.8	65.9 ± 2.8