DNA signatures as a predictor of breast cancer risk

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Medicine by

Roshan Koron

October 2018

Disclaimer

This thesis is the result of my own work. The material contained in the thesis has not been presented, nor is currently being presented, either wholly or in part for any other degree or other qualification. The research was performed in the Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, United Kingdom. All other parties involved in the research presented here, and the nature of their contribution, are listed in the 'Acknowledgements' section of this thesis.

Abstract

Introduction

Breast cancer is the most common cancer in the UK with around 54,900 new cases diagnosed each year and 11,400 breast cancer deaths reported in 2016. Early diagnosis of breast cancer and personalised treatment strategies lead to a more favourable outcome. Biomarkers are essential for disease management in terms of risk calculation, response to treatment and disease outcome.

The aim of this study has been to assess whether two variable number tandem repeats (VNTRs) may help to modify prediction of breast cancer risk in patients with pathogenic germline variants in the *BRCA1* and *BRCA2* tumour suppressor genes. I therefore explored the polymorphism of the AAAG VNTR in the 5′ untranslated region (UTR) of the oestrogen related receptor γ (*ESRRG*) gene and a 15 bp VNTR located in the internal promoter region of the *MIR137* gene as both genes are implicated in breast cancer. Reporter gene analysis was performed on the *ESRRG* AAAG VNTR to observe the effects of varying repeat lengths on differential expression of the gene under basal conditions and following exposure to oestrogen in a luciferase reporter gene assay. VNTRs are often regulatory in nature therefore they may show insight into pathways underlying cancer.

Method

The *ESRRG* AAAG and *MIR137* VNTRs were explored by genotyping analysis on germline DNA in a female breast cancer cohort containing BRCA1/2 positive, BRCA1/2 negative breast cancer patients, and a group of healthy controls. The variance of *ESRRG* AAAG VNTR copy number and its effect on gene transcription with oestrogen stimulation was assessed in a reporter gene assay in a breast cancer cell line (MCF7).

Results

MIR137: *MIR137* VNTR analysis revealed a significant difference in repeat distribution between *BRCA* positive and *BRCA* negative patients and was significantly different between *BRACA1* and *BRACA2* germline mutation carriers using clump statistical analysis.

ESRRG AAAG VNTR: nine different allele variants of the AAAG repeats were observed. The longer copy numbers were more prevalent in the cancer free control group whereas the shorter copy numbers were associated with BRCA1/2 positive cases. Nine copies of the AAAG VNTR showed significantly enhanced reporter gene activity with oestrogen exposure in MCF-7 breast cancer cells.

Conclusion

Significant differences in *MIR137* VNTR length were found between *BRACA1* and *BRACA2* germline mutation carriers and wild type controls. The *ESRRG* AAAG VNTR showed significant distribution differences between the BRCA1/2 positive and control groups.

An increase in ESRRG transcription on exposure to oestrogen in the presence of 9 copies of the VNTR may relate to an increased cancer risk. This may be influenced by environmental factors which contribute to GxE interactions in cancer.

Our findings support the role of the *ESRRG* AAAG VNTRs as a potential biomarker in the prediction of breast cancer risk in *BRCA1/2* pathogenic germline variant carriers and may also add information to genetic risk scores which can further stratify treatment and surveillance in high risk individuals. Although significant differences were found between groups in *MIR137* VNTR the sample sizes were small therefore these results should be interpreted with caution. This data should be extended and validated in a larger cohort.

Acknowledgements

I would like to thank Professor John Quinn for taking me on as a student in his lab and Dr Jill Bubb for her friendship. I am also grateful to Professor Varro for her support and guidance.

I would also like to express my gratitude to Professor Gareth Evans for supplying the patient samples and to Helen Byers for her contribution.

I would like to extend my thanks to all the lab staff and students for their guidance. I couldn't have done it without you. I would especially like to thank (soon to become Dr) Veridiana Pessoa for her emotional support through the highs and lows of research. I hope I have managed to reciprocate.

A special mention goes to my beautiful son Noah and wonderful husband, this project was completed for you both. Last but not least I would like to thank my wonderful family the Korons, Gigers and Whitakers. Thank you mum and dad for standing by me and taking an interest in everything that I do. I may not say it enough but I love you both and appreciate everything you do for me. I'm sorry for involving you in stress management. You can have the old me back now!

Table of contents

1 INTRODUCTION 9

1.1 Breast Cancer Demographics in the UK 9

1.2 Risk Factors for Breast Cancer 9
1.2.1 Modifiable Risk Factors 9
1.2.1.1 Oral Contraceptive Pill and Breast Cancer 9
1.2.1.2 Hormone replacement therapy and breast cancer 10
1.2.1.3 Breastfeeding and Breast Cancer 10
1.2.1.4 Physical activity and Breast Cancer 10
1.2.1.5 Smoking and Breast Cancer 11
1.2.1.6 Alcohol and Breast Cancer 11
1.2.1.7 Breast Cancer and Obesity 12
1.2.2 Non Modifiable Risk Factors 13
1.2.2.1 Aging, Menopause and Breast Cancer 13
1.2.2.2 Menarche, Pregnancies and Breast Cancer 13
1.2.2.3 Breast Density and Breast Cancer 14
1.2.2.4 Ionising Radiation and Breast Cancer 14
1.2.2.5 Oestrogen 14

1.3 Breast Cancer Screening 15

1.4 Family history clinic 16

1.5 The burden of being a BRCA pathogenic mutation carrier 17

1.6 Genetic Factors 18
1.6.1 Hereditary breast cancer syndromes 18
1.6.2 *BRCA1* and *BRCA2* Genes 19

1.7 Microsatellites / VNTRs 21
1.7.1 Microsatellites / VNTR in breast cancer 22

1.8 Single nucleotide polymorphisms 22

1.9 Oestrogen receptors and oestrogen-related receptors 24
1.9.1 Nuclear Receptors 24
1.9.2 Oestrogen receptors (ERs) 25
1.9.3 Oestrogen Related Receptor Gamma (ERRγ) 26
1.9.4 Interaction between Oestrogen Receptors and Oestrogen Related Receptors 27

1.10 Micro RNAs 28

2 MATERIALS AND METHODS 30

2.1 Ethics 30

2.2 Materials 30
2.2.1 Human DNA samples 30
2.2.2 Assessment of DNA quality by spectrophotometry 30
2.2.3 Human cell line 31
2.2.3.1 MCF-7 Cells 31
2.2.4 Commonly used buffers and cell culture reagents 31
2.2.4.1 LB Broth 31
2.2.4.2 LB Agar 31
2.2.4.3 5X TBE Buffer 32
2.2.5 Cell culture media 32
2.2.5.1 Media for MCF-7 cells (MCF-7 media) 32
2.2.5.2 Stripped Media for MCF-7 cells (MCF-7 stripped media (oestrogen free media)) 32

2.3 Genotyping 33
2.3.1 Polymerase chain reaction 33
2.3.1.1 Polymerase chain reaction for genotyping 33
2.3.1.2 PCR for AAAG VNTR in *ESRRG* 33
2.3.1.3 PCR for MIR137 VNTR 35
2.3.2 Agarose Gel Electrophoresis 36

2.4 Generation of reporter gene constructs 37
2.4.1 Agarose gel extraction of DNA fragments 38
2.4.2 Ligation of DNA fragments into pGL3P reporter gene vector 38
2.4.3 Transformation of DH5α competent E. coli cells 39
2.4.4 Miniprep of plasmid DNA 39
2.4.5 pGL3P Plasmid extraction 40
2.4.6 PCR confirmation of plasmid inserts: 41
2.4.7 Sequencing 41

2.5 Cell Culture (Eukaryotic cells) 42
2.5.1 Culturing of MCF-7 Cells 42
2.5.1.1 Cell counting 42
2.5.1.2 Freezing cells for storage 43
2.5.2 Analysis of reporter gene expression 43
2.5.2.1 Transient transfection of reporter gene constructs into MCF-7 cells 43
2.5.2.2 Transfection 44
2.5.2.3 Cell Lysis 44
2.5.2.4 Gene expression analysis by luciferase assay 45
2.5.2.5 Treatment of cells with oestrogen 45

2.6 Statistics 46
2.6.1 Clump statistical analysis 46
2.6.2 Correlation analysis 46

3 AAAG MICROSATELLITE IN THE *ESRRG* GENE AS A POTENTIAL BIOMARKER FOR PREDICTION OF BREAST CANCER RISK **47**

3.1 Introduction **47**

3.2 Aims **50**

3.3 Results **51**
3.3.1 Genotyping of AAAG microsatellite in ESRRG in a breast cancer cohort 51
3.3.2 Analysis of the reporter gene constructs containing varying lengths of the AAAG microsatellite under basal, vehicle (ethanol) and treatment (17-β-oestradiol) conditions in MCF-7 cells 57

3.4 Discussion **59**

4 MIR137 VNTR AS A POTENTIAL BIOMARKER FOR PREDICTION OF BREAST CANCER RISK **64**

4.1 Introduction **64**

4.2 Aims **70**

4.3 Results **70**
4.3.1 Four copy MIR137 VNTR homozygosity is less commonly observed in *BRCA1/2* patients 70
4.3.2 MIR137 VNTR modifies CpG island length *in silico* 77

4.4 Discussion **79**

5 FINAL DISCUSSION **82**

6 FIGURES AND TABLES **88**

6.1 Figures **88**

6.2 Tables **88**

7 REFERENCES **89**

1 Introduction

1.1 Breast Cancer Demographics in the UK

Breast cancer is a complex disease that is caused by abnormal growth and uncontrolled division of cells within the terminal duct and lobules of the breast. It occurs mostly in women but men can also be affected. Although there is an appreciable group of individuals with a hereditary predisposition for breast cancer, the majority are sporadic cases.

Breast cancer is the most common cancer in the UK affecting almost 54,800 women and 390 men each year and is the leading cause of death amongst British women. One in eight women develop breast cancer in the UK during their lifetime compared to one in 870 men (CRUK, 2017).

1.2 Risk Factors for Breast Cancer

The risk of developing cancer is multifactorial and associated with genetics, age and exposure to risk factors. A gene versus environment (GxE) approach should therefore be adopted when assessing risk factors for breast cancer prediction. Modifiable risks can be altered by changes in lifestyle. In the UK 27% of breast cancers in females are lifestyle related (CRUK, 2017).

1.2.1 Modifiable Risk Factors

1.2.1.1 Oral Contraceptive Pill and Breast Cancer

Breast cell proliferation increases during oral contraceptive (OC) use. Oestrogen which is also present in OCs triggers mitosis in epithelial cells in the breast. This may be a harmful side effect which could imply that the OC behaves as a tumour promoter (Isaksson et al., 2001). The exact mechanism of carcinogenesis with OC use is not clear. The current generation of OC preparations are linked with an increased breast cancer risk in an oestrogen and progesterone dose dependent manner. Women aged 20–49 years are

particularly vulnerable to it (Beaber et al., 2014). The risk does however diminish after OC use has been stopped with no apparent residual risk beyond 10 years following cessation (Collaborative Group on Hormonal Factors in Breast, 1996).

1.2.1.2 Hormone replacement therapy and breast cancer

The risk of postmenopausal breast cancer increases with the use of hormone replacement therapy (HRT). Combined forms of HRT containing both oestrogen and progestogen (a synthetic form of progesterone) rather than oestrogen alone further increase the risk. Duration of HRT use up to 15 years also increases the relative risk. This information was reported by Jones et al who conducted follow up on their investigated cohort and were also mindful of each woman's potential change in HRT history. Only women with a known age at menopause as opposed to inferred age were included in their study (Jones et al., 2016). Inferred age refers to women with artificially induced menopause such as after a hysterectomy rather than as a physiological progression. This is important as it eliminates uncertainties in hormone levels.

1.2.1.3 Breastfeeding and Breast Cancer

Women who have breastfed reduce their breast cancer risk by 22% compared with women who have no breastfeeding history. This risk is further reduced with longer periods of breastfeeding (Lambertini et al., 2016). Breastfeeding reduces ER/PR positive and triple-negative (ER-/PR-/HER2-) breast cancer risk by 23% and 21% respectively. It does not however seem to have an effect on HER2 positive breast cancer. The association between breastfeeding and cancer may be linked with the presence of lactation-induced amenorrhea which reduces exposure to endogenous sex hormones (Dontu et al., 2004).

1.2.1.4 Physical activity and Breast Cancer

In the UK 3% of all female breast cancers are attributed to inadequate physical activity defined as less than 150 minutes of moderate activity per week. Every

additional 2 hours per week of moderate vigorous activity can reduce the risk of breast cancer by 5% (Wu et al., 2013). This can be explained partly by the fact that high levels of physical activity lower the levels of sex hormones (oestradiol) resulting in lower risk of breast cancer (Parkin, 2011).

1.2.1.5 Smoking and Breast Cancer

A history of smoking is linked with an increase in breast cancer risk. The risk is present whether smoking was commenced prior to or post first full term pregnancy (FFTP). Although the risk is higher if smoking occurred pre pregnancy. Breast tissue prior to FFTP is undifferentiated and more sensitive to carcinogen exposure (Ponten et al., 1990). A higher number of pack years (unit of measurement of amount of cigarettes smoked by a person, calculated by multiplying the number of packs of cigarettes smoked per day by the number of years the person has smoked) also increased the risk of breast cancer (Catsburg et al., 2015).

1.2.1.6 Alcohol and Breast Cancer

Consumption of alcohol has been linked to breast cancer. In the UK 11% of all breast cancers in women are linked to alcohol (Allen et al., 2009). In the USA, the Nurses' Health Study reported that breast cancer risk increases by 7-10% with each unit of alcohol consumed per day and this increased to 15% when the intake increased to 4 to 9 units per week compared with women who had never consumed alcohol (Chen et al., 2011b). Both premenopausal and postmenopausal women who drink have an increased risk of breast cancer compared to teetotal women (Seitz HK, 2012). Alcohol consumption has been shown to moderately increase the levels of oestrogen (Rinaldi et al., 2006). In vitro studies have also revealed increased transcriptional activity of the oestrogen receptorα (ERα). This may alter sensitivity of breast tissue to oestrogen resulting in enhanced ER levels and proliferation in ER positive cell lines (Singletary et al., 2001).

A positive correlation has been made linking alcohol consumption with oestradiol, oestriol and oestrone levels. The presence of oestriol and oestrone have not been consistent between study groups but the elevation of oestradiol has been a common finding (Hartman et al., 2016; Hirko et al., 2014; Reichman et al., 1993). The mechanisms by which alcohol consumption may be linked to oestrogen levels and hence breast cancer risk however remain unclear.

Mechanisms suggested by which alcohol may increase oestrogen levels include the stimulation of oestrogen production (Rettori and McCann, 1997; Sierksma et al., 2004) or inhibition of oestrogen degradation (Eagon, 2010). Chronic alcohol consumption increases the activity of aromatase which is the enzyme responsible for converting testosterone to oestrogen. This results in a fall in testosterone levels and a rise in oestrogen concentration with alcohol consumption (Sierksma et al., 2004). Alcohol may also enhance oestrogen production by inhibition of melatonin secretion (Stevens et al., 2000) or by interfering with the pituitary gland's release of luteinizing hormone which is responsible for the release of oestradiol from the ovaries (Rettori and McCann, 1997). Alcohol on the other hand may also inhibit oestrogen degradation by inactivation of the enzymes sulfotransferase and 2-hydroxylase (Eagon, 2010).

1.2.1.7 Breast Cancer and Obesity

Twenty six percent of the UK population are obese, with an alarming increase in childhood obesity. This could translate into a future increase in obesity associated breast cancer risk. (NHS, 2018). There is an increased risk of breast cancer in postmenopausal women with an increased body mass index (BMI) who have never taken hormone replacement therapy. This is due to an increase in oestrogen exposure, particularly oestradiol (Key et al., 2003),(Nassir, 2012) This risk increases with increasing BMI in this group (Reeves et al., 2007).

1.2.2 Non Modifiable Risk Factors

1.2.2.1 Aging, Menopause and Breast Cancer

The risk of a woman developing cancer increases with age. Reaching the menopause after the age of 55 is associated with an increased risk of breast, ovarian, and uterine cancers. This risk increases further in women who reach early menarche (start of menstruation) before the age of 12 years. This extends oestrogen exposure increasing breast cancer risk (Surakasula et al., 2014).

1.2.2.2 Menarche, Pregnancies and Breast Cancer

Breast tissue undergoes rapid cellular proliferation between menarche and FFTP, and risk increases rapidly until final differentiation with first pregnancy (Colditz et al., 2014). During this time period the breast tissue is particularly vulnerable to carcinogenic stimuli, as already mentioned (1.2.1.5). Following differentiation cells become less sensitive to carcinogens. Breast cancer models show that risk is determined by early life events, therefore preventative strategies should be considered early (Colditz and Frazier, 1995). Menarche leads to the onset of the ovulatory cycles and oestrogen exposure leading to increased mammary cell proliferation. Early menarche activates proliferation at a younger age (Rosner et al., 1994). The risk of breast cancer therefore increases with a longer interval between menarche and FFTP (Li et al., 2008). This may lead to promotion and progression of breast cancer several years later (Ponten et al., 1990). This risk has a greater effect on postmenopausal breast cancer. However, the protective effect of multiparity is stronger on postmenopausal breast cancers (Clavel-Chapelon and Gerber, 2002). In a study, there was a 25% reduction in luminal breast cancer risk in parous women compared to a 15% increase in risk in woman with higher age at FFTP (Colditz and Frazier, 1995; Pike et al., 1983).

1.2.2.3 Breast Density and Breast Cancer

Breast density is genetically defined as shown in twin studies which reported that the variability in breast density was 53-63% (Boyd et al., 2002; Ursin et al., 2009). Higher breast density is linked with breast cancer. Dense breast tissue is composed of more connective and epithelial tissue than fat (McCormack and dos Santos Silva, 2006). Age, parity (number of viable pregnancies), and menopausal status account for around 20 to 30 % of causes of breast density variability (Vachon et al., 2015). A quarter of all single nucleotide polymorphisms (SNPs) also associated with breast cancer are linked to breast density. An example of this is the SNP ZNF365 which is associated with both mammographic density and breast cancer risk (Lindstrom et al., 2011; Odefrey et al., 2010).

1.2.2.4 Ionising Radiation and Breast Cancer

Radiation exposure (medical and natural (background) radiation) accounts for approximately 1% of female breast cancers in the UK (Parkin, 2011). Women who have received radiotherapy for breast cancer are at a 9-11% higher risk of developing breast cancer in the contralateral breast compared with women who have had surgery (Berrington de Gonzalez et al., 2010; Neta et al., 2012). Breast cancer risk is however small in women under the age of 75 years receiving diagnostic x-rays (0.1%) (Berrington de Gonzalez and Darby, 2004) and mammograms (0.03-0.06%) in screening programs (Berrington de Gonzalez et al., 2010; Neta et al., 2012).

1.2.2.5 Oestrogen

Oestrogens are natural hormones involved in the promotion and proliferation of normal and neoplastic breast tissue (Bonkhoff and Berges, 2009; Stingl, 2011).

There are 4 types of oestrogen: oestrone (E1); oestradiol (E2); oestriol (E3) and oestetrol (E4). Oestrone is found in most tissues of the body but is predominantly found in fat in post-menopausal women. It can be converted to

oestradiol and vice versa. Oestradiol is the most potent oestrogen and along with oestrone is associated with risk of breast cancer (EHBCC-Group et al., 2011; EHBCC-Group et al., 2013). There is less production of oestrone and oestradiol during pregnancy resulting in a protective effect on breast cancer risk. Oestetrol and oestriol (a degradation product of oestradiol) are only found during pregnancy (Ali et al., 2017; Liu et al., 2015).

The main oestrogen in women is oestradiol which is vital for the development and maintenance of the female reproductive system and sperm production in men. Oestradiol is also important in the physiology of bone, the cardiovascular, immune and central nervous systems. Alterations in oestrogen levels can lead to diseases such as breast and endometrial cancer, coronary heart disease, osteoporosis as well as a decline in cognitive function. The effects of oestradiol are converted by the nuclear oestrogen receptors alpha and beta (ERα and ERβ) (Giguere, 1999). In post-menopausal women oestrogen synthesis is taken over by the adrenal glands and adipose tissue once the ovaries stop producing oestrogen (Cuzick et al., 2011; Santen et al., 2009).

There will be further details on the interaction of oestrogen with the oestrogen receptor and oestrogen related receptors throughout the thesis.

1.3 Breast Cancer Screening

Early diagnosis of breast cancer is crucial as detection at an advanced stage leads to less favourable outcomes (Fisher et al., 2015). In the UK currently breast screening is conducted between the ages of 50 and 70 years. In some regions screening is performed between the ages of 47 years and 73 years as a trial. Women are invited to receive a mammogram every three years. Breast screening will pick up 95% of breast cancers and will detect lesions at a much earlier stage than clinical examination. One in 25 women will be called back for further investigation following screening and four of these women will be diagnosed with breast cancer. Almost 79% of these detected cancers are

invasive breast cancers, and 21% are in-situ or micro-invasive (BrSP, 2015; Public Health England, 2017a). Some women who may be considered to be at a higher risk of developing breast cancer may be screened at a younger age due to a family history of breast or ovarian cancer. This group are seen at family history clinics and besides a history and examination they will be screened with MRI scans or mammographic imaging (Public Health England, 2016).

1.4 Family history clinic

A hereditary predisposition for breast and ovarian cancer is suspected with the presentation of: a personal history of early-onset breast cancer (≤ 45 years of age); a personal history of triple-negative breast cancer (≤ 60 years of age, this breast cancer subtype is mostly associated with pathogenic *BRCA1* germline variants); a family history of first- or second-degree relatives with breast or ovarian cancers (or other cancers linked to breast or ovarian cancer genes (section 1.6.1); or a personal history of male breast cancer (Graffeo et al., 2016). Patients with these findings are referred to a family history clinic.

A study by Walsh et al. highlights the complexity of breast cancer genetic mutations even in increased risk groups such as the Ashkenazi Jewish population. In this group half the women with mutations in either of the two BRCA genes did not have an associated family history. This was thought to be as a result of the mutation being acquired by the father coupled with small family units culminating in a less obvious inherited mutation (Walsh et al., 2017). The incidence of male breast cancer is much lower than that of females even in male *BRCA1/2* mutation carriers (section 1.6.2) (Tai et al., 2007).

1.5 The burden of being a BRCA pathogenic mutation carrier

The detection of a germline *BRCA1/2* mutation leads to the following treatment options: surveillance, hormone therapy or prophylactic (risk reducing) mastectomy. The choice of risk reducing strategy is highly personal with several factors involved in reaching a decision. In 2017, Johns et al (Johns et al., 2017) explored the factors influencing the risk reducing decisions made by women who were found to have a *BRCA1/2* germline mutation. Younger women (below the age of 30 years) were more likely to opt for surveillance rather than surgery. This decision may have been influenced by the desire to breast feed or for the purpose of body image. A third of female *BRCA* mutation carriers chose risk reducing surgery over surveillance before the age of 50 years. This decision was strongly influenced by whether a relative had been diagnosed with breast cancer or who had died of the disease (Johns et al., 2017). A prophylactic mastectomy will reduce the risk of developing cancer but may also come with complications of surgery and psychological and emotional problems relating to body image. Post-operative complications following a mastectomy include pain, infection, a seroma (serous fluid collection at the site of surgery with the potential to become infected), bleeding, implant failure (if immediate reconstruction performed at time of mastectomy), excessive scarring and lymphoedema (swelling of the arm due to damage to the lymphatic drainage system) (Barton et al., 2005; West, 2004). On the other hand, choosing to be monitored for breast cancer means that the individual is living with the probability of developing breast cancer on a daily basis, which can also be a great psychological burden. The relative risk of developing breast cancer is highest between the ages of 30-60 years therefore this time period would result in the most benefit from risk-reducing mastectomy (Easton et al., 2015).

1.6 Genetic Factors

Five to ten percent of patients with breast cancer have an underlying genetic mutation (Couch et al., 2014). A common disease phenotype featuring within a family may be due to genetic mutations, however this may also reflect a common life-style and environment (Mavaddat et al., 2010). Hereditary breast cancer was first described by the French physician Pierre Paul Broca in 1866 based on personal involvement. Broca described the first pedigree of inheritable breast cancer when his wife developed breast cancer at a young age as did four generations of her family (Krush, 1979; Nielsen et al., 2016).

1.6.1 Hereditary breast cancer syndromes

Several genes are now known to be linked to defined hereditary breast cancer syndromes. These include Cowden (*PTEN*), Li-Fraumeni (*TP53*), Peutz-Jeghers (*STK11* mutation), Lynch (*MLH1, PMS2, MSH2, MSH6, PMS2*) (Roberts et al., 2018) and hereditary diffuse gastric cancer (*CDH1* mutations) syndromes (Berliner et al., 2013). A subgroup of Fanconi anaemia (D1) with a susceptibility to childhood tumours is associated with bi-allelic germline mutations in *BRCA2* (Howlett et al., 2002). Another high-risk gene which is associated with breast cancer is 'partner and localizer of BRCA2' *(PALB2*) which accounts for 1–2 % of inherited breast cancer cases. *PALB2* encodes a protein which is involved in BRCA2-mediated DNA repair by promotion of localisation and stability of BRCA2. Moderate risk genes which give a relative risk of breast cancer of 2-3 fold include *CHEK2* and Ataxia telangiectasia mutated (*ATM)*. They represent around 5% of inherited breast cancer cases and have not been adequately explored as potential risk predictors (Eccles et al., 2013). *CHEK2* is a tumour suppressor gene that encodes CHK2, a serine/threonine kinase. It plays a pivotal role in DNA repair, cell cycle regulation and apoptosis (Apostolou and Papasotiriou, 2017). An abnormal expression of CHEK2 may therefore result in breast cancer (Dufault et al., 2004). DNA double strand break leads to an activation of ATM which phosphorylates CHK2 at Thr68. This induces *CHEK2*

dimerization and activates it's enzymatic activity. CHK2 then induces DNA double-strand break repair by phosphorylating BRCA1 and BRCA2 (Bartkova et al., 2008; Magni et al., 2014).

Most of the pathogenic mutations of *CHEK2* consist of small deletions or insertions that result in translation of a truncated protein. Carriers of truncating mutations with a family history of breast cancer have a greater risk of developing breast cancer (Cybulski et al., 2011; Weischer et al., 2008). This risk is unaltered in individuals who are also *BRCA1* and *BRCA2* mutation carriers as BRCA1, BRCA2 and CHK2 have the same signaling pathways in common (Meijers-Heijboer et al., 2002). The incidence of these mutations are variable in different populations. In the Ashkenazi Jewish population three founder mutations have been reported (King et al., 2003). In view of the many genes involved in breast cancer multigene panel testing is an efficient and cost effective technique for testing many genes simultaneously (Girardi et al., 2018; Prapa et al., 2017).

1.6.2 *BRCA1* and *BRCA2* Genes

In 1994 the *BRCA1* gene (17q21, chromosome 17: base pairs 43,044,294 to 43,125,482) (Miki et al., 1994) was linked to hereditary breast and ovarian cancer and the link to *BRCA2* (13q12.3, chromosome 13: base pairs 32,315,479 to 32,399,671) was made the following year (Wooster et al., 1995). Mutations in *BRCA1* and *BRCA2* mutations are inherited in an autosomal dominant pattern (Easton, 1999). In patients carrying a germline mutation of *BRCA1/2* a 'second hit' in the non-mutant allele leads to a loss of a functional gene product as described in the Knudson double hit model (Knudson, 1971). Loss of heterozygosity (LOH) is the common pathomechanism underlying carcinogenesis in *BRCA1/2* mutation carriers (Osorio et al., 2002).

BRCA1 and *BRCA2* genes are tumour suppressor genes. BRCA1 and BRCA2 are involved in DNA double strand break repair through homologous recombination and therefore of pivotal importance in the maintenance of genomic stability (O'Donovan and Livingston, 2010). A loss of function is directly linked to carcinogenesis (Ahn et al., 2007). They are further involved in cell cycle checkpoint control and regulation of mitosis (Joukov et al., 2006). The risk of breast cancer increases when a germline mutation in the *BRCA* gene leads to a truncated or inactivated protein (Easton, 1999).

While *BRCA1/2* germline mutation carriers are at significantly increased risk for developing breast and/or ovarian cancer, penetrance is incomplete (Kuchenbaecker et al., 2017). Penetrance is highly variable between carriers of identical mutations which implies the involvement of genetic and environmental modifying risk factors (Milne and Antoniou, 2016).

Overall mutations in *BRCA1* and *BRCA2* account for 20 to 25% of all hereditary breast cancers but also significantly increase the risk of ovarian, prostate and pancreatic cancer as well as of malignant melanoma (Easton, 1999).

In women who are *BRCA1* and *BRACA2* mutation carriers 72% and 69% will develop breast cancer respectively by the age of 80 years (Kuchenbaecker et al., 2017). The lifetime risk of developing ovarian cancer in the general population is 1.3% however this risk increases to 44% and 17% in the presence of a germline mutation in *BRCA1* and *BRCA2* by the age of 80 years respectively. In males, the presence of a *BRCA2* mutations increases the overall lifetime risk of 0.1% for breast cancer up to 8% and in the presence of a *BRCA1* mutation to 1% (Evans et al., 2010; Tai et al., 2007).

The aim of screening for *BRCA1/2* mutations is to calculate the risk of an already affected individual from developing one of these cancers or a contralateral breast malignancy and to work towards preventing cancer in their unaffected relatives (Graffeo et al., 2016).

1.7 Microsatellites / VNTRs

Half the human genome is composed of repetitive sequences, a subset of which are short tandem repeats. Most of the tandem repeats are monomorphic but some are polymorphic. These polymorphic repeats are ubiquitous and also known as variable nuclear tandem repeats (VNTRs) or microsatellites (Cui et al., 2016). Microsatellites occur at thousands of locations within an organism's genome and make up approximately 3 % of the entire genome.

Microsatellites are repeats of one to six nucleotide sequences in tandem within DNA. They occur at a much higher frequency than SNPs in the genome of mammals. The term "satellite" DNA relates to initial observation that centrifugation of genomic DNA in a test tube resulted in 2 layers of DNA. There was a prominent layer of bulk DNA and a "satellite" layer of repetitive DNA (Ellegren, 2004; Kit, 1961). As a result of large genetic variability amongst humans, microsatellites may be used in forensic DNA fingerprinting (Gettings et al., 2015).

Microsatellite instability (MSI), defined by varying repeat copy number is a recurrent feature in gastrointestinal (colorectal and gastric in particular), breast, endometrial and lung cancers (Forgacs et al., 2001) and related to mutations or epimutations in DNA mismatch repair proteins (MMR) (Bebenek and Ziuzia-Graczyk, 2018; Veigl et al., 1998). Therefore the presence of MSI in several loci in multiple different cancers, including known tumour suppressor genes (e.g. PTEN), is strong evidence for a MMR deficiency and leading to a high mutational burden and tumour progression of these cancers (Yarchoan et al., 2017). Alterations in repeat unit number in and around coding sequences can have important quantitative and qualitative effects on gene expression (Galindo et al., 2011) and could potentially be directly instrumental in cancer progression (Bontekoe et al., 2001; Fondon and Garner, 2004; Houlston and Tomlinson, 2001).

1.7.1 Microsatellites / VNTR in breast cancer

Mutations in microsatellites have been associated with loss of function in different tumour suppressor genes and oncogenes (Bussaglia et al., 2000; Huff et al., 1995; Weitzel et al., 2000). The association of VNTRs with breast cancer and as modifiers of *BRCA1/2* gene mutations has previously been reported in the context of "candidate genes" (Cui et al., 2016) and have been proposed as a marker of risk for breast cancer for over two decades (Cui et al., 2016; Galindo et al., 2011; Tamimi et al., 2003; Xiang et al., 2012; Yoon et al., 2010; Zhang et al., 2011). The AAAG VNTR in the 5' UTR of the oestrogen related receptor gamma (*ESRRG*) gene in particular has been investigated in relation to breast cancer and will be discussed further in chapter 3. In this study we aim to investigate variation in AAAG VNTR length in a UK cohort of women with *BRCA* 1/2 mutations, BRCA negative breast cancer and healthy women with no cancer history as controls.

1.8 Single nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are the most frequently encountered genetic variation found in both coding and non-coding regions of the human genome. SNPs which are frequently found near particular genes can be a marker for that gene. They can also be a marker of genes associated with a specific disease if only found in individuals with the disease. SNPs found in coding regions could change the amino acid code affecting protein function and disease progression (Vachon et al., 2015).

The Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) have also confirmed a contribution of SNPs to breast cancer risk in *BRCA1* carriers (Couch et al., 2013). Approximately 100 SNPs have now been associated with breast cancer risk however SNPs identified at 18 loci in 2010 account for about two-third of the hereditary component of breast cancer and has been referred to as SNP18 which is in use for risk prediction (Eccles et al., 2013; Turnbull et

al., 2010). This panel was also used to predict risk in *BRCA2* mutation carriers by Ingham et al (Ingham et al., 2013).

Antoniou et al demonstrated an association with altered penetrance in *BRCA2* mutation carriers and nine common breast cancer associated SNPs (at the *TOX3, FGFR2, MAP3K, LSP1*, 2q35, *SLC4A7*, 1p11.2, 5p12, 6q25.1 loci) (Antoniou et al., 2010; Antoniou et al., 2011). Multiple variant testing is perhaps even more informative for individuals attending a family history clinic who are found not to be BRCA mutation carriers. The majority of these women who receive a negative result following germline testing for *BRCA1/2* only slightly reduce their predicted risk of breast cancer. For these women, a SNP Polygenic Risk Score (PRS) would provide a more meaningful result once *BRCA1* and *BRCA2* mutations have been excluded. This was highlighted by Evans et al who reported that the use of SNP18 PRS resulted in over half of BRCA negative women changing the National institute of health and care excellence (NICE) defined risk category (Evans et al., 2017).

Genome-wide association studies (GWAS) are particularly useful for investigation of SNP association with disease penetrance and those frequently found in noncoding regions (Welter et al., 2014; Zhang et al., 2016). Next generation sequencing advances have made it possible to identify further SNPs and to analyse them in more depth (Chan et al., 2017). Panel testing of multiple selected SNPs in combination can therefore be substantially more informative of cancer risk prediction than analysis of individual SNPs. This is essential for stratification of disease risk and treatment provision (Vachon et al., 2015).

1.9 Oestrogen receptors and oestrogen-related receptors

1.9.1 Nuclear Receptors

Nuclear receptors are a family of transcription factors which regulate several genes and physiological functions in a ligand-dependent manner. Nuclear receptors were initially identified as hormone receptors for 17β-oestradiol, corticosteroids, and thyroid hormones (Becker-Andre et al., 1993; Holzer et al., 2017). The nuclear receptors are called a super-family and consist of 48 members and 7 subgroups formed based on their sequence similarities. The Oestrogen receptor like group is of particular significance to this project due to its link with breast pathology. This group is further divided into Oestrogen receptors (ERs), Oestrogen related receptors (ERRs) and 3-ketosteroid receptors (including androgen, progesterone, glucocorticoid and mineralocorticoid receptors) (Riggins et al., 2010; Roshan-Moniri et al., 2014).

Nuclear receptors are characterised by 4 main conserved domains: the N-terminal regulatory domain consisting of activation function (AF)1; the DNA binding domain; hinge domain; and ligand binding domain. The DNA-binding domain consists of two zinc finger domains and demonstrates affinity for response elements (Helsen et al., 2012; Huang et al., 2010; Mullican et al., 2013).

The nuclear receptors have gathered considerable research interest over the years for the treatment of breast cancer. The ligand binding sites have been found to interact with small molecules which have lead to the discovery of drugs which target the oestrogen receptor in breasts cancer prevention and treatment. These molecules act as either antagonists or agonists by competing with endogenous ligands (Roshan-Moniri et al., 2014).

1.9.2 Oestrogen receptors (ERs)

The oestrogen receptors and oestrogen signalling pathways are involved in breast cancer progression (Ariazi and Jordan, 2006). Approximately 70% of breast cancers are oestrogen receptor positive (ER+ or ERα+) and cell proliferation is oestrogen-dependent (Jemal et al., 2009). These receptors activate gene expression in a ligand-dependent manner by acting as ligand regulated transcription factors. They are involved in various cellular events such as cell growth and apoptosis (Holzer et al., 2017; Platet et al., 2004). ERα represents the interface between cancer development, progression and treatment, this makes it the most important target in breast cancer therapy. ER expression in breast cancer is therefore not only a prognostic factor but also a very important predictive factor guiding oncological regiments. (Ariazi and Jordan, 2006).

The fact that the majority of breast cancers are oestrogen receptor positive (ER +) has a favourable impact on anti-oestrogen therapy in most patients as treatment of ER + breast cancers includes blocking oestrogen therapeutically (Ariazi et al., 2002). The two main principles of anti-oestrogen therapy in the oncological setting in breast cancer are to either inhibit oestrogen binding to the ER (tamoxifen) or prevent the production of oestrogen (aromatase inhibitors) (Howell et al., 2004). Current endocrine therapy consists of the selective oestrogen receptor modulators (SERMs) and aromatase inhibitors (AIs) (Crago et al., 2010). The SERM used most frequently in both pre- and post- menopausal women to treat breast cancer is Tamoxifen. Its mechanism of action is as an antagonist for the ER by competing with 17β-estradiol (Clarke et al., 2001). Aromatase inhibitor (AI) therapy is used in post-menopausal women with breast cancer. As the name suggests, it inhibits aromatase which is the enzyme involved in converting androgens to oestradiol. The AIs therefore deprive ER+ breast cancer of oestrogen and as a result blocks oestrogen signalling and cell growth stimulation. AIs currently in use are Exemestane, Letrozole, and Anastrazole (Smollich et al., 2009).

Adjuvant or neoadjuvant endocrine treatment is extremely effective in the management of ER+ breast cancer however a third of patients do not respond to Tamoxifen and 70% of the remaining group are at risk of disease recurrence (Early Breast Cancer Trialists' Collaborative et al., 2011; Riggins et al., 2005). As a result of increasing drug resistance the discovery of alternative proteins and pathways as targets is of interest. The ERRs are therefore being studied as potential alternative drug targets with the aim of reducing the incidence of disease recurrence through drug resistance (Early Breast Cancer Trialists' Collaborative, 2005; Roshan-Moniri et al., 2014).

1.9.3 Oestrogen Related Receptor Gamma (ERRγ)

The oestrogen related receptor (ERR) family consists of 3 isoforms: ERRα, ERRβ and ERRγ (Giguere et al., 1988). *ESRRA* gene encoding ERRα was mapped to chromosome 11q12-q13, the *ESRRB* encoding ERRβ was located to 14q24.3 (Shi et al., 1997; Sladek et al., 1997) and *ESRRG* encoding ERRγ on chromosome 1q41 (Eudy et al., 1998). The ERRs share sequence similarities and biochemical functions with ERs. ERRs however do not bind to oestrogen or endogenous oestrogen receptor ligands. These receptors are therefore called orphan nuclear receptors (Deblois and Giguere, 2013; Roshan-Moniri et al., 2014)

The ligand binding domain of the ERRs contains the activation factor-2 (AF-2) site which is of importance in the function of ERRs. This site is exposed as a result of the structure of the ligand binding domain and as a result it can bind coregulators in the absence of ligand binding. The consequence of this domain structure is a constitutively active receptor. The mediation of the Transcriptional properties of the ERRS is performed by coactivators (PGC-1α, SRC-1, SRC-2, SRC-3) and corepressors (RIP140) (Roshan-Moniri et al., 2014). ER shares structural overlap with ERRα and ERRγ and are therefore considered to have potentially similar functions in breast cancer (Roshan-Moniri et al., 2014).

Higher ERRα expression levels have been identified in aggressive and triple negative breast cancers. Other associated features are high tumour grade,

metastasis and risk of recurrence. ERRα therefore promotes cell migration and invasion and is an indicator of poor prognosis in breast cancer. Inactivation of ERRα impairs *in vitro* migration of breast cancer cells (Ariazi et al., 2002; Stein et al., 2008; Suzuki et al., 2004; Wu et al., 2015). Upregulation of ERRγ on the other hand is associated mainly with ER+/PR+ breast cancers and a favourable outcome in breast cancer. A microsatellite within the ERRγ gene is the subject of this project and will be discussed further in chapter 3.

The role of ERRβ in breast cancer is not fully understood. It is a coregulator of ER, mediates oestrogen response and may function as a tumour suppressor (Misawa and Inoue, 2015; Sengupta et al., 2014). While the exact mechanism of action of both ERRα & ERRγ in breast cancer is still not fully understood, current findings support a potential role of these receptors as therapeutic targets and prognostic predictors in breast cancer due to their similarities to the ER.

1.9.4 Interaction between Oestrogen Receptors and Oestrogen Related Receptors

ERs regulate gene expression by recruiting co-regulators to oestrogen response elements (EREs). This can only in part explain the effects of the oestrogen signal pathway. ERR and its oestrogen related response elements (ERRE) have been investigated as a potential for coregulatory recruitment and an alternative oestrogen signalling pathway (Sun et al., 2006). Coactivators which play important roles in ERR- mediated transcription have been found to be associated with both ERRs and ERs suggesting that transcriptional cofactors are partly shared between both receptor groups. ERRs are also thought to have an effect on ER-mediated signalling because ERRs can bind to EREs as well as ERREs (Knutti and Kralli, 2001; Lin et al., 2003; Puigserver and Spiegelman, 2003).

Promising discoveries are being made in the search for an alternative oestrogen signalling pathway to overcome the problem of resistance to hormone therapy and risk prediction. The roles of the nuclear receptors are not fully understood therefore research continues to explore these orphan nuclear

receptors as well as the search for ligands. ERRs are still currently considered to be orphans despite the discovery of ligands for many previously orphan nuclear receptors now considered to be ‘adopted’ (Benoit et al., 2006)

1.10 Micro RNAs

MicroRNAs (miRs) are a large class of non-coding RNAs regulating many biological processes (Bartel, 2004). miRs are typically 21-25 nucleotides long. Genes encoding miRs have been located in exonic, intronic (i.e. in host genes) and intergenic DNA (Rodriguez et al., 2004). Micro RNA genes are transcribed in the nucleus to pri-miRNA by RNA polymerase II (Cullen, 2004) and then processed by DROSHA, an RNaseIII to pre-miRNA which is then exported from the nucleus by exportin-5 into the cytosol (Lund et al., 2004; Wu et al., 2000). There pre-miRNA is cleaved by a second RNAse III, DICER to form miRNA duplex which after unwinding and degradation of the passenger strand forms the mature RNA interference induced silencing complex (RISC) (Hutvagner et al., 2001; Martinez et al., 2002). The mature miRNA-RISC complex then binds to the 3'-untranslated region (3'-UTR) of mRNA transcripts leading to mRNA degradation and herein post-transcriptional repression or RNA and catalytic RNA cleavage (Ambros, 2004; Filipowicz et al., 2008; Navarro and Lieberman, 2010). Micro RNAs regulate a wide range of cellular functions such as development (Sayed and Abdellatif, 2011), cell proliferation (Lenkala et al., 2014), cell differentiation (Ivey and Srivastava, 2010), apoptosis (Su et al., 2015), signal transduction (Inui et al., 2010) and cell cycle (Bueno and Malumbres, 2011). MicroRNAs may act as tumour suppressors or oncogenes (Dyrskjot et al., 2009; Zhang et al., 2014).

The role of miR137 in malignancies will be discussed in the results chapter as it is of particular interest in this project due to its role in breast cancer and its potential as a biomarker. Approximately 50 genes have been identified to be directly regulated by miR137 and found to be involved with the pathways mentioned above (Wright et al., 2013). A list of targets of miR-137 are outlined in table 1.1 adapted from Mahmoudi et al (Mahmoudi and Cairns, 2017).

Table 1.1: Targets of miR-137 adapted from Mahmoudi et al. (Mahmoudi and Cairns, 2017)

Target gene	***Biological relevance***	***Target gene***	***Biological relevance***
AEG-1	Role in RISC, miRNA functions, ontogenesis	MAPK3	Proliferation, differentiation, cell cycle
AKT2	Protein kinase, cancer development	MET	Embryonic development, wound healing
C10orf26	Potential tumor suppressor	MIB1	Apoptosis, cell cycle
CACNA1C	Regulating contraction, secretion, gene expression neurotransmission	MITF	Cell differentiation, proliferation, survival
CDC42	Cell cycle	MSI1	Stem cell regulator
CDK2	Cell cycle regulation	NCOA2	Cell growth, development, homeostasis
CDK6	Cell cycle	NSF	Fusion of transport vesicles
CPLX1	Synaptic vesicle exocytosis	PAQR3	Raf kinase regulation
CSE1L	Nucleocytoplasmic recycling of importin-α, cell migration, secretion	PTBP1	Pre-mRNA processing, metabolism and transport
CSMD1	Potential tumor suppressor	PTGS2	Dioxygenase and peroxidase
CTBP1	Development, proliferation	PTN	Apoptosis, cell proliferation
E2F6	Cell cycle	PXN	Actin-membrane attachment, cell mobility and migration
ESRRA	Regulator of energy metabolism	RB1	Cell cycle, tumor suppressor
EZH2	DNA methylation repression, embryonic development	RORa	Circadian rhythm, organogenesis and differentiation
FMNL2	Morphogenesis, cytokinesis, cell polarity, adhesion	SPTLC1	Sphingolipid biosynthesis
FXYD6	Na+/K+-ATPase regulation	SYT1	Synaptic vesicle exocytosis
GLIPR1	Differentiation, involved in cancer	TBX3	Development
Gpr88	Neuron development	TCF4	Neuronal development
HTT	Linked to Huntington's disease	TWIST1	Cell lineage determination/differentiation
KDM1A (LSD1)	Cell proliferation, growth differentiation	Wnt7a	Oncogenesis and development
KDM5B	Histone demethylase, cancer development	YBX1	Regulation of translation, transcription, cell migration, proliferation
KLF4	Development, differentiation,	ZNF804A	Neuronal development, cell adhesion
MAPK1	Proliferation, differentiation, development		

2 Materials and Methods

2.1 Ethics

Ethical approval was granted by the Central Manchester Research Ethics Committee reference number 10/H1008/24 for the collection and analysis of patient DNA and associated clinical information for the purpose of research.

2.2 Materials

2.2.1 Human DNA samples

Genomic DNA was extracted and provided by the National Genetics References laboratory, St Mary's Hospital Manchester. The cohort comprised of four groups with 200 female patients in each. The four groups consisted of : patients with wild type alleles of *BRCA1* and *BRCA2* (BRCA negative) all with breast cancer; patients with proven *BRCA1* or *BRCA2* germline mutations some of whom had breast cancer; and a control group of age matched tumour-free women with wild type alleles for *BRCA1* / *BRCA2* (control) .
Additional complete clinical details were not available by the time of completion of this project. Details of the specific types of the BRCA mutations were not provided and due to time constraints with the project these mutations were not explored.

2.2.2 Assessment of DNA quality by spectrophotometry

DNA concentration and quality were assessed by measuring absorbance at 260 and 280 nm using a NanoDrop 2000c spectrophotometer. The quality was considered appropriate when the ratio A_{260}/A_{280} nm was greater than 1.75; A_{280} of 0.1 - 1 was regarded as suitable for further processing.

2.2.3 Human cell line

2.2.3.1 MCF-7 Cells

This cell line was kindly provided by Professor Rudland and Professor Palmieri from the Royal Liverpool University Hospital.

MCF-7 cells were first described by Dr Herbert D. Soule. He successfully cultured metastatic cancer cells from the malignant pleural effusion of a metastatic breast cancer in an American nun (Sister Catherine Frances Mallon). Herbert D. Soule successfully cultured these cells overcoming difficulties with fibroblast overgrowth and nutritional environment experienced at the time by other scientists.

The resulting cell line was called MCF-7, named after the Michigan Cancer Foundation where Dr Soule was based and represented his seventh attempt at generating a cancer cell line (Lee et al., 2015a).

2.2.4 Commonly used buffers and cell culture reagents

2.2.4.1 LB Broth

LB broth, 2.5% (Fluka Analytical) in water. The 25 g LB broth was dissolved in 1000 ml of distilled water, autoclaved and stored at room temperature. The composition of the LB broth per ml is as follows

10 mg Tryptone

5 mg yeast extract

10 mg sodium chloride

2.2.4.2 LB Agar

LB agar, 4% (2.5% LB broth (see above), 1.5% bacteriological agar (Fluka Analytical). 15 g of bacteriological agar was dissolved in 1000 ml of 2.5% LB broth (see above). The LB agar was autoclaved and stored at room temperature.

2.2.4.3 5X TBE Buffer

Tris base, 108 g (Sigma-Aldrich), 55 g boric acid (Sigma-Aldrich), 5.84 g EDTA (Sigma-Aldrich) was made up to a volume of 2000 ml with distilled water.

- **Ampicillin (75μg/ml) / Kanamycin (25 μg/ml) stock solution.**
- **Insulin**

  Insulin solution human 100 μg/mL in PBS, pH 7.2 stock solution (Sigma-Aldrich; 11061-68-0)
- **Trypsin** (Sigma-Aldrich)
- **Oestrogen**

  17β-estradiol (1 mg/ml in ethanol); final concentration of 10nM dissolved in DMEM.
- **Freezing Media for eukaryotic cells**

  The freezing media was composed of 90% foetal bovine serum (Sigma-Aldrich) and 10% Dimethyl sulfoxide (DMSO) (Sigma-Aldrich).
- **1x PBS** (phosphate buffered saline; Sigma-Aldrich)

2.2.5 Cell culture media

2.2.5.1 Media for MCF-7 cells (MCF-7 media)

Cells were cultured in Dulbecco's modified Eagles medium (DMEM) with 4500 mg glucose/L (Sigma-Aldrich, cat no D5796), 10% foetal bovine serum (FBS; Sigma-Aldrich), penicillin (100 units per ml), streptomycin (0.1 mg/ml) and insulin (10 *μ*g/ml).

2.2.5.2 Stripped Media for MCF-7 cells (MCF-7 stripped media (oestrogen free media))

Stripped media consisted of DMEM with 4500 mg glucose/L free of phenol red (Sigma-Aldrich, cat no D1145), 5% (v/v) charcoal stripped FBS (Sigma-Aldrich), penicillin (100 units per ml), streptomycin (0.1 mg/ml) and L-glutamine (2mM) (Sigma-Aldrich). Phenol red has oestrogenic activity and stimulates the

proliferation of oestrogen receptor positive MCF-7 breast cancer cells (Berthois et al., 1986). Phenol red in tissue culture media acts as a weak oestrogen: Implications concerning the study of oestrogen-responsive cells in tissue culture (Berthois et al., 1986) and charcoal treatment of the serum causes endogenous hormone deprivation (Sikora et al., 2016).

2.3 Genotyping

2.3.1 Polymerase chain reaction

2.3.1.1 Polymerase chain reaction for genotyping

Polymerase chain reaction (PCR) (Kleppe et al., 1971; Mullis et al., 1986; Mullis and Faloona, 1987) was performed to amplify VNTRs within the DNA of the Manchester cohort of breast cancer subjects and controls using Go Taq flexi DNA polymerase (Promega).

A thermocycler: QB-96 (Quanta Biotech) and Peqstar 2X (Peqlab) were used for PCR. The optimisation process was carried out by performing temperature gradient PCRs. The PCR protocol for each gene is outlined in the table below.

2.3.1.2 PCR for AAAG VNTR in *ESRRG*

The sequence of interest in the 5' UTR of the oestrogen receptor gamma gene surrounding the 'AAAG' tetramer tandem repeat defined by genomic position GRCh38.p12, Sequence ID: NC 000001.11 217053466 – 217053500 was entered into BLAST for homology comparison: (https: //blast.ncbi.nlm. nih.gov/Blast.cgi). Primers were designed using UCSC genome (Hg 19) browser (http://genome.ucsc.edu/index.html) and Primer 3 programme (bioinfo.ut.ee/primer3). The primers were finally checked for specificity in the *in silico* PCR programme of the UCSC genome browser (http: //rohsdb.cmb.usc.edu/GBshape/cgi-bin/hgPcr). The designed primers were then produced by Eurofins (MWG).

ESSRG AAAG VNTR primer set:

Forward primer: 5'-acctaggagatagaggttgc-'3

Reverse primer: 5'-cttcttctgcactatcaggg-'3

The predicted PCR product was 369 bp long (chr1:217,053 ,376-217,053,745).

PCR reactions were carried out in a volume of 25 μl. The PCR master mix contained 5μl of 1x Go Taq flexi reaction buffer, dNTPs (400 μM), Taq Polymerase (0.05 U/μl), primers (0.4 μM), additional magnesium if required (2 mM) and DNA (5 ng/μl). Reactions were initiated with a 300 sec enzyme activation and DNA denaturation step at 95°C followed by 35 cycles consisting of an initial denaturation step at 95°C (30 sec), an annealing step at 59.3°C (30 sec) and finalised by an elongation step at 72°C (30 sec). The reaction was completed by a final elongation step for 600 sec at 72°C. The PCR product was finally stored at 4°C (figure 2-1).

Master mix for AAAG VNTR in *ESRRG*

	Volume	Concentration
5X Go Taq flexi buffer (undyed)	5 μl	1x
Mgcl$_2$ (25mM)	2 μl	2 mM
dNTP(10mM)	1 μl	0.4 mM
Forward Primer (20 μM)	0.5 μl	0.4 μM
Reverse Primer (20 μM)	0.5 μl	0.4 μM
Go Taq polymerase (5 U/μl)	0.25 μl	0.05 U/μl
Nuclease free water	14.75 μl	
DNA(5 ng/μl)	1 μl	
Final Volume	25 μl	

Master mix for each reaction (n=1)

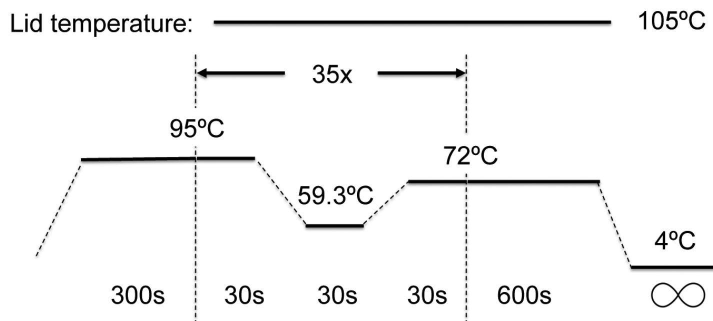


Figure 2-1: ESRRG AAAG VNTR thermo cycler protocol: Schematic outline of the thermo cycler protocol as described in 2.3.1.2

2.3.1.3 PCR for MIR137 VNTR

PCR reactions were carried out in a volume of 25 µl. The PCR master mix contained 5µl of 1x Go Taq flexi reaction buffer, dNTPs (400 µM), Taq Polymerase (0.05 U/µl), primers (0.4 µM), additional magnesium if required (2 mM) and DNA (5 ng/µl). Reactions were initiated with a 300 sec enzyme activation and DNA denaturation step at 95°C followed by 35 cycles consisting of an initial denaturation step at 95°C (30 sec), an annealing step at 68°C (30 sec) and finalised by an elongation step at 72°C (30 sec). The reaction was completed by a final elongation step for 600 sec at 72°C. The PCR product was finally stored at 4°C (figure 2-2).

MIR137 VNTR primer set:

MIR137-F: `5’-cacccaagaatacccgtcac-‘3`

MIR137-R: `5’-tgggagagcaccaggtaaac-‘3`

The shortest predicted PCR product was 370 bp. For each additional copy of the VNTR multiples of 15 bp were expected to be added to the PCR product.

Mastermix for MIR137 PCR (Warburton et al., 2015)

	Concentration	Volume
5X Go Taq flexi buffer (undyed)	5μl	1x
$MgCl_2$ (25mM)	4 μl	4 mM
dNTP(10mM)	1 μl	0.4 mM
Forward Primer(20 μM)	0.25 μl	0.2 μM
Reverse Primer(20 μM)	0.25 μl	0.2 μM
Go Taq polymerase(5u/μl)	0.25 μl	0.05 U/μl
Nuclease free water	12.25 μl	
DNA(5ng/μl)	2 μl	
Final Volume	25 μl	

Master mix for each reaction (n=1)

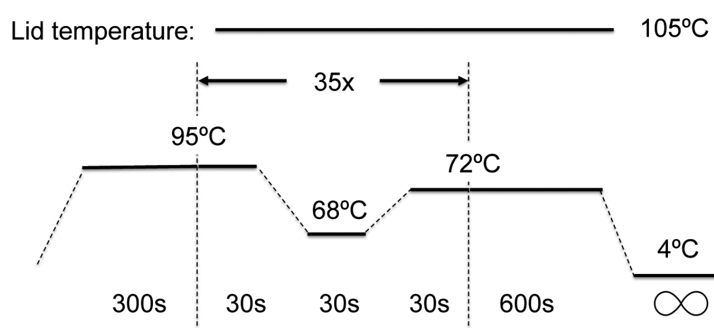


Figure 2-2: MIR137 VNTR thermo cycler protocol: Schematic outline of the thermo cycler protocol as described in 2.3.1.3

2.3.2 Agarose Gel Electrophoresis

Gel electrophoresis was used to separate DNA fragments by size in order to identify PCR products and allele variants. It was also used in order to segregate PCR products prior to DNA extractions for cloning. The amount of agarose (Bioline) used was determined by the percentage of the gel required. Gels were varied between 0.8% and 2% depending on the fragment size.

Example: 1.5% agarose gel: 0.9 g agarose powder in 60 ml 0.5x TBE buffer

Agarose powder was dissolved in the 0.5 x TBE and DNA intercalating dye (gel red (Biotium 10000X; 0.1 μl/ml TBE) was added. The warm liquid gel solution was poured into a casting tray and the wells within the gel were created by combs placed in the tray until the gel had set at room temperature. The casting tray containing the set gel was placed into a gel tank filled with 0.5 x buffer. Commonly, 15 μl of sample and 3 μl of loading dye (Promega 6x) were added to each well. DNA ladders of 100 bp and 1 kb (Promega) were used. The ladders were placed into the first and last wells of the gel. The gel was run in 0.5x TBE at a rate of 5 volts per centimetre.

The fragments of DNA were separated as an electric current passed through the gel migrating towards the cathode. The level of separation and the speed at which this occurred was dependent on the size of the fragments and the percentage of the gel used. DNA in gels was visualised using a UV transilluminator (BioDoc-it Imaging system).

2.4 Generation of reporter gene constructs

The 16 copies of the AAAG VNTR in the *ESRRG* were excised from the pGL3-control (pGL3C) plasmid by restriction digest and inserted it into a pGL3-promoter (pGL3P) plasmid containing a luciferase reporter gene (figure 2.1). pGL3-promoter was chosen as a vector as it has been designed for cloning and testing putative enhancer regions. It has also previously been successfully used by our group (Galindo et al., 2011).

The restriction reaction was performed in a digestion mix as follows:

10 x Digestion buffer	2 μl
DNA (1 μg)	x μl
Restriction enzymes (10 U/μl) (Acc 65I / Bgl II)	0.5 μl
10 x BSA (10 μg/μl)	2 μl
Nuclease free water	z μl
Final Volume	25 μl

The digests were incubated for 4 hours at 37ºC. In brief, the enzymes used gave products with end overhangs for use in DNA ligation reactions. The digestion products were run on an agarose gel as described above, visualised under a UV-transilluminator and the DNA bands of interest excised from the gel.

2.4.1 Agarose gel extraction of DNA fragments

DNA isolated from agarose gels and purified using the QIAquick Gel Extraction Kit (Qiagen) as outlined in the manufacturer's protocol. In brief: the gel containing the DNA of interest was dissolved in the solubilisation buffer 'QC' provided by the supplier. The dissolved DNA was mounted on the filter columns and then purified using wash buffers provided. At the end DNA was eluted from the filter columns and used for further processing (i.e. ligation experiments) or was frozen at -20°C.

2.4.2 Ligation of DNA fragments into pGL3P reporter gene vector

The digested vector and insert were used at a molar ratio of 1:3. The following equation was used to calculate quantities to be used.

$$\text{Insert (ng)} = \frac{\text{Vector (ng) x size of the insert}}{\text{Size of the vector (kb)}}$$

Insert	x μl
Vector	y μl
2 x ligation buffer	5 or 10 μl (depending on final volume)
T4 DNA ligase (100 units)	1 μl
Nuclease free water	z μl
Final Volume	10-20 μl

The above reaction was incubated at room temperature for 3 hours.

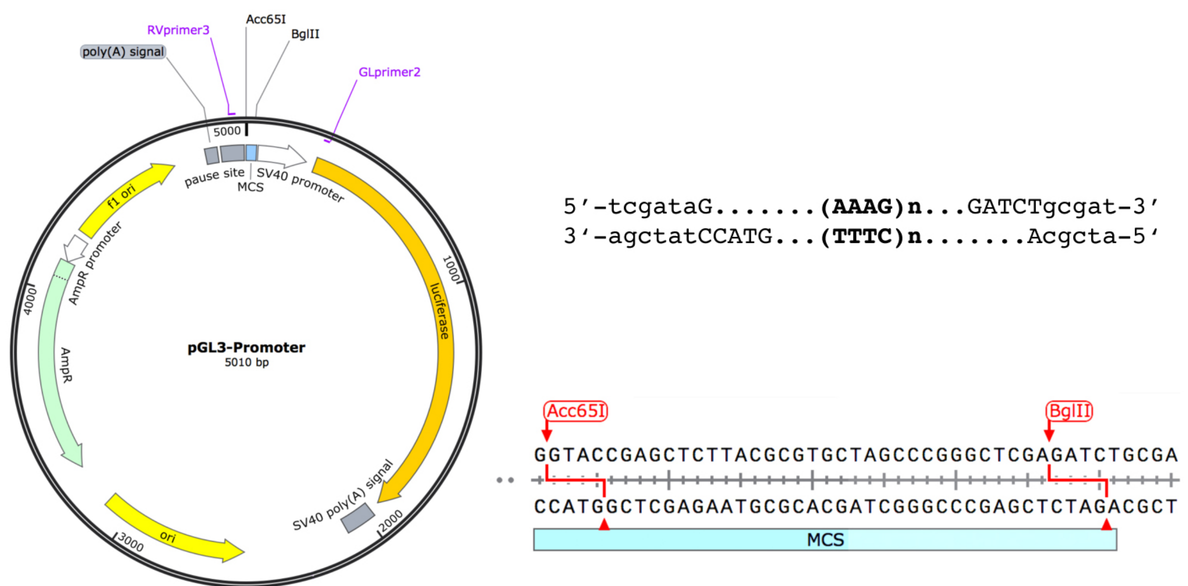


Figure 2.3: pGL3P construct: the multiple cloning site (MCS) highlighted with restriction enzyme cutting sites (Acc 65I and Bgl II) and site of AAAG insertion. Sequencing primers (RVprimer3 and GLprimer2, see section 2.3.7; sequencing) are indicated in purple.

2.4.3 Transformation of DH5α competent E. coli cells

50 μl of DH5α competent E. coli cells (Invitrogen) were thawed on ice for each transformation reaction. Each 50 μl of the thawed cells were placed into a 1.5 ml microcentrifuge tube and 5 μl (10 ng) of each ligation product were added to it and then mixed by gentle tapping. The reaction was incubated for 30 min on ice and then heat shocked at 42°C for 20 sec in a water bath following which it was placed on ice again for a further 2 min. 950 μl of pre-warmed LB broth was added to each tube and shook at 225 rpm for 1 hour at 37°C in an incubator. Then 200 μl of the above mixture was spread on warmed LB agar plates and incubated at 37°C overnight. The plates contained 100 μg/ml Ampicillin.

2.4.4 Miniprep of plasmid DNA

DNA was extracted from E. coli cells using QIAprep miniprep kit and was performed according to manufacturer's instruction. In brief: colonies were picked from the LB agar plates which had been previously spread with the

transformation mixture. Individual colonies were picked, added to 5 ml of LB broth with 5 μl of ampicillin (100 μg/ml) and grown in a shaking incubator overnight at 37°C. DNA isolation involved lysing the bacterial culture, adsorption of the plasmid DNA on a silica membrane and finally elution of purified DNA.

2.4.5 pGL3P Plasmid extraction

Single E. coli colonies from an inoculated LB agar plate or a scraping from a glycerol stock of the required plasmid DNA were added to a starter culture consisting of 5 ml LB broth and 5 μl of ampicillin. This was placed into a shaking incubator set at 37°C for 8 hours. 200 μl of the starter culture was then added to 100 ml of LB broth and 100 μl of ampicillin the next day. This was grown at 37°C in a shaking incubator for 12-16 hours. The cells where then pelleted by centrifugation and the pellet washed with PBS. The plasmid was then extracted using the Qiagen plasmid maxi prep kit according to the manufacturer's instructions. In brief, the cells were lysed using the provided lysis buffer. The lysis reaction was stopped by the addition of a neutralising buffer. The solution was then centrifuged and the DNA in the aqueous phase was transferred on the maxi prep columns. The plasmid DNA was washed and eluted from the columns. Isopropanol was added to the eluted DNA to precipitate it. The DNA was pelleted by centrifugation and washed with 70% ethanol to remove precipitated salt and to replace the Isopropanol. The ethanol was decanted and the pellet air-dried. The DNA was then resuspended in 50 μl of elution buffer. This process provides highly purified high yield plasmid DNA.

2.4.6 PCR confirmation of plasmid inserts:

High fidelity PCR Phusion kit (Thermoscientific) was used for plasmid sequencing. A standard PCR master mix included:

	Volume	Final concentration
Buffer (5x)	10 μl	1X
dNTPs (10mM each)	1 μl	200 μM
Upstream Primer (20 μM)	2.5 μl	0.5 μM
Downstream Primer (20 μM)	2.5 μl	0.5 μM
Nuclease free water	32.5 μl	
DNA template	1 μl	1 ng, 10 ng,
1 pg		
Phusion DNA polymerase	0.5 μl	0.02U/ μl
Final volume	50 μl	

2.4.7 Sequencing

Plasmids were then sent for sequencing to Dundee DNA Sequencing Service (Dundee, UK).

Sequencing primers (RV primer 3 (5'-CTAGCAAAATAGGCTGTCCC-3') and GL primer 2 (5'-CTTTATGTTTTTGGCGTCTTCCA-3') were used as recommended by the manufacturers protocol (figure 2.3). All electropherograms were manually read and samples with correct inserts used for further processing.

2.5 Cell Culture (Eukaryotic cells)

2.5.1 Culturing of MCF-7 Cells

MCF-7 cells were grown in modified DMEM in T175 tissue culture flasks. They were incubated at 37ºC / 5% CO_2 and passaged once 70% confluence was reached. To passage the cells the media was aspirated from the T175 flask. The cells were rinsed with 10 ml of sterile PBS (Sigma-Aldrich, cat no D8537) and then with 5 ml trypsin (Sigma-Aldrich). The trypsin was aspirated and the flask containing the cells was incubated at 37°C for 5 minutes until most of the cells lost adherence to the flask. The cells were washed from the flask with 10 ml of media. A 25 ml pipette was used to break up clumps of cells by pipetting up and down. 1 ml of cell suspension in media was added to 50 ml of media containing 50 μl of insulin in a new 175 cm^2 ventilated cap flask.
A Mycoalert Mycoplasma Detection Kit (Lonza) was used to test the cell lines for mycoplasma infection every 6 months.

2.5.1.1 Cell counting

The cell count was determined prior to plating the cells for transfection experiments. The cells in a T175 flask were passaged at 70 % confluency up to the stage were the cells were harvested in 10 ml of media as explained (section 2.5.1). 20 μl of cell solution was removed and placed at the edge of a coverslip positioned over the central chamber of a haemocytometer. The cells then filled the central chamber. The haemocytometer was then viewed under a light microscope at 10x magnification. The cells were counted within a 25 square grid in the haemocytometer which had a volume of 1 μl. Cells on the left and upper border of each square were included in the count whereas cell on the lower and right border were not. The total cell count was the number of cells in 1 μl of cell suspension. The number of cells in 1 ml of cell suspension were derived by multiplying the total cell count by 10^4. To convert this number into the number of cells per μl of cell suspension this figure was divided by 1000. In order to calculate the amount of cell suspension to add to each well to

give 100,000 cells per well the number of cells per μl of suspension was divided by 100,000.

2.5.1.2 Freezing cells for storage

Once cells had reached 70-80% confluency in a T175 flask they were trypsinised as described (section 2.5.1) until the step of washing the cells off the flask wall. At this point the cells were washed from the flask wall with 10 ml of freezing media. The freezing media containing the cells were divided between multiple cryovials with 1.8 ml placed in each. The cryovials were placed into a cell freezing container (Mr Frosty, Sigma-Aldrich) containing isopropanol and placed in a -80°C freezer for 24 hours. After 24 hours the cryovials were transferred to a liquid nitrogen tank for definitive storage.

2.5.2 Analysis of reporter gene expression

2.5.2.1 Transient transfection of reporter gene constructs into MCF-7 cells

Twenty-four well plates were used. 100,000 cells were plated per well suspended in 1 ml MCF-7 media supplemented with 1 μl insulin stock solution. The cells were incubated at 37°C and 5% CO_2 for 24 hours.

2.5.2.2 Transfection

Twenty four hours later when the cells were 70% confluent a master mix was made up as outlined below.

Table 2-1: plasmids

Construct	Volume of DNA (1µg/µl)	No of reactions	Total volume of DNA	Total Turbofect volume 2µl/ reaction	Serum free media (make up to 100 µl/ reaction)	HSV-thymidine kinase (TK) Renillin (20 ng/ µl)
pGL3C	1 µl	5	5 µl	10 µl	476.5 µl	2.5 µl
pGL3P	1 µl	5	5 µl	10 µl	476.5 µl	2.5 µl
19 copy	1 µl	5	5 µl	10 µl	476.5 µl	2.5 µl
7 copy	1 µl	5	5 µl	10 µl	476.5 µl	2.5 µl
9 copy	1 µl	5	5 µl	10 µl	476.5 µl	2.5 µl
16 copy	1 µl	5	5 µl	10 µl	476.5 µl	2.5 µl

The media was added first and the Turbofect (transfection reagent for efficient transfection of plasmid DNA) last. To make up the media is required to be FBS free, 0.5 µl of insulin was added to 50 ml of the medium.

The master mix solution was vortexed for 10-15 seconds then left for 20 minutes at room temperature to allow transfection complexes to form. After ligation 100 µl of the master mix was added to each of the 24 wells of the plate and the plate was then placed back in the incubator. After 4 hours the medium was changed to 1ml/well of media with FBS and the plate was placed back into the incubator.

2.5.2.3 Cell Lysis

The final stages of the transfection was completed 48 hours after the cells had been transfected. Gene expression was quantified using the Promega luciferase assay kit and performed as recommended by the manufacturers protocol: the media was removed from each well containing cultured MCF-7 cells and the cells then washed with 1 ml of sterile PBS. Cell lysis took place in 100 µl of PLB/reporter lysis buffer. The 24 well plate was then placed on a

rocking platform at room temperature for 15 min. Following this 20 μl of lysate was transferred from each well and placed into an opaque 96 well luminometer plate. The first and last two wells were left empty as controls.

2.5.2.4 Gene expression analysis by luciferase assay

The 96 well luminometer plate was then placed into a Glomax luminometer (Promega). The injectors were flushed with distilled water, 70% ethanol, distilled water again, air to clean them. One of the two injectors was then primed with luciferase assay reagent II (LARII) and the other injector with stop and go. Finally the programme injected luciferase assay reagents into each well and luminescence from the firefly luciferase reaction and then from the renilla luciferase reaction was measured using the Glomax luminometer. Gene expression levels of the constructs were measured, normalised and compared.

2.5.2.5 Treatment of cells with oestrogen

The protocol outlined above was predominantly followed with a few exceptions: two days prior to plating the cells the normal media was replaced with MCF-7 phenol red free media and charcoal stripped serum which was used for the remainder of the experiment. The experiment was then carried out as outlined (2.5.2.2) until the media was changed 4 hours post transfection when the following changes were performed (this media was used for the remainder of the experiment):

- For the basal conditions the media was replaced with fresh media
- For the vehicle control the media used contained 1μl/ml of molecular grade ethanol
- For the treatment the media contained 10nm of 17-β-oestradiol dissolved in molecular grade ethanol.

The cells in media were then incubated for 18 h following which the media was changed. After 48 hours of incubation the dual luciferase assay was performed.

2.6 Statistics

2.6.1 Clump statistical analysis

The genotyping data and allele frequency was statistically analysed using the clump 24 software which is available at http://www.davecurtis.net /dcurtis/software.html.

The Clump statistical analysis which has been used in over 100 articles of peer reviewed journals (examples: (Fan and Song, 2017; Hsieh et al., 2014; Lavu et al., 2015; Wang et al., 2015)) is designed for maximising the probability of significance with data presented in sparse contingency tables where this may not be possible in conventional Pearson chi-squared tests. It reduces the power of finding associations in studies involving large numbers of rare alleles in highly polymorphic loci. It generates 4 chi-squared tests using the Monte Carlo method, three of the tests 'clumping' alleles together. Clumping together the rare alleles and also comparing them individually to the rest of the alleles avoids expected small values whilst detecting meaningful associations (Sham and Curtis, 1995).

The four rows created by the program are as follows:

T1: The 2-by-N table provided by the user

T2: The original table including small numbers 'clumped' together

T3: Compares each non rare column of the original table against the total of all the columns

T4: A 2-by-2 table as a result of 'clumping' the columns of the original table. This maximises the chi-squared value

2.6.2 Correlation analysis

Correlation analysis for allele distribution between the groups was performed using a two-tailed Spearman rank analysis assuming non-Gaussian distribution. The analysis was carried out using Graphpat Prism 6.0.

3 AAAG microsatellite in the *ESRRG* gene as a potential biomarker for prediction of breast cancer risk

3.1 Introduction

The oestrogen related receptor gamma (*ESRRG*) gene encodes for the ERRγ protein and is on chromosome 1q41 (Figure 3-1). A notably polymorphic AAAG microsatellite in the 5' untranslated region (UTR) of this gene has previously been investigated. An association was found between the presence of long AAAG repeat lengths in the ERRγ gene and the risk of developing breast cancer by our group (Galindo et al., 2011). Our group has also previously demonstrated that 9 copies of the AAAG VNTR drives reporter gene expression in MCF-7 breast cancer cells (Galindo et al., 2011). In the present study I explored reporter gene expression with constructs containing 7, 9, 16 and 19 copies of the AAAG VNTR. With the knowledge that oestrogen can drive the expression of ER+ breast cancer I also investigated the influence of the various repeats on reporter gene expression with exposure to oestrogen. In the quest for a novel biomarker to predict breast cancer risk I genotyped various alleles of the AAAG VNTR within a UK breast cancer cohort to find a link between allelic variation and breast cancer risk.

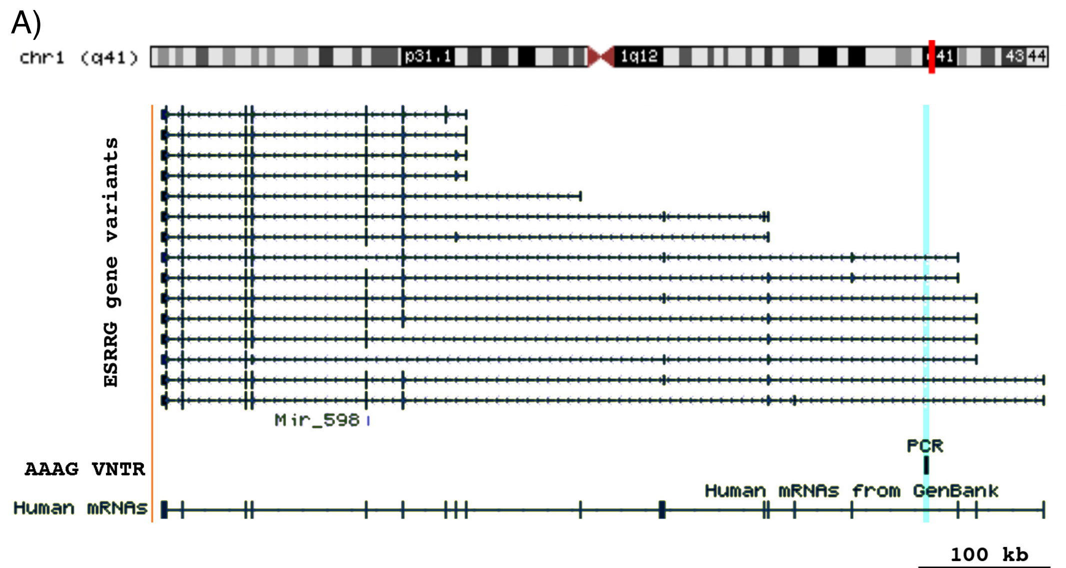
Other studies have revealed germline variations in microsatellite regions linked to glioblastoma, ovarian and breast cancer (Fonville et al., 2015; Karunasena et al., 2014; McIver et al., 2014). Both breast and ovarian cancer account for the majority of cancer related deaths in women. Novel biomarkers could therefore help identify individuals at risk of these cancers and guide surveillance as early cancer detection is associated with a better overall outcome. Germline whole exome sequencing of both breast and ovarian cancer patients have revealed loci associated with cancer risk for both breast

and ovarian tumours. 30% of variants found in breast cancer exome sequencing analysis were also predictive of ovarian cancer risk and 70% of those identified for ovarian cancer are also associated with breast cancer risk (Fonville et al., 2015; Karunasena et al., 2014; McIver et al., 2014)

Previous studies have examined the role of AAAG microsatellites in breast cancer in two cohorts of different ethnic backgrounds with conflicting results. In a study of an Iranian cohort breast cancer was linked to a lower copy number of the AAAG polymorphism in the intron of *ESRRG*. The allelic distribution ranged from six to eighteen repeats with nine repeats being the most common. AAAG repeats of six and seven were only found in the breast cancer patients linking these findings to breast cancer risk (Karimi et al., 2013). Whereas in an American study AAAG tandem repeats of 13 copies or longer located in the 5′UTR region of the *ESRRG* gene was observed in women with breast cancer (allelic distribution ranging from five to 21 repeats with seven repeats being the most prevalent). Their cohort consisted of a population of mixed ethnicity (African-American, Hispanics, Caucasians, and Chinese women). No significant difference was found between the different ethnic backgrounds (Galindo et al., 2011). These conflicting reports may be attributed in part to differences in ethnicity. Further research needs to be undertaken in order to support this theory (Karimi et al., 2013).

Interestingly varying copy numbers of the AAAG repeat within the P3 promoter region of the parathyroid hormone receptor 1 (*PTHR1*) gene have been shown to have a significant association with markers of bone desorption and adult height in an allele specific manner (Minagawa et al., 2002).
Several studies have reported increased risk of breast cancer related to high ERRγ expression levels (Ao et al., 2008; Ariazi et al., 2002; Ijichi et al., 2011) whereas others report a negative correlation with breast cancer (Eichner et al., 2010; Tiraby et al., 2011).

A)



B)

```
     >>>>> ESRRG fwd >>>>
TGAACCTAGGAGATAGAGGTTGCAGTGAGCCGAGATTGCGCCACTGCACTCCAGCCTGGG
             *         *         *         *         *         *

CAACAGAGTGAGATTCTGTCCCAAAGCCAAAAA[AAAG]AAATGCCTGCTGTAATGGGGT
             *         *         *           *         *

TGCAAAAACAAAATCTCAAAATGCTGTAGAAAAAATGCCTGCTGTAATGGGGTTGCAAAA
  *         *         *         *         *         *

ACAAAATCTCAAAATGCTGTAGAAAAATTATGTCATTTGTGTTTTAGTCATTAGATGTCC
  *         *         *         *         *         *

ATAGATTATTCAAGAATAAATATTAATACAAAAAATTGCATGTTACCAAAATGAGTTCTT
  *         *         *         *         *         *

GTGTTCCAAATAATTTGCATGGCATGACAAATAAATTATTGGCCTGGTTTCTAAAATGTG
  *         *         *         *         *         *
                   <<<<<< ESRRG rev <<<<<
GTAGTTGTTTTGTTTTTTACCCTGATAGTGCAGAAGAAGAGTCC
  *         *         *         *         *         *
```

Figure 3-1: Gene map of *ESRRG* : A) (UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly chr1: 216,675,576-217,500,000. Chromosomal position 1q41 highlighted on top. Graph adapted from UCSC Genome Browser): *ESRRG* gene locus with variants. The gene is located on the negative strand. Exons appear as vertical bars The $(AAAG)_n$ VNTR (chr1:217053466-217053469) spanning PCR is highlighted as a blue line. Possible mRNA transcripts collated at the bottom of the figure.
B) ESRRG VNTR PCR: Sequence targeted by PCR primers for the AAAG VNTR in the first intron of ESRRG, PCR amplicon highlighted in blue. PCR primer (ESRRG fwd and ESRRG rev are indicated as arrows. The repetitive VNTR element $(AAAG)_n$ is highlighted in yellow and delineated by brackets.

The mechanism of activity of the ERRγ may be explained by the sequence overlap of the ERR and DNA binding and ligand binding domains of the ER (Huss et al., 2015). Both ERR*α* and ERR*γ* can substitute for the oestrogen receptor (ER) (Ariazi et al., 2006). The main difference between ER and ERR is that ERRs do not directly bind endogenous ER ligands or oestrogen but they can bind to oestrogen response elements (ERE) and oestrogen related response elements (ERRE). Co-activators such as peroxisome proliferator-activated receptor γ (PPARγ), coactivator-1α (PGC-1α) and PGC-1β are found to be involved in transcription and modulation of ER signalling pathways and genes with functional ERREs. ERRγ which is a downstream target of the ER recognises ERRE and EREs which are bound to ER and modulates ER mediated transcription and the oestrogen signalling pathway. ERR control of target gene transcription, coregulatory proteins, ligands and site of action partly overlaps with the ER's implemented processes. In other words ERR can coregulate target genes with ER as well as regulating genes distinct from ER. Their exact mechanisms of action are yet to be fully elucidated (Foulkes et al., 2004; Giguere, 2002; Howell et al., 2004; Ijichi et al., 2011; Lu et al., 2012).

.

3.2 Aims

The aim of this chapter is to explore whether the AAAG VNTR could be a potential biomarker for breast cancer risk. I therefore performed genotype analysis of the AAAG VNTR in a female breast cancer cohort and matched controls to explore whether a difference in AAAG VNTR copy number differentiates between the groups. A second aim was to investigate whether the length of the AAAG VNTR affects *ESRRG* gene transcription in an oestadiol dependent manner. I addressed this by performing an *in vitro* luciferase reporter gene assay investigating the effect of different AAAG VNTR copy numbers on gene transcription with exposure to oestradiol.

3.3 Results

3.3.1 Genotyping of AAAG microsatellite in ESRRG in a breast cancer cohort

Genotype analysis was performed by simple observation in all four patient groups. Details of the groups are outlined in section 2.1.1. Allele calls were achieved for 44 cases in the BRCA1, 18 in the BRCA2, 54 in the BRCA negative and 26 in the control group. Due to low DNA concentrations and quality some of the samples could not be used and others could not be amplified by PCR. Nine different alleles were found which were numbered 1 to 9, where 1 represented the shortest and 9 the longest VNTR length (figure 3-1). The PCR products were not sequenced therefore the exact copy number is not known.

I observed the frequency of individual alleles and genotypes (tables 3-1 and 3-2, figure 3-3) in order to replicate the only two studies published on this subject which looked at the individual alleles in the AAAG VNTR and their association with breast cancer (Galindo et al., 2011) (Karimi et al., 2013). I found a clear pattern between specific alleles and particular groups within the cohort. The presence of alleles 2 and 3 were strongly associated with pathogenic *BRCA1* germline variant carriers. Allele 4 was significantly more common in the group of patients with pathogenic germline variants in *BRCA2* . The BRCA negative group demonstrated a more widespread distribution. The control group were mostly in the 5 to 9 allele category with the number 9 allele only observed in the control group. The larger alleles were therefore more of a feature in the control group, mainly with allele number 6 and 7.

The overall observation was that shorter repeats of the AAAG VNTR were associated with BRCA germline mutation carriers (BRCA1 and BRCA2) (alleles numbered 2 to 4) and longer repeats were a finding with the control group. The BRCA negative individuals did not appear to show a strong affinity with any particular allele number.

A significant correlation for alleles was observed for the BRCA1 group in a Spearman Rank correlation analysis (p=0.0209) and there was a strong trend for a correlation between allele distribution and the control group (p=0.0878). No correlation was found for the other groups.

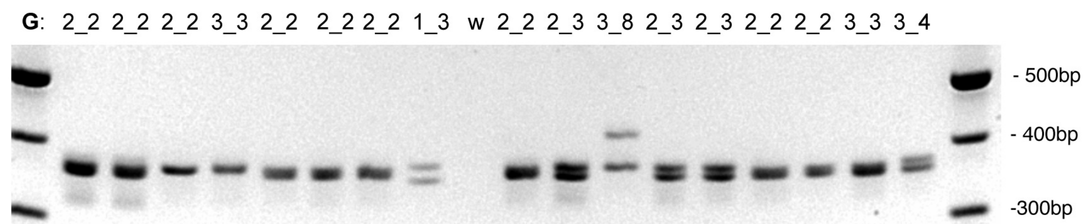


Figure 3-2: Genotyping of *ESRRG* AAAG VNTR PCR in a breast cohort: Representative agarose gel for the AAAG VNTR PCR (BRCA1 samples). A 2% gel was used to separate amplicons and a 100 bp DNA ladder (on right of gel image) was used to determine size. The shortest observed PCR amplicon spanning the VNTR in this cohort was called allele 1. G: Genotype, w : water control.

Table 3-1: Allele pair frequency (including percentages) of the AAAG VNTR in the ERRG gene. Nine different alleles were observed where 1 is the shortest and 9 the longest. Genotype comprises the two alleles found in a sample. The allele numbering does not directly correspond to the number of AAAG 4bp VNTR repeats. Frequency distribution of observed patient genotype defined by the two alleles listed in the first column (x_y).

Genotype	BRCA1 CA	BRCA1 no CA	BRCA2 CA	BRCA2 no CA	BRCA neg	control
1_3	1 (4.35%)	0	0	0	0	0
2_2	3 (13.04%)	6 (28.57%)	0	0	4 (7.14%)	0
2_3	0	3 (14.29%)	0	0	0	1 (4%)
3_3	13 (56.52%)	4 (19.05%)	0	0	12 (21.43%)	0
3_4	1 (4.35%)	0	0	0	1 (1.79%)	2 (8%)
3_5	1 (4.35%)	0	0	0	1 (1.79%)	0
3_6	1 (4.35%)	1 (4.76%)	0	0	0	0
3_8	0	1 (4.76%)	0	1 (11.11%)	0	0
4_4	1 (4.35%)	5 (23.81%)	8 (88.89%)	5 (55.56%)	8 (14.29%)	0
4_6	0	0	0	0	1 (1.79%)	0
4_7	0	0	0	0	2 (3.57%)	0
4_8	0	0	0	1 (11.11%)	1 (1.79%)	0
5_5	2 (8.7%)	1 (4.76%)	0	1 (11.11%)	10 (17.86%)	0
5_6	0	0	0	0	1 (1.79%)	2 (8%)
5_7	0	0	0	0	2 (7.14%)	0
5_8	0	0	0	0	3 (5.36%)	0
6_6	0	0	0	0	6 (10.71%)	2 (8%)
6_7	0	0	0	0	0	6 (24%)
6_8	0	0	0	0	0	1 (4%)
7_7	0	0	1 (11.11%)	0	2 (3.57%)	1 (4%)
7_9	0	0	0	0	0	8 (32%)
8_8	0	0	0	1 (11.11%)	0	1 (4%)
8_9	0	0	0	0	0	2 (8%)
Total	23	21	9	9	54	26

Table 3-2: Individual allele frequencies (including percentages) of the AAAG VNTR in the ERRG gene. The BRCA1 and 2 groups are divided into subgroups with and without breast cancer (br-CA: breast cancer).

Allele	BRCA1 no CA	BRCA 1 br-CA	BRCA2 no CA	BRCA2 br-CA	Control	BRCA neg
1	0	1 (2.17 %)	0	0	0	0
2	15 (35.71 %)	6 (13.04 %)	0	0	2 (3.85 %)	8 (7.41 %)
3	13 (30.95 %)	30 (65.22 %)	1 (5.56 %)	0	4 (7.69 %)	26 (24.07 %)
4	10 (23.81 %)	3 (6.52 %)	11 (61.11 %)	16 (88.89 %)	0	21 (19.44 %)
5	2 (4.76 %)	5 (10.87 %)	2 (11.11 %)	0	5 (9.62 %)	27 (25 %)
6	1 (2.38 %)	1 (2.17 %)	0	0	14 (26.92 %)	14 (12.96 %)
7	0	0	0	2 (11.11 %)	18 (34.62 %)	8 (7.41 %)
8	1 (2.38 %)	0	4 (22.22 %)	0	6 (11.54 %)	4 (3.7 %)
9	0	0	0	0	3 (5.77 %)	0
total	42	46	18	18	52	108

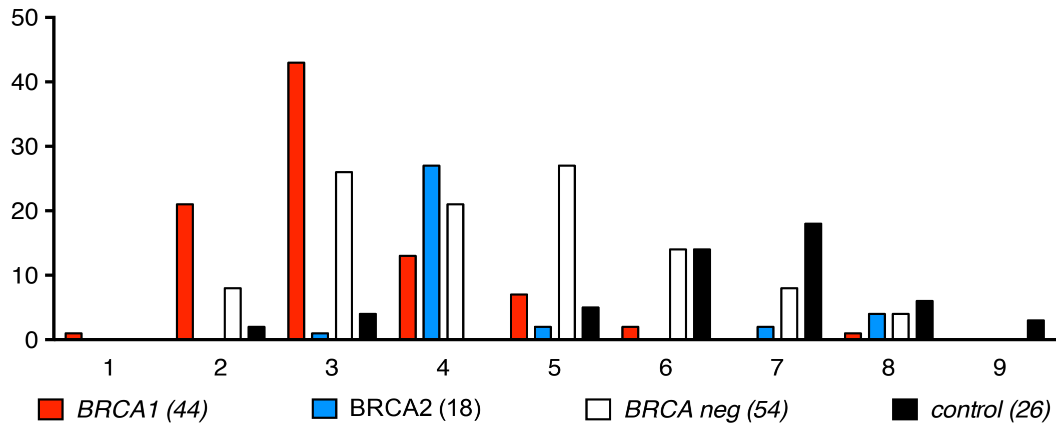


Figure 3-3: Graphs of allele frequency occurrence of the AAAG VNTR. Number of individual alleles and their distribution in each group. Numbers in brackets indicate patient group size)

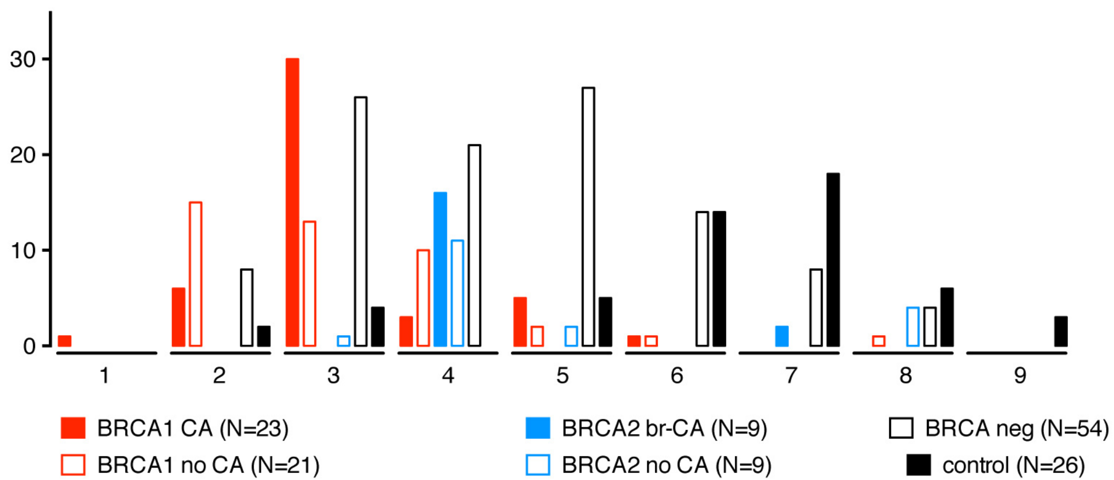


Figure 3-4: Graph of single allele distribution percentage frequencies with BRCA1 and 2 groups subdivided into groups with and without breast cancer.

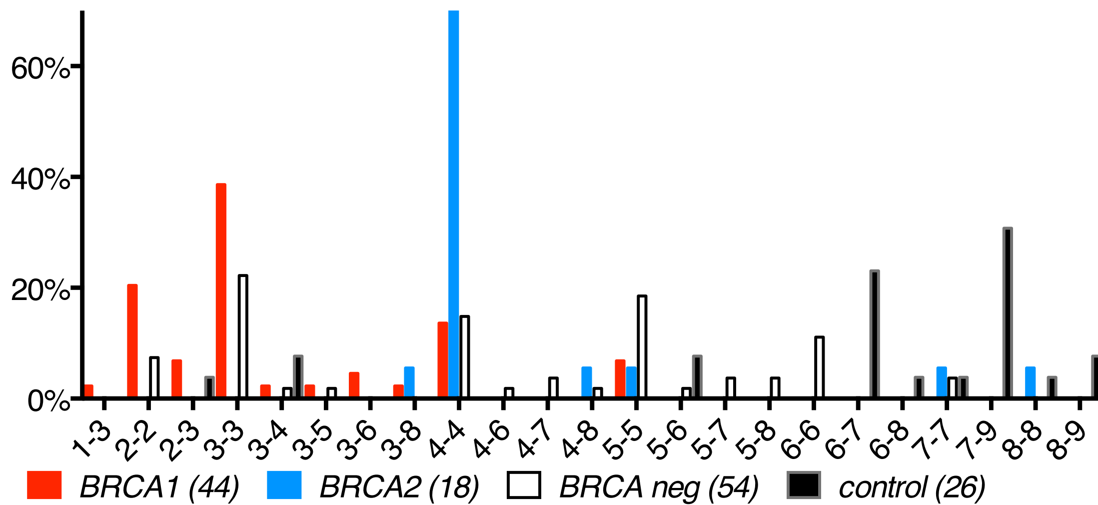


Figure 3-5: Graph of allele pair distribution percentage frequencies.

On assessment of allele pairs, I found homozygosity of the 2-2 and 3-3 alleles were the main feature of BRCA1, homozygosity with 4-4 for BRCA 2 and heterozygosity for 5-6, 6-7 and 7-9 allele combinations in the control group with allele number 9 only found in the control group in combination with alleles number 7 and 8 (figure 3-5).

Table 3-3: Comparison of AAAG alleles within BRCA1, BRCA2, BRCA -ve and the control groups. The BRCA +ve groups were compared with each other in cancer and disease free subgroups of BRCA1 and 2 and in a combined BRCA +ve group with and without cancer. Statistical testing was performed using clump analysis using the Monte Carlo approach.

AAAG VNTR groups	T1	T2	T3	T4
CTL vs BRCA neg	0.000002	0.36069	1	0.0001
CTL vs BRCA1	0.000013	-	0.000442	0.0001
CTL vs BRCA2	0.019049	-	0.000001	0.0001
BRCA neg vs BRCA1	0.112872	0.021657	0.06327	0.0046
BRCA neg vs BRCA2	0.051279	-	0.000005	0.0002
BRCA1 vs BRCA2	0.030738	-	0.000006	0.0001
BRCA1 CA vs BRCA1 no CA	0.014629	0.000555	0.001327	0.0006
BRCA2 CA vs BRCA2 no CA	0.193092	1	0.054349	0.013899
BRCA1&2 CA vs BRCA1&2 no CA	0.013134	0.00783	0.006202	0.021198

The BRCA 1 and 2 groups were further divided into individuals with and without a personal history of breast cancer (tables 3-1, 3-2, figure 3-4). Although this created even smaller numbers significant associations were made using a clump analysis (Table 3-3). Significant distributions were seen in alleles associated with breast cancer in both, the BRCA 1 and 2 positive groups. These findings mirrored the findings in the combined BRCA groups with homozygous 2-2, and 3-3 alleles in BRCA1 and homozygosity for allele 4 in BRCA 2 carriers (figure 3-5). An interesting observation was that a significant proportion of the number 3 allele in the BRCA 1 group was found in individuals with breast cancer (tables 3-1, 3-2 and figure 3-4). Clump analysis showed significant differences between all comparisons (table 3-3).

Overall the findings of the distribution of AAAG VNTR in the different patient groups are interesting and could suggest an association for cancer risk which may be used as an additional predictive marker for breast cancer risk. As the main cancer finding was in the BRCA positive group clinically this may not add to the currently established, successful diagnostic tool used for identifying the BRCA mutation. However BRCA mutation carriers do not demonstrate a 100% disease penetrance therefore our findings could add additional evidence in calculating the cancer risk in the BRCA mutation carrier. This could give even more information to BRCA carriers about risk prediction during genetic counselling.

The study will require reproduction in a larger cohort which may give more BRCA mutation carriers with cancer in order to make more statistically significant associations within this group. Allele analysis using capillary electrophoresis or sequencing to determine exact AAAG repeat numbers within each allele should also be carried out. In the current study a large number of comparisons have been made in a small sized cohort which could lead to false positive or negative results.

If there had been an overwhelming characteristic allele distribution specifically for all individuals with cancer including the BRCA-ve group then a conclusion

may have been made to associate specific alleles to a general cancer risk prediction. This was not found on this occasion but has been reported by other groups (Galindo et al., 2011).

3.3.2 Analysis of the reporter gene constructs containing varying lengths of the AAAG microsatellite under basal, vehicle (ethanol) and treatment (17-β-oestradiol) conditions in MCF-7 cells

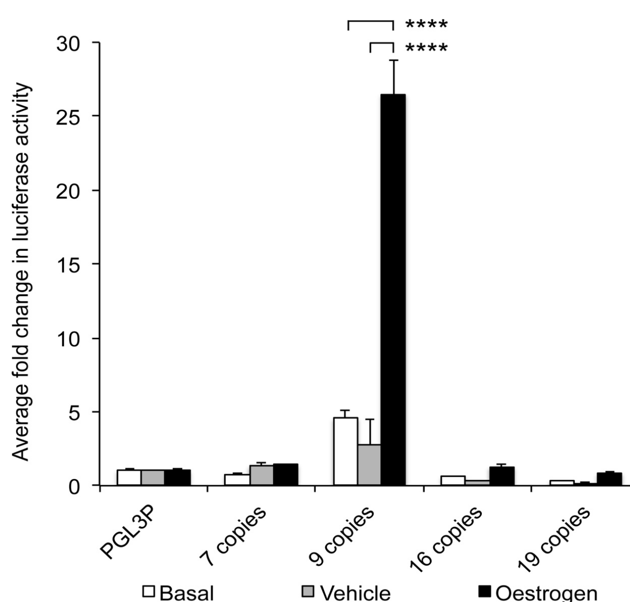


Figure 3-6: Effect of AAAG copy number repeat on reporter gene expression. The graph displays the average fold difference in activity of the reporter gene constructs containing 7,9,16 and 19 copies of the AAAG microsatellite over the pGL3P (minimal SV 40 promoter) normalised to TK *renilla* luciferase activity in MCF-7 cells. N=4. Two-way ANOVA (analysis of variance) was used to calculate the significance of fold change between all copy numbers and pGL3P in basal, vehicle and treatment (10nM 17β-oestradiol). (See table 3-4 for statistical results). Significance in differences within the nine copy number group are indicated with '*', where **** indicates a statistical p value smaller than 0.001 in a two-way ANOVA analysis (table 3-4).

Table 3-4: statistical significance between transfection groups. Table of significance of average fold change comparing the 7, 9, 16 and 19 copies of the AAAG VNTRs compared with each other and pGL3P. Statistical significance was calculated using two-way ANOVA.

	Adjusted P value
basal	
PGL3P vs. 9 copies	< 0.0001
7 copies vs. 9 copies	< 0.0001
9 copies vs. 16 copies	< 0.0001
9 copies vs. 19 copies	< 0.0001
vehicle	
PGL3P vs. 9 copies	0.0002
7 copies vs. 9 copies	0.0007
9 copies vs. 16 copies	0.0002
9 copies vs. 19 copies	< 0.0001
treatment	
PGL3P vs. 9 copies	< 0.0001
7 copies vs. 9 copies	< 0.0001
9 copies vs. 16 copies	< 0.0001
9 copies vs. 19 copies	< 0.0001

	Adjusted P Value
9 copies	
vehicle vs. basal	0.0178
treatment vs. basal	< 0.0001
treatment vs. vehicle	< 0.0001

Four AAAG microsatellite variants (7, 9, 16 and 19 copies) were cloned into a vector with a minimal SV40 promoter (pGL3P) and analysed using reporter gene (luciferase) expression which was normalised to TK *renilla* in an MCF-7 cell line. Transfections were performed in quadruplicate. The reporter gene constructs demonstrated variable levels of expression which were minimal in

the 7, 16 and 19 AAAG repeats. Activity was however significantly enhanced with the 9 copies of the AAAG repeat (figure 3-6).

The reporter gene activity of the construct containing the 9 copies of the AAAG VNTR in MCF-7 cells showed a significant increase in basal conditions over pGL3P alone ($P<0.0001$). Activity with the vehicle conditions also significantly increased ($P<0.0002$) over pGL3P. When the MCF-7 cells were treated with 17β-oestradiol the 9 copy AAAG VNTR showed a 26.5 fold ($P< 0.0001$) increase in reporter gene activity. The activity of the 9 copy repeat in all 3 conditions (basal, vehicle, treatment) compared with the activity of the 7, 16 and 19 VNTR repeats under the three conditions was found to be highly significant (table 3-4).

3.4 Discussion

This study has shown that the AAAG microsatellite in the 5' UTR of the *ERRG* gene can support promoter gene activity. A variation of the VNTR copy number alone lead to an altered transcriptional activity, which was further enhanced upon stimulation with 17β-oestradiol. This supports findings of previous studies which also report similar gene expression in MCF-7 cells (Galindo et al., 2011) and an association between the repeat length of the AAAG microsatellite and transcription (Minagawa et al., 2002).

We found significantly higher reporter gene activity specifically with the 9 copies of the AAAG repeat in MCF-7 (ER+) breast cancer cells especially with exposure to oestrogen compared to the 7,16 and 19 repeat lengths. Ijichi et al explored ERRγ upregulation in various oestrogen environments. Their data showed that ERRγ was overexpressed in an oestrogen dose dependant manner, with oestrogen deprivation, in ER+ (MCF-7 cells) and ER- (MDA-MB-23 cells) breast cancer cells (Ijichi et al., 2011).

As Ijichi et al showed ERRγ expression in an oestrogen dose dependent manner the various lengths may also show differential reporter gene activity as a result of differing hormonal sensitivities and varying oestrogen dose

exposure. Variable hormone sensitivity could explain why 9 copies of the VNTR was the 2nd most common allele in the study by Galindo et al. (Galindo et al., 2011) and was the most commonly occurring allele in the cohort published by Karimi et al (Karimi et al., 2013) yet this did not translate into a marker for breast cancer risk in either study. In the American cohort published by Galindo 13 AAAG repeats and longer were found to be associated with breast cancer (Galindo et al., 2011). However in the Iranian study by Karimi shorter copy numbers (6 and 7 copies) were linked with breast cancer (Karimi et al., 2013). Genetic differences between ethnic backgrounds may be due to differences in genetic composition, lifestyle and environmental exposures. Patients who are carriers of the 9 copy allele will have varying exposure to oestrogen through lifestyle and reproductive history differences. This may translate into the resulting ERRγ activity caused by the 9 copy allele and therefore increase the risk of developing ER+ breast cancer. This also raises the question of whether the 7, 16 and 19 copies of the AAAG microsatellite may show greater activity in ER negative cells which may be explained by differences in transcription factor complements between ER positive and ER negative cell lines (Goldenberg and Froese, 1982).

There is mounting evidence that ERRγ may be a marker for favourable outcome in breast cancer. ERRγ over expression is seen in less aggressive and ER/PR positive tumours. Its overexpression is also associated with overexpression of ErbB4 which is thought to indicate good clinical outcome as well (Ariazi et al., 2002). 4-hydroxytamoxifen (4-OHT) has been identified as a high affinity ligand for ERRγ causing the complex created by ERRγ and steroid receptor coactivator-1 to dissociate and as a result inhibit transcriptional activity of ERRγ. The discovery of 4-OHT antagonising ERRγ was made by peptide sensor, direct binding and cell based assays (Coward et al., 2001; Tremblay et al., 2001). The discovery of an alternative tamoxifen pathway involving ERRγ could aid patient selection for hormone therapy in the future as well as providing an alternative pathway to overcome tamoxifen resistance. It would be interesting to test the effect of the individual lengths of the VNTR on

this pathway to explore which of the repeat lengths specifically may be sensitive to anti-hormone therapy.

Published data as to whether ERRγ expression has a positive or negative effect on breast cancer are conflicting (Ariazi et al., 2002; Galindo et al., 2011; Tiraby et al., 2011). One study which was performed using CHIP analysis and reporter gene activity in MDA-ERRγ cells found ERRγ to positively regulate E-cadherin expression by direct binding of ERRγ via two ERREs in the second intron of the E-cadherin gene (*CDH1*) (Tiraby et al., 2011). In the same study an ERRγ promoted E-cadherin expression and mesenchymal-to-epithelial transition (MET) was observed in tumour xenografts leading to diminished invasiveness and suppression of tumour growth. This finding also supports the role of the ERRγ as a marker of good prognosis.

In summary ERRγ may play an active role in breast cancer through its overexpression specifically in the 9 copy VNTR (Galindo et al., 2011). Its response to oestrogen may also be a useful marker for favourable clinical outcome and hormone therapy response in ER positive breast cancer.

Follow on work from this project should explore reporter gene activity of the various repeat lengths with exposure to varying concentrations of oestrogen in both ER+ and ER- breast cancer cells. This study should also be linked with the investigation of the effect of varying lengths of the VNTR on the tamoxifen pathway and breast cancer cells.

Our genotyping analysis of the AAAG VNTR in a breast cancer cohort revealed an association between *BRCA1* mutation carriers and allele numbers 2 and 3, *BRCA2* mutation carriers and allele number 4 and the control group and allele numbers 5 to 9. Allele 9 was only observed in the control group. Allele number 3 was observed in individuals with breast cancer within the BRCA1 group and allele number 4 was seen in breast cancer patients with a *BRCA2* mutation.

Overall therefore longer repeats were observed in the control group and shorter VNTRs were found in the BRCA positive group.

Our discovery in the pathogenic *BRCA1* germline mutation carriers is of interest as these individuals tend to develop ER negative breast cancers which are difficult to treat due to their hormone insensitivity. *BRCA1* related breast cancers also tend to present at a younger age and are generally of a higher grade. This finding may elude to a potential marker of breast cancers which may be a challenge to treatment. Further analysis would be required not only to attempt to reproduce the findings in the BRCA1 group but also to identify exactly how many copies of the VNTR are represented by allele number 2 and 3.

ERRγ expression in the brain, skeletal muscle, heart, kidney, and retina suggests its involvement in maintenance and differentiation of these tissues in humans. There are several alternatively splice variants of the ERRγ mRNA in the 5' UTR which are tissue specific. Over six different *ERRG* transcripts, some of which result in alternative protein isoforms have been reported (Heard et al., 2000). The mechanisms involved are unclear. Two isoforms have been described to occur in the 5' UTR and the first exon of *ERRG*. These may lead to alternative gene regulation which is dependent on the variable AAAG microsatellite lengths. This effect has been described as a possible explanation for the parathyroid hormone receptor 1 (*PTHR1*) gene which contains AAAG repeats in the promoter region which co-varies with adult height (Scillitani et al., 2006). Another transcription factor which contains the AAAG VNTR which may bind to the 5′UTR of *ERRG is the* paired domain transcription factor PAX2 which has been shown to activate specific target genes by direct binding (Bouchard et al., 2005; Galindo et al., 2011). Future studies could explore the AAAG microsatellite and its effect on transcription factor binding sites in the region for potential mechanisms of action.

Another explanation for differences in gene regulation and phenotype in cancer could be ‘over expansion’ of the microsatellite leading to misregulation of splicing. Ancestral lengthening of the VNTR coupled with mutations in the *ERRG* have also been suggested as a cause for variations in gene regulation. Alternatively the AAAG repeat may be located in an active RNA which has not been identified and may result in variation in activity based on repeat length (Galindo et al., 2011).

In summary characteristic, significant allele distributions were identified for the BRCA positive and control group. Substantially more 2-2 and 3-3 homozygosity was seen in the BRCA 1 group with significantly more BRCA 1 individuals with breast cancer presenting with the 3 allele. The BRCA 2 group were homozygous for the 4 allele with a higher proportion of individuals with cancer compared to cancer free individuals with the mutation. Allele 4 may therefore be associated with a higher penetrance of cancer in individuals who are carriers of a BRCA 2 mutation. The length of the AAAG VNTR in ERRγ may also act as a biomarker to predict risk between the subtypes of breast cancer with differing hormonal status and sensitivities to oestrogen.

A larger cohort with cross reference to clinical data and course of disease is required as a follow on to this research so that more in depth clinical associations can be made if present, with more statistical significance.

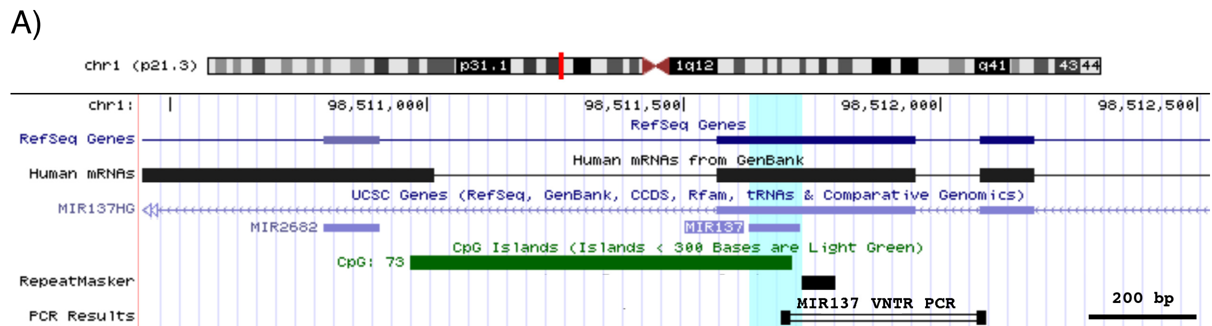
4 MIR137 VNTR as a potential biomarker for prediction of breast cancer risk

4.1 Introduction

In this chapter I address the potential role of a VNTR associated with MIR137 in relation to breast cancer. Micro RNAs play a pivotal role in carcinogenesis and have also been discussed as potential therapeutic targets (Lee et al., 2015b). *MIR137* (1p22) is located in a CpG island (figure 4-1) and is of particular interest as it has been studied by our group in the context of schizophrenia (Warburton et al., 2016a; Warburton et al., 2016b; Warburton, 2015) and extensively by others in relation to cancer. Some of these findings in cancer have been summarised and referenced in table 4.1. MIR137 has the hallmarks of a tumour suppressor and the downregulation of MIR137 has been proposed as being an early event in carcinogenesis (Huang et al., 2016). It is related to epigenetic regulation by promoter hypermethylation in a variety of cancers (Langevin et al., 2010; Zhang et al., 2018). Epigenetic modulation of micro RNAs has been described for a variety of cancers (table 4-1).

Overexpression of *MIR137* in cancer has demonstrated inhibition of cell invasion, cell cycle arrest, apoptosis and improved chemosensitivity through the effects of various targets (table 4-1). The silencing of miR-137 has been associated with aberrant DNA methylation in the region of the CpG island in oral squamous cell carcinoma (Kozaki et al., 2008) and lung cancer (Kang et al., 2015). *In vitro* experiments have shown that miR137 impairs proliferation and migration of breast cancer cells by targeting expression of the nuclear receptor oestrogen-related receptor alpha (ERRα) (Zhao et al., 2012). The possibility of reactivation of silenced MIR137 through EpiDrugs may present new therapy options based on the exploration of its role in the clinical setting. Epigenetic modifiers, have recently been introduced for anti-tumour therapy (Verma and Banerjee, 2015).

A)



B)

```
                   >>>>> MIR137-F >>>>>
GTAATCCGTATTATCCACCCAAGAATACCCGTCACCGAAGAGAGTCAGAGGACCAAGCTG
           *         *         *         *         *         *

[CCGCTGCCGCTGCTA]CTGCCGCCGCCGCCGCCACCAGAACTCTTGCTGCTCGCTGAGC
          *          *         *         *         *

CCGCCCCTGCCTGGGGATGGGCTGAGCTTGACCGGGACCATAAATCCATAACTCGATTTC
 *         *         *         *         *         *

CCTAAAGAAGGATCCAAAGCTGTGCTCGGCTGCTTCCTGCCCAAATCCAAATGGCCGCTC
 *         *         *         *         *         *

TATTTCCAATTctgaaagacaaatcacaaaatcaaatacttaacacagaagagtgctgtc
 *         *         *         *         *         *

tttccagtgaggagggaagatgtggcaaggcttttgggcacaagtgggagtgacaagta
 *         *         *         *         *         *
       <<<<< MIR137-R >>>>>
accttcagtttacCTGGTGCTCTCCCAGTGCTTTCCT
 *         *         *         *
```

Figure 4-1: *MIR137* gene map: A) (UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly chr1:98,510,446-98,512,572): chromosomal region flanking MIR137 (highlighted in light blue) featuring the CpG island (green), VNTRs and MIR137 amplicon (bottom line). B) MIR137 VNTR PCR: Sequence targeted by PCR primers for the internal MIR137 promoter (Imir137) (adapted genomic reference sequence: GRCh38.p12: 98046542-98046118). Upper case font indicates exons, lower case introns. PCR primer (MIR137-F and MIR137-R are indicated as arrows. The repetitive VNTR element is highlighted in yellow and delineated by brackets, the two CpG dinucleotides within the VNTR in bold and double underlined. Single underlined text marks the MIR-137 sequence.

Colleagues within our group have identified and validated an internal promoter (Imir137) adjacent to *MIR137* in the *MIR137-H*G (Warburton et al., 2016a). This internal promoter contains a 15 bp VNTR which was found to contain several variants ranging from 4 to 12 copies. The most common variant identified was homozygosity for 4 copies of the VNTR in 55% of their schizophrenia cohort. The 12 copy variant was however rare therefore its role could not be further explored with significance. The 12 copy repeat within the ImiR VNTR was also found to be functionally distinct by reporter gene analysis. The copy number differentially modulated the ImiR137 reporter gene function: Mamdani and colleagues showed that 9 and 12 copies of the ImiR VNTR were associated with a decreased *MIR137* transcription (Mamdani et al., 2013). This result was also partly reproduced by Warburton and colleagues who showed that 12 copies were differentially expressed (Warburton et al., 2016a).

Other groups have also found repeat number variation (3 and 12 copies) to differentially alter the processing of mature miR-137 in a melanoma cell line model (Bemis et al., 2008).

Research performed by our group has created an interface between the role of MIR137 in both schizophrenia and breast cancer. Work originally performed by colleagues on schizophrenia has formed the bases for and inspired further work to investigate the role of MIR137 VNTR in breast cancer.

Table 4-1: Table of targets for MIR137 and their observed effects in cancer

Cancer Type	Validation method	MIR137 Target	Effect of MIR137 up-regulation	References
Pancreatic	luciferase reporter assays, western blot in cell lines, tumour cells	PTN	inhibition of cell division, increased sensitivity to 5FU chemotherapy, inhibition of tumour formation	(Xiao et al., 2014)
Breast	luciferase reporter assays in breast cells	ERRα	impaired proliferation and migration of cancer cells	(Lee et al., 2015b; Zhao et al., 2012)
Osteo-sarcoma	luciferase reporter assays in human tissue	EZH2 mRNA	reduced cell viability and invasion	(Li et al., 2016)
Gastric	gastric cancer tissue	Cdc42	Apoptosis, cell cycle G1 arrest	(Chen et al., 2011a)
Oral	western blot, reporter assays	BRD4	reduced cell proliferation, invasion and migration	(Langevin et al., 2010)
Colorectal	luciferase reporter assay, methylation-specific PCR, immune blotting in cell model, xenograft animal model	Aurora A, MSI1	inhibition of cell proliferation and tumour formation, apoptosis, G2/M arrest	(Huang et al., 2016; Smith et al., 2015)
Lung	luciferase activity in lung cancer cells and tissue (in vivo and in vitro)	NUCKS1, Cdc42, CdK6	inhibition of cell proliferation and migration, cell cycle arrest in G1 phase, chemotherapy, inhibition of angiogenesis	(Shen et al., 2016), (Zhu et al., 2013)
Ovarian	ovarian cancer cell line	AEG-1	reduced in vitro cell growth, clonogenicity, cell cycle arrest (G1 phase)	(Guo et al., 2013)
Malignant melanoma	Bioinformatics and S in vivo/vitro labelling analysis for dynamic proteomics (SiLAD) in melanoma cells	PAK2	inhibition of cell proliferation	(Hao et al., 2015)
Neuro-blastoma	luciferase reporter assays in SHEP and HEK293 cells	KDM1A mRNA	apoptosis , decreased cell viability and proliferation	(Althoff et al., 2013)

The association of VNTR polymorphisms in carriers of pathogenic variants of the *BRCA1/2* gene has been suggested as a modifier of ovarian and breast cancer risk (Cui et al., 2016; Phelan et al., 1996). Variations in a VNTR located in the promoter of the XRCC5 gene showed differential reporter gene expression and associations between BRCA1/2 groups and specific repeat numbers by genotype analysis (Cui et al., 2016).

My study of the *MIR137* VNTR in breast cancer has therefore been based on past studies which associate the role of MIR137 as a tumour suppressor in cancer including breast and the differential reporter gene expression of the MIR137 by variations in the VNTR repeat length shown by our group and others (Bemis et al., 2008; Warburton et al., 2015). Bemis *et al* reported this VNTR to contain two CpG dinucleotides and that variation in number of this 15 bp repeat to be associated with CpG island length and modulation of miR-137 function in a melanoma cell line (Bemis et al., 2008). The variations in the VNTR could affect the susceptibility of CpG islands to become hypermethylated or to be protected from this change (Ronneberg et al., 2008).
A role for miR137 as a tumour suppressor is evident in several malignancies including, breast cancer (Lee et al., 2015b), pancreatic cancer (Xiao et al., 2014), osteosarcoma cancer (Li et al., 2016), gastric cancer (Chen et al., 2011a), oral cancer (Langevin et al., 2010), ovarian cancer (Guo et al., 2013), lung cancer (Zhu et al., 2013) and in colon cancer (Smith et al., 2015). miR137 is also involved in several significant pathways in melanoma development and progression (Hao et al., 2015).
Downregulation of miR137 by promoter hypermethylation has been described in a variety of cancers and is believed to significantly contribute to carcinogenesis. miR137 has also been shown to be downregulated in neuroblastoma (Chen and Stallings, 2007) and high-grade astrocytic tumours further supporting its role in carcinogenesis (Godlewski et al., 2008). miR137 may therefore have a protective role with regards to breast cancer (Lee et al., 2015b).

The *MIR137* VNTR is located at the internal promoter of the MIR137 host gene and consists of a 15 base pair sequence containing two CpG dinucleotides (Warburton, 2015).

DNA methylation is a key feature of transcriptional regulation. In particular, methylation of gene promoters regulate transcription (Bird, 1986). High levels of DNA methylation in CpG islands located in promoter regions are typically associated with gene silencing and vice versa (Merlo et al., 1995; Veigl et al., 1998). CpG islands are variably methylated however aberrant methylation at some CpG sites are linked with disease, in particular with cancer (Portela and Esteller, 2010; Robertson and Wolffe, 2000).

The importance of epigenetics in cancers has been described for altered DNA methylation and histone modification. In cancer cells entire chromosomal segments may be densely packed as heterochromatin. Genes in these regions are inaccessible for the translation apparatus which consecutively leads to silencing of genes encoded on these stretches in the genome. Silencing by altered packing has been described for tumour suppressor genes and their target genes (Soria et al., 2010; Zhu et al., 2011). DNA methylation during replication is maintained by DNA methyl transferase (DNMT)1 whereas DNMT3a & DNMT3b are involved in *de novo* methylation (Paska and Hudler, 2015; Sorm and Vesely, 1968). DNA methylation has been established as diagnostic, prognostic and predictive tumour markers (Kristensen and Hansen, 2009; Mikeska et al., 2012).

It is now also accepted that altered DNA methylation pattern may function as biomarkers for early detection of cancer. For instance in breast cancer *BRCA1* promoter hypermethylation detected in circulating DNA was associated with an increased risk for early onset of breast cancer (Wong et al., 2011). Epigenetic regulation of MIR137 through promoter hypermethylation has been described in a variety of cancers (Min et al., 2018; Vrba et al., 2013; Zhang et al., 2018) (table 4.1). The MIR137 internal promoter is located adjacent to a CpG island within the MIR137-HG (Warburton et al., 2015). The effect of the length of the VNTR on the CpG island was not addressed in this study. Varying lengths of CpG islands in promoter regions of tissue-specific genes are associated with

modulation of gene expression. Longer CpG islands appear to contain a high number of transcriptional start sites which may be responsible for regulation of the gene (Elango and Yi, 2011).

4.2 Aims

The aim of this chapter is to explore whether the MIR137 VNTR could be a potential biomarker for breast cancer risk. I therefore performed genotype analysis of the MIR137 VNTR by PCR in a female breast cancer cohort and matched controls.

4.3 Results

4.3.1 Four copy MIR137 VNTR homozygosity is less commonly observed in *BRCA1/2* patients

Genotype analysis was performed by simple observation in all four patient groups. Details of the groups are outlined in section 2.1.1. Allele calls were achieved for 40 samples in the BRCA1, 49 in the BRCA2, 97 in the BRCA negative and 98 in the control group. In the breast cancer cohort I observed 3 to 12 repeats of the MIR137 VNTR (Figure 4.2). Initial inspection of the genotype figures showed that homozygosity for the 4 copy variant of the VNTR was the most frequent finding within all of the groups. The presence of the larger repeat number (12 copies) was rare (Figure 4.3 A; Table 4.2).

The biallelic four copy VNTR of MIR137 was significantly less frequently observed in the BRCA1 cohort when compared to BRCA negative breast cancer patients, BRCA2 mutation carriers and healthy controls (two-way ANOVA) (Figure 4.3 B). Slightly more subjects were heterozygous for the 4 copy and larger (10-12) copies within the BRCA1 cohort (7/40 (17.50%) subjects) compared to the control, BRCA negative and BRCA2 groups (2/98 (2.04%), 2/97 (2.06%) and 0/49 (0%) respectively). This was only a trend and not found to be statistically significant (Table 4-2; Figure 4-3B)

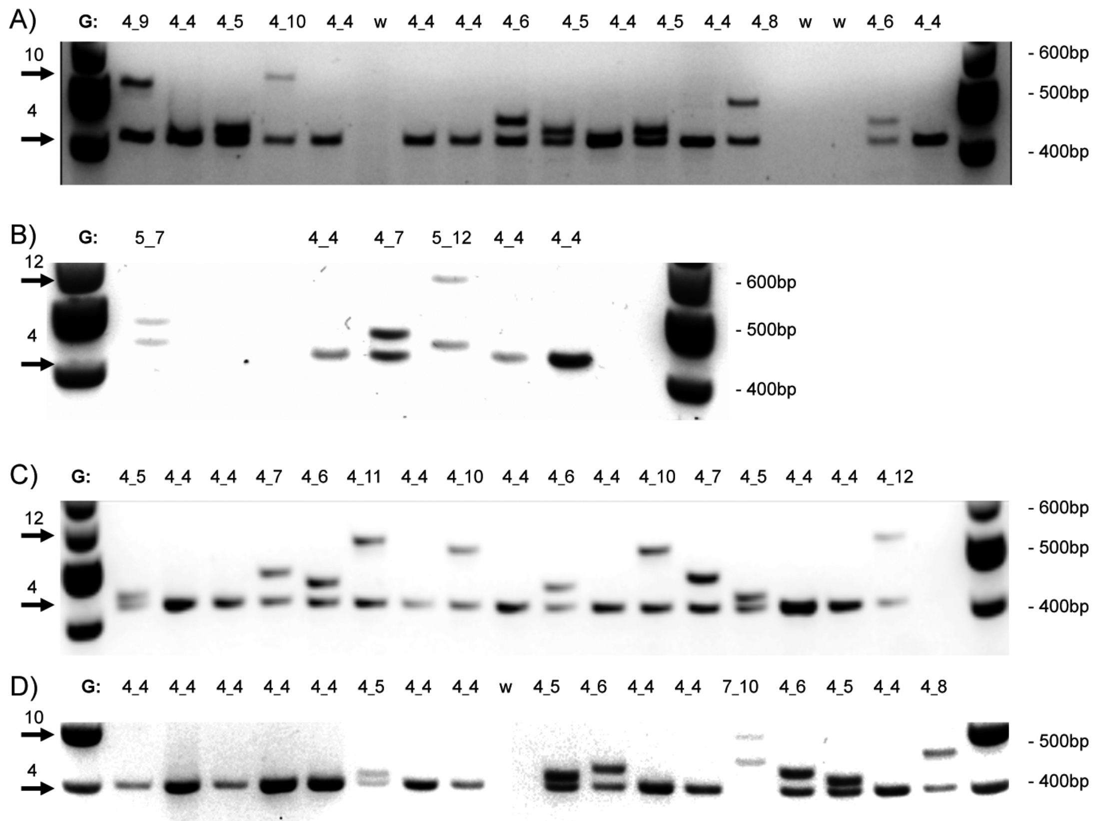


Figure 4-2 Genotyping of MIR137 VNTR PCR in a breast cohort: DNA of healthy controls, BRCA negative patients and BRCA1/2 positive individuals were amplified using PCR. A 2% gel was used to separate amplicons and a 100 bp DNA ladder (on right of gel image) was used to determine size. Genotype numbers represent number of repeats of the MIR137 VNTR. A) Example of genotype in the control group, B) Example of genotype in the BRCA negative group, C) Example of genotype in the BRCA1 positive group, D) Example of genotype in the BRCA2 positive group. w:water control.

Table 4-2: MIR137 VNTR Genotype analysis. Twelve different alleles were observed where 3 is the shortest and 12 the longest. Genotype designates the two alleles found in a sample. The allele numbering corresponds to the number of 15 bp VNTR repeats and are numbered 3-12.
Frequency distribution of observed patient genotype defined by the two alleles listed in the first column (x_y).

Genotype	BRCA1	BRCA2	BRCA neg	control
3_4	0	0	0	1 (1.02%)
3_7	0	1 (2.04%)	0	0
4_4	15 (37.5%)	22 (44.8%)	52 (53.6%)	51 (52.0%)
4_5	5 (12.5%)	8 (16.3%)	11 (11.3%)	9 (9.18%)
4_6	5 (12.5%)	5 (10.2%)	12 (12.3%)	9 (9.18%)
4_7	4 (10%)	7 (14.2%)	4 (4.12%)	9 (9.18%)
4_8	1 (2.5%)	2 (4.08%)	4 (4.12%)	6 (6.12%)
4_9	2 (5%)	1 (2.04%)	6 (6.18%)	3 (3.06%)
4_10	4 (10%)	0	2 (2.06%)	1 (1.02%)
4_11	1 (2.5%)	0	0	1 (1.02%)
4_12	2 (5%)	0	0	0
5_5	1 (2.5%)	0	0	0
5_6	0	0	0	3 (3.06%)
5_7	0	0	1 (1.03%)	1 (1.02%)
5_8	0	0	1 (1.03%)	1 (1.02%)
5_9	0	0	1 (1.03%)	0
5_12	0	1 (2.04%)	0	0
6_8	0	0	1 (1.03%)	0
6_9	0	0	1 (1.03%)	1 (1.02%)
6_11	0	1 (2.04%)	0	0
7_10	0	1 (2.04%)	0	2 (2.04%)
8_10	0	0	1 (1.03%)	0
Total	40	49	97	98

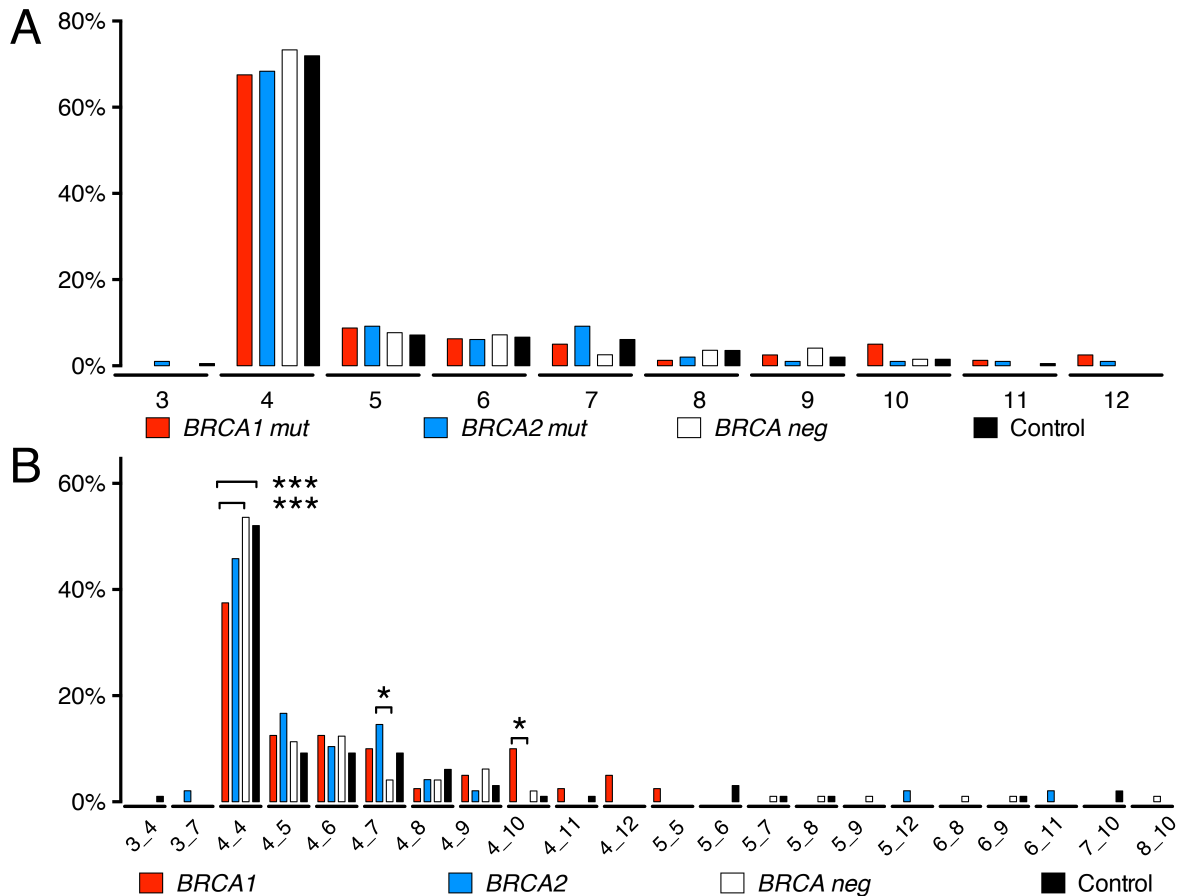


Figure 4-3: Graph of allele frequency percentages of the VNTR in the MIR137 promoter. A) depicts the percentage frequency distribution of individual alleles in each group. B) shows the percentage of allele pairs (genotype) observed in each group listed in table 4.2. Significant variation of distribution between groups (two-way ANOVA: * $p \leq 0.05$; *** $p \leq 0.001$). Detailed allele pairs detected in BRCA1/2 positive, BRCA negative and control samples. BRCA1 N=40; BRCA2 N=49; BRCA negative N=97 and Controls N=98.

Table 4-3: MIR137 genotype of BRCA1 and 2 positive individuals with and without cancer. Frequency distribution of observed patient genotype defined by the two alleles listed in the first column (x_y). The allele numbers correspond to the number of copies of the 15 bp VNTR and are numbered 3-12.

Allele pair	BRCA1 CA	BRCA1 no CA	BRCA2 CA	BRCA2 no CA	control
3_4	0	0	0	0	1 (1.02%)
4_4	6 (40%)	9 (36%)	10 (43.4%)	12 (46.1%)	51 (52.0%)
4_5	2 (13.3%)	3 (12%)	3 (13.0%)	5 (19.2%)	9 (9.18%)
5_5	0	1 (4%)	0	0	0
4_6	2 (13.3%)	3 (12%)	2 (8.69%)	3 (11.5%)	9 (9.18%)
5_6	0	0	0	0	3 (3.06%)
3_7	0	0	0	1 (3.84%)	0
4_7	1 (6.66%)	3 (12%)	4 (17.3%)	3 (11.5%)	9 (9.18%)
5_7	0	0	0	0	1 (1.02%)
4_8	0	1 (4%)	1 (4.34%)	1 (3.84%)	6 (6.12%)
5_8	0	0	0	0	1 (1.02%)
6_8	0	0	0	0	0
4_9	1 (6.66%)	1 (4%)	1 (4.34%)	0	3 (3.06%)
5_9	0	0	0	0	0
6_9	0	0	0	0	1 (1.02%)
4_10	1 (6.66%)	3 (12%)	0	0	1 (1.02%)
7_10	0	0	1 (4.34%)	0	2 (2.04%)
8_10	0	0	0	0	0
6_11	0	0	0	1 (3.84%)	0
4_11	0	1 (4%)	0	0	1 (1.02%)
4_12	2 (13.3%)	0	0	0	0
5_12	0	0	1 (4.34%)	0	0
Total	15	25	23	26	98

4.3.2 Correlation of MIR137 VNTR with *BRCA1* and *BRCA2* pathogenic germline variant carriers and the occurrence of breast cancer in these patients

Clump analysis (section 2.2.10, materials and methods chapter) was performed to maximise the chi-squared value of the multiple rare allele copy numbers (Table 4.4). Significance was reached when BRCA1 and BRCA2 positive groups were combined and compared with the BRCA negative group (p=0.035). Statistical significance was also observed when the difference in allele frequency was compared between *BRCA1* and *BRCA2* mutation carriers (p=0.008). As it stands because of the small sample size these results will

have to be interpreted with caution. No further significance was reached when BRCA1 and 2 groups were further divided into individuals with and without cancer (Table 4.3). Again a larger cohort is required for future work on these groups.

The data above shows a trend towards the presence of higher copy numbers (10-12 copies) of the VNTR in the BRCA1 positive group. This group was further divided into individuals with and without breast cancer for further comparison (table 4.3). Clump analysis was performed on individual BRCA1 and 2 groups, on pooled data of BRCA1 and BRCA2 (i.e. BRCA positive) for all with and without cancer (table 4.4). None of the comparisons reached statistical significance. Some results within this table had a smaller sample mean compared to the hypothesized mean therefore a p value could not be calculated for these groups.

Table 4-4: Clump analysis for MIR137 VNTR genotype in a breast cohort

Sample	T1	T2	T3	T4
Control vs BRCA negative	0.849	0.430	0.151	0.633
Control vs BRCA1/2 positive	0.550	0.402	0.173	0.553
Control vs BRCA1 positive	0.540	-	0.121	0.375
Control vs BRCA2 positive	0.919	0.215	0.202	0.838
BRCA negative vs BRCA1/2 positive	0.492	0.052	0.035	0.187
BRCA negative vs BRCA1 positive	0.569	-	0.097	0.102
BRCA negative vs BRCA2 positive	0.621	0.076	0.351	0.258
BRCA2 positive vs BRCA1 positive	0.873	0.008	0.481	0.167
Control vs BRCA1/2 positive w/o CA	0.786	0.501	0.208	0.621
Control vs BRCA1/2 positive with CA	0.908	-	0.298	0.826
Control vs BRCA1 positive w/o CA	0.793	-	0.152	0.639
Control vs BRCA1 positive with CA	0.629	-	0.385	0.175
Control vs BRCA2 positive w/o CA	0.913	-	0.594	0.500
Control vs BRCA2 postive with CA	0.995	-	0.460	0.919
BRCA1/2 positive w/o CA vs BRCA1/2 positive with CA	0.988	0.520	0.738	1.000
BRCA1/2 positive w/o CA vs BRCA1 positive with CA	0.988	-	0.935	0.470
BRCA1/2 positive w/o CA vs BRCA2 positive with CA	0.995	-	0.853	1.000
BRCA1/2 positive withCA vs BRCA1 positive w/o CA	0.997	-	0.628	0.714
BRCA1/2 positive with CA vs BRCA2 positive w/o CA	0.996	-	0.748	0.729
BRCA1 positive w/o CA vs BRCA1 positive with CA	1.000	-	0.800	1.000
BRCA1 positive w/o CA vs BRCA2 positive w/o CA	0.994	-	0.461	0.403
BRCA1 positive w/o CA vs BRCA2 positive with CA	0.999	-	0.597	0.727
BRCA1 positive with CA vs BRCA2 positive w/o CA	0.992	-	0.702	0.278
BRCA1 positive with CA vs BRCA2 positive with CA	0.998	-	0.832	0.657
BRCA2 positive w/o CA vs BRCA2 positive with CA	1.000	-	0.851	0.905

Legend for table: Clump analysis using the Monte Carlo approach

Significance was reached when *BRCA* negative was compared with the *BRCA* positive (*BRCA1* and *BRCA 2* cohorts were grouped together). P=0.035 (evidence). This gives an association with the MIR137 VNTR when BRCA positive is compared with BRCA negative. T3 (2x2 table constructed by comparing each non-rare column against the total of all remaining columns.), 10,000 simulations, chi-squared = 4.43, 1 degree of freedom (df). Further significance was associated with the BRCA1 cohort when compared with the BRCA2 group. P=0.008 (strong evidence). T2 (table compiled from an amalgamation of columns containing rare alleles), 10,000 simulations, chi-squared=7.068, 1 df. The sections within the table that do not contain a *p* value are groups where the analysis could not calculate such values due to the small size of those groups.

Abbreviations within table: w/o- without; CA- cancer (with reference to breast cancer); BRCA1 / 2 positive- BRCA1 and BRCA 2 cohorts grouped together.

4.3.2 MIR137 VNTR modifies CpG island length *in silico*

In order to establish whether the VNTR copy number may influence the length of the CpG island which extends into *MIR137* an *in silico* analysis was performed. Sequences including 5 kb upstream of the first copy of the VNTR and 5 kb downstream of *MIR137* with four or twelve repeats of the VNTR were entered into EMBOSS CpGplot (Rice et al., 2000). A higher VNTR copy number gave a longer predicted CpG island, as would be expected. A predicted CPG island of 262 bp was calculated for the 4 copy VNTR whereas twelve repeats extended the predicted CpG island to 382 bp (Figure 4.4).

A second simple nine base pair VNTR containing one CpG dinucleotide is located downstream of *MIR137* in the CpG island in this region (RepeatMasker v4.0.7 Dfam_2.0 chr1:98045853-98045903). The influence of this repeat on the CpG island was not further investigated.

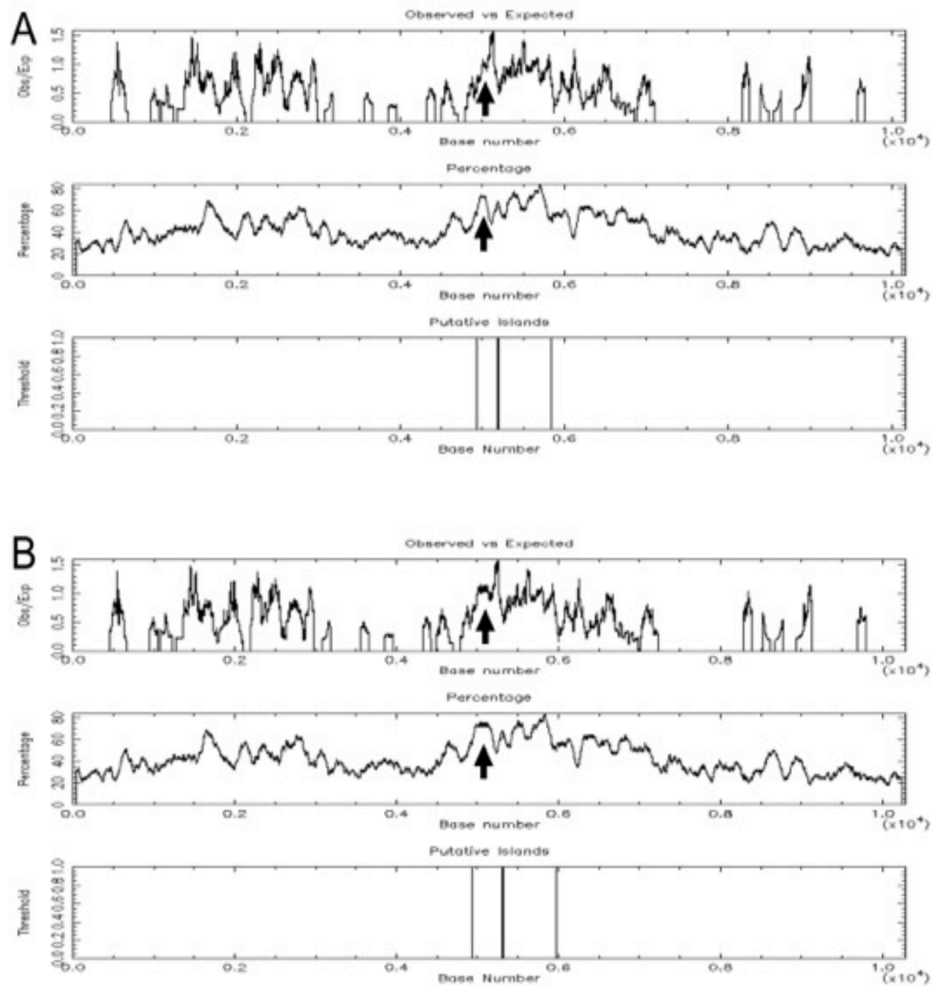


Figure 4-4: *in silico* CpG island prediction in MIR137: in silico CpG island prediction by EMBOSS Cpgplot (Rice et al., 2000) of the MIR137 sequence including 5kb upstream of the first VNTR and downstream of the MIR137 coding sequence. A) MIR137 with four copies of the VNTR and B) twelve copies of the VNTR. First row: observed vs expected CpG; second row: percentage of CpGs; third line CpG site. Arrows indicate the region of the MIR137 VNTR.

4.4 Discussion

I have addressed the repeat length variation of the VNTR in the internal promoter upstream of *MIR137 (ImiR137)*. *In silico* analysis showed that the copy number variation of the *MIR137* VNTR leads to an elongation of the predicted CpG island. This expansion of the CpG island as a result of a higher copy number of the VNTR which contains two CpG dinucleotides may be a risk predictor for carcinogenesis in carriers of pathogenic *BRCA1/2* germline variants. A longer CpG island may predispose to hypermethylation of the *ImiR137* promoter and a subsequent down regulation of miR137 and therefore of a loss of its tumour suppressor function.

When addressing the VNTR allele types between our four patient cohorts we observed that the *BRCA1* germline variant carriers showed a trend towards *MIR137* VNTR alleles with higher copy numbers. This might be associated with a down regulation of miR137 as outlined above especially as the mechanisms of penetrance for *BRCA1/2* pathogenic germline variant carriers with regards to development of breast cancer are still not entirely understood. A longer VNTR in the promoter of *MIR137* may therefore be an additional risk factor for developing breast cancer in *BRCA1/2* pathogenic germline variant carriers especially in the case of a loss of short allele as with loss of heterozygosity. The biallelic presence of the 4 copies of the VNTR was more frequently observed in the BRCA negative and the control group highlighting that these two groups may be less prone to a loss of miR137 expression related to *ImiR137* promoter hypermethylation. Somatic miR137 promoter methylation analysis in BRCA1/2 and BRCA negative cancer patients would therefore be of interest.

An association was observed between heterozygous higher copy numbers (10-12 copies) and 4 copies in *BRCA1* mutation carriers (table 4.2, figure 4.2). This observation however was similar in *BRCA1* mutation carriers with and without a cancer diagnosis. Significance was not achieved on statistical analysis of the MIR137 VNTR and an association with cancer diagnosis using clump analysis.

In terms of clinical impact the results of the BRCA1/2 positive groups with and without breast cancer are of particular interest as it could give an insight into

the effect of various repeat numbers on BRCA1/2 mutation penetrance. A larger study group in future work may give a more complete picture of BRCA1/2 penetrance if more individuals with breast cancer are found within the *BRCA1/2* mutation carriers. Therefore if an association between specific VNTR lengths of the *lmiR137* and breast cancer are found in *BRCA1/2* mutation carriers in future work then a biomarker to predict risk of breast cancer in this group could be a possibility.

Functional analysis was also performed *in vitro* using reporter gene constructs to address the action of the MIR137 VNTRs as a regulatory domain by Mamdani an coworkers (Mamdani et al., 2013) and within our unit (Warburton et al., 2016a). Reporter gene analysis of the internal promoter demonstrated distinct VNTR repeat number supported expression in a neuroblastoma (SH-SY5Y) cell line. The rare 12 copy variant was found to be functionally distinct from the 4 copy variant (Mamdani et al., 2013; Warburton et al., 2016a). These findings support further work to explore the potential role of miR137 in breast cancer and may link the effects of *MIR137* silencing in breast and other cancers (table 4.1) through allele specific reporter gene activity, *MIR137* promoter methylation analysis and transcription levels. Other groups have also found repeat number variation (3 and 12 copies) to differentially alter the processing of mature miR137 in a melanoma cell line model (Bemis et al., 2008).

In conclusion VNTR copy number in the lmiR137 significantly differentiates between BRCA1/2 germline variant carriers and BRCA wild type individuals. However the observed lmiR137 VNTR copy number length did not differentiate between the cancer affected and tumour free BRCA1/2 individuals. Patient data and sample numbers were limited especially for the *BRCA* positive groups. The clinical significance of any future finding in a breast cancer group would be strengthened in a larger cohort with more clinical details.

The 4 copy allele was the most frequently occurring VNTR in all groups and the 12 copy was rare. This mirrors the distribution of allele frequencies observed in a schizophrenia cohort investigated in our laboratory. Twelve copy alleles were seen in a heterozygous combination with lower copy numbers (mainly with 4 copies). Twelve copies may be linked with cancer as the CpG island within the promoter region is increased in length with the presence of 12 copies. Longer CpG islands are associated with an increased likelihood of promoter hypermethylation and therefore an expansion of a CpG island may relate to silencing of a tumour suppressor gene (Elango and Yi, 2011).

The hypothesis that the MIR137 VNTR may serve as a potential modifier for prediction of breast cancer in patients with pathogenic *BRCA1/2* germline variants could not be verified in this study. However the results of this pilot study would suggest that further investigation for utility in a larger cohort should be conducted.

5 Final discussion

Women have a 12% lifetime risk of developing breast cancer and this increases up to 80% in individuals carrying a *BRCA1* or *BRCA2* germline mutation (Rebbeck and Domchek, 2008). Screening for *BRCA1/2* mutations is well established for prediction of familial breast and ovarian cancer risk. The penetrance of these mutations however are incomplete (Kuchenbaecker et al., 2017). Other additional mutations and genetic variations may influence the level of penetrance as well as modifying the risk of developing breast cancer. One of the aims of my study was to explore the role of polymorphism in two candidate genes that may influence breast cancer progression.

In both the MIR137 and AAAG VNTR studies associations were found mainly with specific VNTR variants and BRCA1/2 mutation carriers and in particular with females with cancer who are BRCA1 mutation carriers (AAAG VNTR). Breast cancer in *BRCA1* mutation carriers can be difficult to treat as their cancers tend to be oestrogen receptor negative and therefore unlikely to respond to hormone therapy. They also present at a younger age often with a more aggressive tumour type. These women may benefit not only from more accurate prediction for cancer risk but also from even closer surveillance or risk reducing surgery at a younger age. Choosing a treatment option is a difficult decision to make for a multitude of reasons as previously outlined (section 1.5). The discovery of additional biomarkers to aid further risk determination is of great importance especially in the younger age group of BRCA positive individuals who may choose surveillance over risk reducing surgery (Johns et al., 2017). My study therefore highlights particularly promising discoveries where replication in a larger cohort is justified.

In a reporter gene assay addressing the AAAG VNTR copy number length in *ESRRG* I found that oestrogen plays an important role in transcription regulation, which supports the important role of GxE in breast cancer. This may give further insight into the understanding of the biology of oestrogen related

carcinogenesis in breast tissue. This is significant in relation to women and their varying relationship and response to endogenous and exogenous sex hormones. Variations in breast cancer risk in general and oestrogen exposure in different ethnic groups are likely to be due to cultural differences, genetic variations and differing sensitivities to oestrogen (Gilliland et al., 1998; Hall et al., 2005; Rebbeck and Domchek, 2008). There is even more variability in reports of breast cancer risk amongst individuals with identical *BRCA1/2* mutations (Rebbeck and Domchek, 2008; Xie et al., 2018). Factors which can increase the risk of developing breast cancer due to increased exposure to oestrogen are very individual but may mirror the socioeconomic and cultural environment. This may also explain the differences in data for AAAG repeat number variation described in Iranian, American and British (this study) cohorts.

The mechanisms involved in VNTR driven reporter gene expression in *MIR137* and *ESRRG* are still not clear. In 2011 in a similar study the *ESRRG* gene was described to have two isoforms which were thought to result in alternate regulation of the gene under the influence of the varying length of the VNTR. Today eight transcription start sites and two additional isoforms have been found giving further possibilities for mechanism of action. This was also described with the parathyroid hormone gene where a variance of the VNTR co-varies with height (Scillitani et al., 2006). There are several candidate transcription factors and transcription start sites that could bind to the gene in the VNTR region of *ESRRG* based on repeat length. One transcription factor which has been suggested to bind directly to the AAAG VNTR is PAX2 (Galindo et al., 2011). PAX2 has also been shown to bind to the ER and to supress *ERRB* transcription. This mechanism was found when the interaction of PAX2 with HER2 was investigated in relation to tamoxifen resistance (Hurtado et al., 2008). This could explain a possible mechanism for how variance in the AAAG VNTR of *ESRRG* and its gene product are involved in oestrogen signalling pathways. Further investigation to explore possible mechanisms of action is required.

Discovery of genetic variants to date is just the tip of the iceberg which represents the human genome and disease processes involved. Even more important than the identification of genetic polymorphisms is the continuing progress in technological capabilities to make these discoveries. Many genetic variants discovered so far remain of uncertain clinical relevance with regards to their risk related to cancer development. This might result from a selection bias of patients involved in these studies (i.e. recruited from specialist clinics) and therefore some variants may be associated with an over estimated cancer risk (Easton et al., 2015). A significant number of studies have also been performed on European women and this raises doubt as to whether relative risks calculated in this population can be applied to other nationalities who may have differences in breast cancer incidence. Genetic counselling and risk prediction is difficult and also requires consideration of non-genetic factors such as ethnicity as well as lifestyle and hormonal exposure.

Testing for multiple mutations can be informative in updating risk prediction scoring (Easton et al., 2015). In a GWAS study on *BRCA1* mutation carriers Couch and colleagues demonstrated that the absolute risk was significantly modified compared to published figures when they identified new loci associated with breast cancer risk (Couch et al., 2013). Panel testing of multiple genes are currently in clinical use and can be quite informative especially when combined with clinical and epidemiological data. With the falling costs and increased speed of whole exome and whole genome sequencing (WES / WGS) and the improvement of bioinformatics capacities required for analysis of the large data generated the identification of individuals with increased cancer risk and disease penetrance will become identifiable. A concern with such thorough analysis is the identification of several germline variants of uncertain significance which may be difficult to interpret in the clinical setting unless combined with other mutations to give a more meaningful risk prediction (Easton et al., 2015; Evans et al., 2017). We are already entering the next generation of medical practice with the 100,000 Genomes project (Public Health England, 2017b). This is a project which has been set up

by the NHS with the aim of transforming the way we diagnose and deliver patient care. The target set is to sequence 100,000 genomes from approximately 70,000 patients with cancer, infectious and rare diseases as well as their families. These specific groups have been selected as such advances in genomic technology would benefit these patients and the NHS the most. This project and genome sequencing is already leading to new and more accurate diagnoses as well as changes in treatment provision, in other words more stratification of patient care.

Whole exome sequencing is cheaper than whole genome sequencing and can therefore be applied to a broader cohort. Even though WES will not cover the same extend as WGS it is a powerful tool and a suitable investigation for discovery of novel genes and for mutation analysis. Niguidula et al performed a survey which assessed the effect of WES data on clinical decision making and healthcare impact. They found that WES data directly lead to changes in treatment management, cessation of unnecessary diagnostic testing, recruitment of patients into appropriate clinical trials and referral of patients to more appropriate specialist care based on their diagnosis. It is of note that 90% of patients also opted for further testing for secondary/incidental findings which was beneficial for testing potential family members who were genetic mutation carriers. In patients with a definitive diagnosis derived from WES unnecessary, expensive, time consuming, potentially invasive testing was stopped. Even though the single test may appear expensive the financial overall benefit from thorough testing may reduce the overall costs (Niguidula et al., 2018).

The findings of this project and related work performed by others within our group has revealed further possibilities for future work to continue the search for novel biomarkers for the prediction of breast cancer risk. Investigating the tumour suppressor RE1 Silencing Transcription Factor (REST) also known as Neuron Restrictive Silencing Factor (NRSF) would be another natural progression from work already performed by our group on breast cancer and neurodegenerative disorders and with the overlap of biology between the two. Our team have found that the internal *ImiR137* promoter can be regulated

partly by REST and miR137 expression can be affected by the genotype of the VNTR in the promoter as well as by environmental factors. This mechanism could play an influential role in schizophrenia (Warburton et al., 2015). Loss of REST has also been associated with a subset of aggressive breast cancers which have a particularly high rate of recurrence and lymph node metastasis. Wagoner and colleagues investigated the expression of REST and it's target genes in aggressive breast cancer. They found loss of function of REST and expression of a truncated REST splice variants associated with increased breast cancer risk. Interestingly this truncated REST isoform is usually found in the brain following epileptic fits (Wagoner et al., 2010).

Cancer is a complex disease which is influenced by genetic variations, epigenetic factors (e.g. DNA methylation) as well as environmental/life style choices such as oestrogen exposure. In order to establish any further biomarkers modifying factors other than genetic variations would need to also be considered such as ethnic background, lifestyle, hormone and reproductive history as well as the age of the cohort.

In conclusion, my findings suggest that VNTRs within *MIR137* and the *ESRRG* could be potential modifiers for prediction of breast cancer risk in *BRCA1* and *BRCA2* pathogenic mutation carriers. The data will require replication with additional significance in a larger cohort before it may be considered for implementation in clinical practice.

Despite new genetic advances within the NHS which will revolutionise medicine we still need to individualise risk by considering modifying factors. This will not only guide the healthcare professionals in offering stratified treatment options but more importantly it will empower patients whilst they make the most important and difficult decisions of their lives. It will be interesting to observe how changes in risk prediction with more accurate risk calculation will affect the clinician's advice to patients and the patient's ultimate decision. Rebbeck and colleagues summarised such uncertainties rather

precisely. ‘Will a mutation carrier make a different decision based on a 40% lifetime risk of breast cancer versus a 60% lifetime risk?’ Will doctors ever give the advice of a mastectomy or the option of surveillance with great certainty? (Rebbeck and Domchek, 2008). Only time will tell how the outcome of these advances will evolve.

6 Figures and Tables

6.1 Figures

2-1: ESRRG AAAG VNTR thermo cycler protocol p. 35
2-2: MIR137 VNTR thermo cycler protocol p. 36
2.3: PGLP construct p. 39
3-1: Gene map of *ESRRG* p. 49
3-2: Genotyping of *ESRRG* AAAG VNTR PCR in a breast cohort p. 52
3-3: Graphs of allele frequency occurrence of the AAAG VNTR p. 54
3-4: Graph of single allele distribution percentage frequencies with BRCA1 and 2 groups subdivided into groups with and without breast cancer p. 54
3-5: Graph of allele pairs distribution percentage frequencies p. 55
3-6: Effect of AAAG copy number repeat on reporter gene expression. p. 57
4-1: *MIR137* gene map p. 65
4-2: Genotyping of MIR137 VNTR PCR in a breast cohort p. 71
4-3: Graph of allele frequency percentages of the VNTR in the MIR137 promoter p. 73
4-4: *In silico* CpG island prediction in MIR137 p. 78

6.2 Tables

1-1: Targets of miR-137 p. 29
2-1: Plasmids p. 44
3-1: Allele pair frequency (including percentages) of the AAAG VNTR in the ERRG gene. p. 53
3-2: Individual allele frequencies (including percentages) of the AAAG VNTR in the ERRG gene. p. 53
3-3: Comparison of AAAG alleles within BRCA1, BRCA2, BRCA -ve and a control groups. p. 55
3-4: Statistical significance between transfection groups. p. 58
4-1: Table of targets for MIR137 and their observed effects in cancer p. 67
4-2: MIR137 VNTR Genotype analysis p. 72
4-3: MIR137 genotype of BRCA1 and 2 positive individuals with and without cancer p. 74
4-4: Clump analysis for MIR137 VNTR genotype in a breast cohort p. 76

7 References

Ahn, S.H., Son, B.H., Yoon, K.S., Noh, D.Y., Han, W., Kim, S.W., Lee, E.S., Park, H.L., Hong, Y.J., Choi, J.J., et al. (2007). BRCA1 and BRCA2 germline mutations in Korean breast cancer patients at high risk of carrying mutations. Cancer Lett 245, 90-95.

Ali, E.S., Mangold, C., and Peiris, A.N. (2017). Estriol: emerging clinical benefits. Menopause 24, 1081-1085.

Allen, N.E., Beral, V., Casabonne, D., Kan, S.W., Reeves, G.K., Brown, A., Green, J., and Million Women Study, C. (2009). Moderate alcohol intake and cancer incidence in women. J Natl Cancer Inst 101, 296-305.

Althoff, K., Beckers, A., Odersky, A., Mestdagh, P., Koster, J., Bray, I.M., Bryan, K., Vandesompele, J., Speleman, F., Stallings, R.L., et al. (2013). MiR-137 functions as a tumor suppressor in neuroblastoma by downregulating KDM1A. Int J Cancer 133, 1064-1073.

Ambros, V. (2004). The functions of animal microRNAs. Nature 431, 350-355.

Antoniou, A.C., Beesley, J., McGuffog, L., Sinilnikova, O.M., Healey, S., Neuhausen, S.L., Ding, Y.C., Rebbeck, T.R., Weitzel, J.N., Lynch, H.T., et al. (2010). Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: implications for risk prediction. Cancer Res 70, 9742-9754.

Antoniou, A.C., Kartsonaki, C., Sinilnikova, O.M., Soucy, P., McGuffog, L., Healey, S., Lee, A., Peterlongo, P., Manoukian, S., Peissel, B., et al. (2011). Common alleles at 6q25.1 and 1p11.2 are associated with breast cancer risk for BRCA1 and BRCA2 mutation carriers. Hum Mol Genet 20, 3304-3321.

Ao, A., Wang, H., Kamarajugadda, S., and Lu, J. (2008). Involvement of estrogen-related receptors in transcriptional response to hypoxia and growth of solid tumors. Proc Natl Acad Sci U S A 105, 7821-7826.

Apostolou, P., and Papasotiriou, I. (2017). Current perspectives on CHEK2 mutations in breast cancer. Breast Cancer (Dove Med Press) 9, 331-335.

Ariazi, E.A., Ariazi, J.L., Cordera, F., and Jordan, V.C. (2006). Estrogen receptors as therapeutic targets in breast cancer. Curr Top Med Chem 6, 181-202.

Ariazi, E.A., Clark, G.M., and Mertz, J.E. (2002). Estrogen-related receptor alpha and estrogen-related receptor gamma associate with unfavorable and favorable biomarkers, respectively, in human breast cancer. Cancer Res 62, 6510-6518.

Ariazi, E.A., and Jordan, V.C. (2006). Estrogen-related receptors as emerging targets in cancer and metabolic disorders. Curr Top Med Chem 6, 203-215.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281-297.

Bartkova, J., Tommiska, J., Oplustilova, L., Aaltonen, K., Tamminen, A., Heikkinen, T., Mistrik, M., Aittomaki, K., Blomqvist, C., Heikkila, P., et al. (2008). Aberrations of the MRE11-RAD50-NBS1 DNA damage sensor complex in human breast cancer: MRE11 as a candidate familial cancer-predisposing gene. Mol Oncol 2, 296-316.

Barton, M.B., West, C.N., Liu, I.L., Harris, E.L., Rolnick, S.J., Elmore, J.G., Herrinton, L.J., Greene, S.M., Nekhlyudov, L., Fletcher, S.W., et al. (2005). Complications following bilateral prophylactic mastectomy. J Natl Cancer Inst Monogr, 61-66.

Beaber, E.F., Buist, D.S., Barlow, W.E., Malone, K.E., Reed, S.D., and Li, C.I. (2014). Recent oral contraceptive use by formulation and breast cancer risk among women 20 to 49 years of age. Cancer Res 74, 4078-4089.

Bebenek, A., and Ziuzia-Graczyk, I. (2018). Fidelity of DNA replication-a matter of proofreading. Curr Genet 64, 985-996.

Becker-Andre, M., Andre, E., and DeLamarter, J.F. (1993). Identification of nuclear receptor mRNAs by RT-PCR amplification of conserved zinc-finger motif sequences. Biochem Biophys Res Commun 194, 1371-1379.

Bemis, L.T., Chen, R., Amato, C.M., Classen, E.H., Robinson, S.E., Coffey, D.G., Erickson, P.F., Shellman, Y.G., and Robinson, W.A. (2008). MicroRNA-137 targets microphthalmia-associated transcription factor in melanoma cell lines. Cancer Res 68, 1362-1368.

Benoit, G., Cooney, A., Giguere, V., Ingraham, H., Lazar, M., Muscat, G., Perlmann, T., Renaud, J.P., Schwabe, J., Sladek, F., et al. (2006). International Union of Pharmacology. LXVI. Orphan nuclear receptors. Pharmacol Rev 58, 798-836.

Berliner, J.L., Fay, A.M., Cummings, S.A., Burnett, B., and Tillmanns, T. (2013). NSGC practice guideline: risk assessment and genetic counseling for hereditary breast and ovarian cancer. J Genet Couns 22, 155-163.

Berrington de Gonzalez, A., Curtis, R.E., Gilbert, E., Berg, C.D., Smith, S.A., Stovall, M., and Ron, E. (2010). Second solid cancers after radiotherapy for breast cancer in SEER cancer registries. Br J Cancer 102, 220-226.

Berrington de Gonzalez, A., and Darby, S. (2004). Risk of cancer from diagnostic X-rays: estimates for the UK and 14 other countries. Lancet 363, 345-351.

Berthois, Y., Katzenellenbogen, J.A., and Katzenellenbogen, B.S. (1986). Phenol red in tissue culture media is a weak estrogen: implications concerning the study of estrogen-responsive cells in culture. Proc Natl Acad Sci U S A 83, 2496-2500.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature 321, 209-213.

Bonkhoff, H., and Berges, R. (2009). The evolving role of oestrogens and their receptors in the development and progression of prostate cancer. Eur Urol 55, 533-542.

Bontekoe, C.J., Bakker, C.E., Nieuwenhuizen, I.M., van der Linde, H., Lans, H., de Lange, D., Hirst, M.C., and Oostra, B.A. (2001). Instability of a (CGG)98 repeat in the Fmr1 promoter. Hum Mol Genet 10, 1693-1699.

Bouchard, M., Grote, D., Craven, S.E., Sun, Q., Steinlein, P., and Busslinger, M. (2005). Identification of Pax2-regulated genes by expression profiling of the mid-hindbrain organizer region. Development 132, 2633-2643.

Boyd, N.F., Dite, G.S., Stone, J., Gunasekara, A., English, D.R., McCredie, M.R., Giles, G.G., Tritchler, D., Chiarelli, A., Yaffe, M.J., et al. (2002). Heritability of mammographic density, a risk factor for breast cancer. N Engl J Med 347, 886-894.

BrSP (2015). Breast Screening Programme, England; Statistics for 2013-14. Health and Social Care Information Centre.

Bueno, M.J., and Malumbres, M. (2011). MicroRNAs and the cell cycle. Biochim Biophys Acta 1812, 592-601.

Bussaglia, E., del Rio, E., Matias-Guiu, X., and Prat, J. (2000). PTEN mutations in endometrial carcinomas: a molecular and clinicopathologic analysis of 38 cases. Hum Pathol 31, 312-317.

Catsburg, C., Miller, A.B., and Rohan, T.E. (2015). Active cigarette smoking and risk of breast cancer. Int J Cancer 136, 2204-2209.

Chan, C.H.T., Munusamy, P., Loke, S.Y., Koh, G.L., Wong, E.S.Y., Law, H.Y., Yoon, C.S., Tan, M.H., Yap, Y.S., Ang, P., et al. (2017). Identification of Novel Breast Cancer Risk Loci. Cancer Res 77, 5428-5437.

Chen, Q., Chen, X., Zhang, M., Fan, Q., Luo, S., and Cao, X. (2011a). miR-137 is frequently down-regulated in gastric cancer and is a negative regulator of Cdc42. Dig Dis Sci 56, 2009-2016.

Chen, W.Y., Rosner, B., Hankinson, S.E., Colditz, G.A., and Willett, W.C. (2011b). Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk. JAMA 306, 1884-1890.

Chen, Y., and Stallings, R.L. (2007). Differential patterns of microRNA expression in neuroblastoma are correlated with prognosis, differentiation, and apoptosis. Cancer Res 67, 976-983.

Clarke, R., Leonessa, F., Welch, J.N., and Skaar, T.C. (2001). Cellular and molecular pharmacology of antiestrogen action and resistance. Pharmacol Rev 53, 25-71.

Clavel-Chapelon, F., and Gerber, M. (2002). Reproductive factors and breast cancer risk. Do they differ according to age at diagnosis? Breast cancer research and treatment 72, 107-115.

Colditz, G.A., Bohlke, K., and Berkey, C.S. (2014). Breast cancer risk accumulation starts early: prevention must also. Breast cancer research and treatment 145, 567-579.

Colditz, G.A., and Frazier, A.L. (1995). Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. Cancer Epidemiol Biomarkers Prev 4, 567-571.

Collaborative Group on Hormonal Factors in Breast, C. (1996). Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Lancet 347, 1713-1727.

Couch, F.J., Nathanson, K.L., and Offit, K. (2014). Two decades after BRCA: setting paradigms in personalized cancer care and prevention. Science 343, 1466-1470.

Couch, F.J., Wang, X., McGuffog, L., Lee, A., Olswold, C., Kuchenbaecker, K.B., Soucy, P., Fredericksen, Z., Barrowdale, D., Dennis, J., et al. (2013). Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS Genet 9, e1003212.

Coward, P., Lee, D., Hull, M.V., and Lehmann, J.M. (2001). 4-Hydroxytamoxifen binds to and deactivates the estrogen-related receptor gamma. Proc Natl Acad Sci U S A 98, 8880-8884.

Crago, A.M., Azu, M., Tierney, S., and Morrow, M. (2010). Randomized clinical trials in breast cancer. Surg Oncol Clin N Am 19, 33-58.

CRUK (2017). Cancer Statistics UK.

Cui, J., Luo, J., Kim, Y.C., Snyder, C., Becirovic, D., Downs, B., Lynch, H., and Wang, S.M. (2016). Differences of Variable Number Tandem Repeats in XRCC5 Promoter Are Associated with Increased or Decreased Risk of Breast Cancer in BRCA Gene Mutation Carriers. Front Oncol 6, 92.

Cullen, B.R. (2004). Transcription and processing of human microRNA precursors. Mol Cell 16, 861-865.

Cuzick, J., Warwick, J., Pinney, E., Duffy, S.W., Cawthorn, S., Howell, A., Forbes, J.F., and Warren, R.M. (2011). Tamoxifen-induced reduction in mammographic density and breast cancer risk reduction: a nested case-control study. J Natl Cancer Inst 103, 744-752.

Cybulski, C., Wokolorczyk, D., Jakubowska, A., Huzarski, T., Byrski, T., Gronwald, J., Masojc, B., Deebniak, T., Gorski, B., Blecharz, P., et al. (2011). Risk of breast cancer in women with a CHEK2 mutation with and without a family history of breast cancer. J Clin Oncol 29, 3747-3752.

Deblois, G., and Giguere, V. (2013). Oestrogen-related receptors in breast cancer: control of cellular metabolism and beyond. Nat Rev Cancer 13, 27-36.

Dontu, G., El-Ashry, D., and Wicha, M.S. (2004). Breast cancer, stem/progenitor cells and the estrogen receptor. Trends Endocrinol Metab 15, 193-197.

Dufault, M.R., Betz, B., Wappenschmidt, B., Hofmann, W., Bandick, K., Golla, A., Pietschmann, A., Nestle-Kramling, C., Rhiem, K., Huttner, C., et al. (2004). Limited relevance of the CHEK2 gene in hereditary breast cancer. Int J Cancer 110, 320-325.

Dyrskjot, L., Ostenfeld, M.S., Bramsen, J.B., Silahtaroglu, A.N., Lamy, P., Ramanathan, R., Fristrup, N., Jensen, J.L., Andersen, C.L., Zieger, K., et al. (2009). Genomic profiling of microRNAs in bladder cancer: miR-129 is associated with poor outcome and promotes cell death in vitro. Cancer Res 69, 4851-4860.

Eagon, P.K. (2010). Alcoholic liver injury: influence of gender and hormones. World J Gastroenterol 16, 1377-1384.

Early Breast Cancer Trialists' Collaborative, G. (2005). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365, 1687-1717.

Early Breast Cancer Trialists' Collaborative, G., Davies, C., Godwin, J., Gray, R., Clarke, M., Cutter, D., Darby, S., McGale, P., Pan, H.C., Taylor, C., et al. (2011). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet 378, 771-784.

Easton, D.F. (1999). How many more breast cancer predisposition genes are there? Breast cancer research : BCR 1, 14-17.

Easton, D.F., Pharoah, P.D., Antoniou, A.C., Tischkowitz, M., Tavtigian, S.V., Nathanson, K.L., Devilee, P., Meindl, A., Couch, F.J., Southey, M., et al. (2015). Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med 372, 2243-2257.

Eccles, S.A., Aboagye, E.O., Ali, S., Anderson, A.S., Armes, J., Berditchevski, F., Blaydes, J.P., Brennan, K., Brown, N.J., Bryant, H.E., et al. (2013). Critical research gaps and translational priorities for the successful prevention and treatment of breast cancer. Breast cancer research : BCR 15, R92.

EHBCC-Group, E., Hormones, Breast Cancer Collaborative, G., Key, T.J., Appleby, P.N., Reeves, G.K., Roddam, A.W., Helzlsouer, K.J., Alberg, A.J., Rollison, D.E., Dorgan, J.F., et al. (2011). Circulating sex hormones and breast cancer risk factors in postmenopausal women: reanalysis of 13 studies. Br J Cancer 105, 709-722.

EHBCC-Group, E.H., Breast Cancer Collaborative, G., Key, T.J., Appleby, P.N., Reeves, G.K., Travis, R.C., Alberg, A.J., Barricarte, A., Berrino, F., Krogh, V., et al. (2013). Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. Lancet Oncol 14, 1009-1019.

Eichner, L.J., Perry, M.C., Dufour, C.R., Bertos, N., Park, M., St-Pierre, J., and Giguere, V. (2010). miR-378(*) mediates metabolic shift in breast cancer cells via the PGC-1beta/ERRgamma transcriptional pathway. Cell Metab 12, 352-361.

Elango, N., and Yi, S.V. (2011). Functional relevance of CpG island length for regulation of gene expression. Genetics 187, 1077-1083.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5, 435-445.

Public Health England (2017a). 100,000 Genome Project.

Public Health England (2016). Breast cancer in women

Public Health England (2017b). NHS breast screening (BSP) programme

Eudy, J.D., Yao, S., Weston, M.D., Ma-Edmonds, M., Talmadge, C.B., Cheng, J.J., Kimberling, W.J., and Sumegi, J. (1998). Isolation of a gene encoding a

novel member of the nuclear receptor superfamily from the critical region of Usher syndrome type IIa at 1q41. Genomics 50, 382-384.

Evans, D.G., Brentnall, A., Byers, H., Harkness, E., Stavrinos, P., Howell, A., Group, F.H.-r.s., Newman, W.G., and Cuzick, J. (2017). The impact of a panel of 18 SNPs on breast cancer risk in women attending a UK familial screening clinic: a case-control study. J Med Genet 54, 111-113.

Evans, D.G., Susnerwala, I., Dawson, J., Woodward, E., Maher, E.R., and Lalloo, F. (2010). Risk of breast cancer in male BRCA2 carriers. J Med Genet 47, 710-711.

Fan, Y., and Song, Y.Q. (2017). PyHLA: tests for the association between HLA alleles and diseases. BMC Bioinformatics 18, 90.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 9, 102-114.

Fisher, S., Gao, H., Yasui, Y., Dabbs, K., and Winget, M. (2015). Survival in stage I-III breast cancer patients by surgical treatment in a publicly funded health care system. Ann Oncol 26, 1161-1169.

Fondon, J.W., 3rd, and Garner, H.R. (2004). Molecular origins of rapid and continuous morphological evolution. Proc Natl Acad Sci U S A 101, 18058-18063.

Fonville, N.C., Vaksman, Z., McIver, L.J., and Garner, H.R. (2015). Population analysis of microsatellite genotypes reveals a signature associated with ovarian cancer. Oncotarget 6, 11407-11420.

Forgacs, E., Wren, J.D., Kamibayashi, C., Kondo, M., Xu, X.L., Markowitz, S., Tomlinson, G.E., Muller, C.Y., Gazdar, A.F., Garner, H.R., et al. (2001). Searching for microsatellite mutations in coding regions in lung, breast, ovarian and colorectal cancers. Oncogene 20, 1005-1009.

Foulkes, W.D., Metcalfe, K., Sun, P., Hanna, W.M., Lynch, H.T., Ghadirian, P., Tung, N., Olopade, O.I., Weber, B.L., McLennan, J., et al. (2004). Estrogen receptor status in BRCA1- and BRCA2-related breast cancer: the influence of age, grade, and histological type. Clin Cancer Res 10, 2029-2034.

Galindo, C.L., McCormick, J.F., Bubb, V.J., Abid Alkadem, D.H., Li, L.S., McIver, L.J., George, A.C., Boothman, D.A., Quinn, J.P., Skinner, M.A., et al. (2011). A long AAAG repeat allele in the 5' UTR of the ERR-gamma gene is correlated with breast cancer predisposition and drives promoter activity in MCF-7 breast cancer cells. Breast cancer research and treatment 130, 41-48.

Gettings, K.B., Aponte, R.A., Vallone, P.M., and Butler, J.M. (2015). STR allele sequence variation: Current knowledge and future issues. Forensic Sci Int Genet 18, 118-130.

Giguere, V. (1999). Orphan nuclear receptors: from gene to function. Endocr Rev 20, 689-725.

Giguere, V. (2002). To ERR in the estrogen pathway. Trends Endocrinol Metab 13, 220-225.

Giguere, V., Yang, N., Segui, P., and Evans, R.M. (1988). Identification of a new class of steroid hormone receptors. Nature 331, 91-94.

Gilliland, F.D., Hunt, W.C., Baumgartner, K.B., Crumley, D., Nicholson, C.S., Fetherolf, J., and Samet, J.M. (1998). Reproductive risk factors for breast cancer in Hispanic and non-Hispanic white women: the New Mexico Women's Health Study. Am J Epidemiol 148, 683-692.

Girardi, F., Barnes, D.R., Barrowdale, D., Frost, D., Brady, A.F., Miller, C., Henderson, A., Donaldson, A., Murray, A., Brewer, C., et al. (2018). Risks of breast or ovarian cancer in BRCA1 or BRCA2 predictive test negatives: findings from the EMBRACE study. Genet Med.

Godlewski, J., Nowicki, M.O., Bronisz, A., Williams, S., Otsuki, A., Nuovo, G., Raychaudhury, A., Newton, H.B., Chiocca, E.A., and Lawler, S. (2008). Targeting of the Bmi-1 oncogene/stem cell renewal factor by microRNA-128 inhibits glioma proliferation and self-renewal. Cancer Res 68, 9125-9130.

Goldenberg, G.J., and Froese, E.K. (1982). Drug and hormone sensitivity of estrogen receptor-positive and -negative human breast cancer cells in vitro. Cancer Res 42, 5147-5151.

Graffeo, R., Livraghi, L., Pagani, O., Goldhirsch, A., Partridge, A.H., and Garber, J.E. (2016). Time to incorporate germline multigene panel testing

into breast and ovarian cancer patient care. Breast cancer research and treatment 160, 393-410.

Guo, J., Xia, B., Meng, F., and Lou, G. (2013). miR-137 suppresses cell growth in ovarian cancer by targeting AEG-1. Biochem Biophys Res Commun 441, 357-363.

Hall, I.J., Moorman, P.G., Millikan, R.C., and Newman, B. (2005). Comparative analysis of breast cancer risk factors among African-American women and White women. Am J Epidemiol 161, 40-51.

Hao, S., Luo, C., Abukiwan, A., Wang, G., He, J., Huang, L., Weber, C.E., Lv, N., Xiao, X., Eichmuller, S.B., et al. (2015). miR-137 inhibits proliferation of melanoma cells by targeting PAK2. Exp Dermatol 24, 947-952.

Hartman, T.J., Sisti, J.S., Hankinson, S.E., Xu, X., Eliassen, A.H., and Ziegler, R. (2016). Alcohol Consumption and Urinary Estrogens and Estrogen Metabolites in Premenopausal Women. Horm Cancer 7, 65-74.

Heard, D.J., Norby, P.L., Holloway, J., and Vissing, H. (2000). Human ERRgamma, a third member of the estrogen receptor-related receptor (ERR) subfamily of orphan nuclear receptors: tissue-specific isoforms are expressed during development and in the adult. Mol Endocrinol 14, 382-392.

Helsen, C., Kerkhofs, S., Clinckemalie, L., Spans, L., Laurent, M., Boonen, S., Vanderschueren, D., and Claessens, F. (2012). Structural basis for nuclear hormone receptor DNA binding. Mol Cell Endocrinol 348, 411-417.

Hirko, K.A., Spiegelman, D., Willett, W.C., Hankinson, S.E., and Eliassen, A.H. (2014). Alcohol consumption in relation to plasma sex hormones, prolactin, and sex hormone-binding globulin in premenopausal women. Cancer Epidemiol Biomarkers Prev 23, 2943-2953.

Holzer, G., Markov, G.V., and Laudet, V. (2017). Evolution of Nuclear Receptors and Ligand Signaling: Toward a Soft Key-Lock Model? Curr Top Dev Biol 125, 1-38.

Houlston, R.S., and Tomlinson, I.P. (2001). Polymorphisms and colorectal tumor risk. Gastroenterology 121, 282-301.

Howell, S.J., Johnston, S.R., and Howell, A. (2004). The use of selective estrogen receptor modulators and selective estrogen receptor down-

regulators in breast cancer. Best Pract Res Clin Endocrinol Metab 18, 47-66.

Howlett, N.G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., De Die-Smulders, C., Persky, N., Grompe, M., Joenje, H., Pals, G., et al. (2002). Biallelic inactivation of BRCA2 in Fanconi anemia. Science 297, 606-609.

Hsieh, A.R., Chang, S.W., Chen, P.L., Chu, C.C., Hsiao, C.L., Yang, W.S., Chang, C.C., Wu, J.Y., Chen, Y.T., Chang, T.C., et al. (2014). Predicting HLA genotypes using unphased and flanking single-nucleotide polymorphisms in Han Chinese population. BMC Genomics 15, 81.

Huang, P., Chandra, V., and Rastinejad, F. (2010). Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. Annu Rev Physiol 72, 247-272.

Huang, Y.C., Lee, C.T., Lee, J.C., Liu, Y.W., Chen, Y.J., Tseng, J.T., Kang, J.W., Sheu, B.S., Lin, B.W., and Hung, L.Y. (2016). Epigenetic silencing of miR-137 contributes to early colorectal carcinogenesis by impaired Aurora-A inhibition. Oncotarget 7, 76852-76866.

Huff, V., Jaffe, N., Saunders, G.F., Strong, L.C., Villalba, F., and Ruteshouser, E.C. (1995). WT1 exon 1 deletion/insertion mutations in Wilms tumor patients, associated with di- and trinucleotide repeats and deletion hotspot consensus sequences. Am J Hum Genet 56, 84-90.

Hurtado, A., Holmes, K.A., Geistlinger, T.R., Hutcheson, I.R., Nicholson, R.I., Brown, M., Jiang, J., Howat, W.J., Ali, S., and Carroll, J.S. (2008). Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. Nature 456, 663-666.

Huss, J.M., Garbacz, W.G., and Xie, W. (2015). Constitutive activities of estrogen-related receptors: Transcriptional regulation of metabolism by the ERR pathways in health and disease. Biochim Biophys Acta 1852, 1912-1927.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science 293, 834-838.

Ijichi, N., Shigekawa, T., Ikeda, K., Horie-Inoue, K., Fujimura, T., Tsuda, H., Osaki, A., Saeki, T., and Inoue, S. (2011). Estrogen-related receptor gamma modulates cell proliferation and estrogen signaling in breast cancer. J Steroid Biochem Mol Biol 123, 1-7.

Ingham, S.L., Warwick, J., Byers, H., Lalloo, F., Newman, W.G., and Evans, D.G. (2013). Is multiple SNP testing in BRCA2 and BRCA1 female carriers ready for use in clinical practice? Results from a large Genetic Centre in the UK. Clin Genet 84, 37-42.

Inui, M., Martello, G., and Piccolo, S. (2010). MicroRNA control of signal transduction. Nat Rev Mol Cell Biol 11, 252-263.

Isaksson, E., von Schoultz, E., Odlind, V., Soderqvist, G., Csemiczky, G., Carlstrom, K., Skoog, L., and von Schoultz, B. (2001). Effects of oral contraceptives on breast epithelial proliferation. Breast cancer research and treatment 65, 163-169.

Ivey, K.N., and Srivastava, D. (2010). MicroRNAs as regulators of differentiation and cell fate decisions. Cell Stem Cell 7, 36-41.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., and Thun, M.J. (2009). Cancer statistics, 2009. CA Cancer J Clin 59, 225-249.

Johns, D., Agarwal, J., Anderson, L., Ying, J., and Kohlmann, W. (2017). Breast Cancer Risk Reduction Decisions of the BRCA-Positive Patient: An Observational Study at a Single Institution. J Womens Health (Larchmt) 26, 702-706.

Jones, M.E., Schoemaker, M.J., Wright, L., McFadden, E., Griffin, J., Thomas, D., Hemming, J., Wright, K., Ashworth, A., and Swerdlow, A.J. (2016). Menopausal hormone therapy and breast cancer: what is the true size of the increased risk? Br J Cancer 115, 607-615.

Joukov, V., Groen, A.C., Prokhorova, T., Gerson, R., White, E., Rodriguez, A., Walter, J.C., and Livingston, D.M. (2006). The BRCA1/BARD1 heterodimer modulates ran-dependent mitotic spindle assembly. Cell 127, 539-552.

Kang, N., Choi, S.Y., Kim, Y.K., Yoo Ie, R., Han, D.H., Lee, D.S., Kim, Y.S., Hong, S.H., Kang, J.H., Lee, K.Y., et al. (2015). Silencing of miR-137 by

aberrant promoter hypermethylation in surgically resected lung cancer. Lung Cancer 89, 99-103.

Karimi, P., Hematti, S., Safari, F., and Tavassoli, M. (2013). Polymorphic AAAG repeat length in estrogen-related receptor gamma (ERRgamma) and risk of breast cancer in Iranian women. Cancer investigation 31, 600-603.

Karunasena, E., McIver, L.J., Rood, B.R., Wu, X., Zhu, H., Bavarva, J.H., and Garner, H.R. (2014). Somatic intronic microsatellite loci differentiate glioblastoma from lower-grade gliomas. Oncotarget 5, 6003-6014.

Key, T.J., Appleby, P.N., Reeves, G.K., Roddam, A., Dorgan, J.F., Longcope, C., Stanczyk, F.Z., Stephenson, H.E., Jr., Falk, R.T., Miller, R., et al. (2003). Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women. J Natl Cancer Inst 95, 1218-1226.

King, M.C., Marks, J.H., Mandell, J.B., and New York Breast Cancer Study, G. (2003). Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science 302, 643-646.

Kit, S. (1961). Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. J Mol Biol 3, 711-716.

Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I., and Khorana, H.G. (1971). Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. J Mol Biol 56, 341-361.

Knudson, A.G., Jr. (1971). Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 68, 820-823.

Knutti, D., and Kralli, A. (2001). PGC-1, a versatile coactivator. Trends Endocrinol Metab 12, 360-365.

Kozaki, K., Imoto, I., Mogi, S., Omura, K., and Inazawa, J. (2008). Exploration of tumor-suppressive microRNAs silenced by DNA hypermethylation in oral cancer. Cancer Res 68, 2094-2105.

Kristensen, L.S., and Hansen, L.L. (2009). PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. Clin Chem 55, 1471-1483.

Krush, A.J. (1979). Contributions of Pierre Paul Broca to Cancer Genetics. Transactions of the Nebraska Academz of Sciences VII, 125-129.

Kuchenbaecker, K.B., Hopper, J.L., Barnes, D.R., Phillips, K.A., Mooij, T.M., Roos-Blom, M.J., Jervis, S., van Leeuwen, F.E., Milne, R.L., Andrieu, N., et al. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA 317, 2402-2416.

Lambertini, M., Santoro, L., Del Mastro, L., Nguyen, B., Livraghi, L., Ugolini, D., Peccatori, F.A., and Azim, H.A., Jr. (2016). Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. Cancer Treat Rev 49, 65-76.

Langevin, S.M., Stone, R.A., Bunker, C.H., Grandis, J.R., Sobol, R.W., and Taioli, E. (2010). MicroRNA-137 promoter methylation in oral rinses from patients with squamous cell carcinoma of the head and neck is associated with gender and body mass index. Carcinogenesis 31, 864-870.

Lavu, V., Venkatesan, V., Venkata Kameswara Subrahmanya Lakkakula, B., Venugopal, P., Paul, S.F., and Rao, S.R. (2015). Polymorphic regions in the interleukin-1 gene and susceptibility to chronic periodontitis: a genetic association study. Genet Test Mol Biomarkers 19, 175-181.

Lee, A.V., Oesterreich, S., and Davidson, N.E. (2015a). MCF-7 cells--changing the course of breast cancer research and care for 45 years. J Natl Cancer Inst 107.

Lee, J.M., Cho, K.W., Kim, E.J., Tang, Q., Kim, K.S., Tickle, C., and Jung, H.S. (2015b). A contrasting function for miR-137 in embryonic mammogenesis and adult breast carcinogenesis. Oncotarget 6, 22048-22059.

Lenkala, D., LaCroix, B., Gamazon, E.R., Geeleher, P., Im, H.K., and Huang, R.S. (2014). The impact of microRNA expression on cellular proliferation. Hum Genet 133, 931-938.

Li, C.I., Malone, K.E., Daling, J.R., Potter, J.D., Bernstein, L., Marchbanks, P.A., Strom, B.L., Simon, M.S., Press, M.F., Ursin, G., et al. (2008). Timing of menarche and first full-term birth in relation to breast cancer risk. Am J Epidemiol 167, 230-239.

Li, Z.M., Zhang, H.Y., Wang, Y.X., and Wang, W.B. (2016). MicroRNA-137 is downregulated in human osteosarcoma and regulates cell proliferation and migration through targeting FXYD6. J Drug Target 24, 102-110.

Lin, J., Tarr, P.T., Yang, R., Rhee, J., Puigserver, P., Newgard, C.B., and Spiegelman, B.M. (2003). PGC-1beta in the regulation of hepatic glucose and energy metabolism. J Biol Chem 278, 30843-30848.

Lindstrom, S., Vachon, C.M., Li, J., Varghese, J., Thompson, D., Warren, R., Brown, J., Leyland, J., Audley, T., Wareham, N.J., et al. (2011). Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. Nat Genet 43, 185-187.

Liu, S., Ruan, X., Schultz, S., Neubauer, H., Fehm, T., Seeger, H., and Mueck, A.O. (2015). Oestetrol stimulates proliferation and oestrogen receptor expression in breast cancer cell lines: comparison of four oestrogens. Eur J Contracept Reprod Health Care 20, 29-35.

Lu, X., Peng, L., Lv, M., and ding, K. (2012). Recent advance in the design of small molecular modulators of estrogen-related receptors. Curr Pharm Des 18, 3421-3431.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science 303, 95-98.

Magni, M., Ruscica, V., Buscemi, G., Kim, J.E., Nachimuthu, B.T., Fontanella, E., Delia, D., and Zannini, L. (2014). Chk2 and REGgamma-dependent DBC1 regulation in DNA damage induced apoptosis. Nucleic Acids Res 42, 13150-13160.

Mahmoudi, E., and Cairns, M.J. (2017). MiR-137: an important player in neural development and neoplastic transformation. Mol Psychiatry 22, 44-55.

Mamdani, M., McMichael, G.O., Gadepalli, V., Williamson, V., Parker, E.K., Haroutunian, V., and Vladimirov, V.I. (2013). Differential regulation of schizophrenia-associated microRNA gene function by variable number tandem repeats (VNTR) polymorphism. Schizophr Res 151, 284-286.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell 110, 563-574.

Mavaddat, N., Antoniou, A.C., Easton, D.F., and Garcia-Closas, M. (2010). Genetic susceptibility to breast cancer. Mol Oncol 4, 174-191.

McCormack, V.A., and dos Santos Silva, I. (2006). Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 15, 1159-1169.

McIver, L.J., Fonville, N.C., Karunasena, E., and Garner, H.R. (2014). Microsatellite genotyping reveals a signature in breast cancer exomes. Breast cancer research and treatment 145, 791-798.

Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., Hollestelle, A., Houben, M., Crepin, E., van Veghel-Plandsoen, M., et al. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet 31, 55-59.

Merlo, A., Herman, J.G., Mao, L., Lee, D.J., Gabrielson, E., Burger, P.C., Baylin, S.B., and Sidransky, D. (1995). 5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nature medicine 1, 686-692.

Mikeska, T., Bock, C., Do, H., and Dobrovic, A. (2012). DNA methylation biomarkers in cancer: progress towards clinical implementation. Expert Rev Mol Diagn 12, 473-487.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., et al. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266, 66-71.

Milne, R.L., and Antoniou, A.C. (2016). Modifiers of breast and ovarian cancer risks for BRCA1 and BRCA2 mutation carriers. Endocr Relat Cancer 23, T69-84.

Min, L., Wang, F., Hu, S., Chen, Y., Yang, J., Liang, S., and Xu, X. (2018). Aberrant microRNA-137 promoter methylation is associated with lymph node metastasis and poor clinical outcomes in non-small cell lung cancer. Oncol Lett 15, 7744-7750.

Minagawa, M., Yasuda, T., Watanabe, T., Minamitani, K., Takahashi, Y., Goltzman, D., White, J.H., Hendy, G.N., and Kohno, Y. (2002). Association between AAAG repeat polymorphism in the P3 promoter of the human parathyroid hormone (PTH)/PTH-related peptide receptor gene and adult height, urinary pyridinoline excretion, and promoter activity. J Clin Endocrinol Metab 87, 1791-1796.

Misawa, A., and Inoue, S. (2015). Estrogen-Related Receptors in Breast Cancer and Prostate Cancer. Front Endocrinol (Lausanne) 6, 83.

Mullican, S.E., Dispirito, J.R., and Lazar, M.A. (2013). The orphan nuclear receptors at their 25-year reunion. J Mol Endocrinol 51, T115-140.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor symposia on quantitative biology 51 Pt 1, 263-273.

Mullis, K.B., and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in enzymology 155, 335-350.

Navarro, F., and Lieberman, J. (2010). Small RNAs guide hematopoietic cell differentiation and function. J Immunol 184, 5939-5947.

Neta, G., Anderson, W.F., Gilbert, E., and Berrington, A. (2012). Variation in the risk of radiation-related contralateral breast cancer by histology and estrogen receptor expression in SEER. Breast cancer research and treatment 131, 1021-1027.

NHS (2018). Statistics on Obesity, Physical Activity and Diet - England, 2018. In NHS England (National Institute of Statistics).

Nielsen, F.C., van Overeem Hansen, T., and Sorensen, C.S. (2016). Hereditary breast and ovarian cancer: new genes in confined pathways. Nat Rev Cancer 16, 599-612.

Niguidula, N., Alamillo, C., Shahmirzadi Mowlavi, L., Powis, Z., Cohen, J.S., and Farwell Hagman, K.D. (2018). Clinical whole-exome sequencing results impact medical management. Mol Genet Genomic Med.

O'Donovan, P.J., and Livingston, D.M. (2010). BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. Carcinogenesis 31, 961-967.

Odefrey, F., Stone, J., Gurrin, L.C., Byrnes, G.B., Apicella, C., Dite, G.S., Cawson, J.N., Giles, G.G., Treloar, S.A., English, D.R., et al. (2010). Common genetic variants associated with breast cancer and mammographic density measures that predict disease. Cancer Res 70, 1449-1458.

Osorio, A., de la Hoya, M., Rodriguez-Lopez, R., Martinez-Ramirez, A., Cazorla, A., Granizo, J.J., Esteller, M., Rivas, C., Caldes, T., and Benitez, J. (2002). Loss of heterozygosity analysis at the BRCA loci in tumor samples from patients with familial breast cancer. Int J Cancer 99, 305-309.

Parkin, D.M. (2011). 9. Cancers attributable to inadequate physical exercise in the UK in 2010. Br J Cancer 105 Suppl 2, S38-41.

Paska, A.V., and Hudler, P. (2015). Aberrant methylation patterns in cancer: a clinical view. Biochem Med (Zagreb) 25, 161-176.

Phelan, C.M., Rebbeck, T.R., Weber, B.L., Devilee, P., Ruttledge, M.H., Lynch, H.T., Lenoir, G.M., Stratton, M.R., Easton, D.F., Ponder, B.A., et al. (1996). Ovarian cancer risk in BRCA1 carriers is modified by the HRAS1 variable number of tandem repeat (VNTR) locus. Nat Genet 12, 309-311.

Pike, M.C., Krailo, M.D., Henderson, B.E., Casagrande, J.T., and Hoel, D.G. (1983). 'Hormonal' risk factors, 'breast tissue age' and the age-incidence of breast cancer. Nature 303, 767-770.

Platet, N., Cathiard, A.M., Gleizes, M., and Garcia, M. (2004). Estrogens and their receptors in breast cancer progression: a dual role in cancer proliferation and invasion. Crit Rev Oncol Hematol 51, 55-67.

Ponten, J., Holmberg, L., Trichopoulos, D., Kallioniemi, O.P., Kvale, G., Wallgren, A., and Taylor-Papadimitriou, J. (1990). Biology and natural history of breast cancer. Int J Cancer Suppl 5, 5-21.

Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nat Biotechnol 28, 1057-1068.

Prapa, M., Solomons, J., and Tischkowitz, M. (2017). The use of panel testing in familial breast and ovarian cancer. Clin Med (Lond) 17, 568-572.

Puigserver, P., and Spiegelman, B.M. (2003). Peroxisome proliferator-activated receptor-gamma coactivator 1 alpha (PGC-1 alpha): transcriptional coactivator and metabolic regulator. Endocr Rev 24, 78-90.

Rebbeck, T.R., and Domchek, S.M. (2008). Variation in breast cancer risk in BRCA1 and BRCA2 mutation carriers. Breast cancer research : BCR 10, 108.

Reeves, G.K., Pirie, K., Beral, V., Green, J., Spencer, E., Bull, D., and Million Women Study, C. (2007). Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. BMJ 335, 1134.

Reichman, M.E., Judd, J.T., Longcope, C., Schatzkin, A., Clevidence, B.A., Nair, P.P., Campbell, W.S., and Taylor, P.R. (1993). Effects of alcohol consumption on plasma and urinary hormone concentrations in premenopausal women. J Natl Cancer Inst 85, 722-727.

Rettori, V., and McCann, S.M. (1997). The mechanism of action of alcohol to suppress gonadotropin secretion. Mol Psychiatry 2, 350-354.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16, 276-277.

Riggins, R.B., Bouton, A.H., Liu, M.C., and Clarke, R. (2005). Antiestrogens, aromatase inhibitors, and apoptosis in breast cancer. Vitam Horm 71, 201-237.

Riggins, R.B., Mazzotta, M.M., Maniya, O.Z., and Clarke, R. (2010). Orphan nuclear receptors in breast cancer pathogenesis and therapeutic response. Endocr Relat Cancer 17, R213-231.

Rinaldi, S., Peeters, P.H., Bezemer, I.D., Dossus, L., Biessy, C., Sacerdote, C., Berrino, F., Panico, S., Palli, D., Tumino, R., et al. (2006). Relationship of alcohol intake and sex steroid concentrations in blood in pre- and post-menopausal women: the European Prospective Investigation into Cancer and Nutrition. Cancer Causes Control 17, 1033-1043.

Roberts, M.E., Jackson, S.A., Susswein, L.R., Zeinomar, N., Ma, X., Marshall, M.L., Stettner, A.R., Milewski, B., Xu, Z., Solomon, B.D., et al. (2018). MSH6

and PMS2 germ-line pathogenic variants implicated in Lynch syndrome are associated with breast cancer. Genet Med.

Robertson, K.D., and Wolffe, A.P. (2000). DNA methylation in health and disease. Nat Rev Genet 1, 11-19.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J.L., and Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. Genome Res 14, 1902-1910.

Ronneberg, J.A., Tost, J., Solvang, H.K., Alnaes, G.I., Johansen, F.E., Brendeford, E.M., Yakhini, Z., Gut, I.G., Lonning, P.E., Borresen-Dale, A.L., et al. (2008). GSTP1 promoter haplotypes affect DNA methylation levels and promoter activity in breast carcinomas. Cancer Res 68, 5562-5571.

Roshan-Moniri, M., Hsing, M., Butler, M.S., Cherkasov, A., and Rennie, P.S. (2014). Orphan nuclear receptors as drug targets for the treatment of prostate and breast cancers. Cancer Treat Rev 40, 1137-1152.

Rosner, B., Colditz, G.A., and Willett, W.C. (1994). Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. Am J Epidemiol 139, 819-835.

Santen, R.J., Brodie, H., Simpson, E.R., Siiteri, P.K., and Brodie, A. (2009). History of aromatase: saga of an important biological mediator and therapeutic target. Endocr Rev 30, 343-375.

Sayed, D., and Abdellatif, M. (2011). MicroRNAs in development and disease. Physiol Rev 91, 827-887.

Scillitani, A., Jang, C., Wong, B.Y., Hendy, G.N., and Cole, D.E. (2006). A functional polymorphism in the PTHR1 promoter region is associated with adult height and BMD measured at the femoral neck in a large cohort of young caucasian women. Hum Genet 119, 416-421.

Seitz HK, P.C., Bagnardi V, La Vecchia C. (2012). Epidemiology and pathophysiology of alcohol and breast cancer: Update 2012. Alcohol and Alcoholism 4, 204-212.

Sengupta, D., Bhargava, D.K., Dixit, A., Sahoo, B.S., Biswas, S., Biswas, G., and Mishra, S.K. (2014). ERRbeta signalling through FST and BCAS2

inhibits cellular proliferation in breast cancer cells. Br J Cancer 110, 2144-2158.

Sham, P.C., and Curtis, D. (1995). Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Ann Hum Genet 59, 97-105.

Shen, H., Wang, L., Ge, X., Jiang, C.F., Shi, Z.M., Li, D.M., Liu, W.T., Yu, X., and Shu, Y.Q. (2016). MicroRNA-137 inhibits tumor growth and sensitizes chemosensitivity to paclitaxel and cisplatin in lung cancer. Oncotarget 7, 20728-20742.

Shi, H., Shigeta, H., Yang, N., Fu, K., O'Brian, G., and Teng, C.T. (1997). Human estrogen receptor-like 1 (ESRL1) gene: genomic organization, chromosomal localization, and promoter characterization. Genomics 44, 52-60.

Sierksma, A., Sarkola, T., Eriksson, C.J., van der Gaag, M.S., Grobbee, D.E., and Hendriks, H.F. (2004). Effect of moderate alcohol consumption on plasma dehydroepiandrosterone sulfate, testosterone, and estradiol levels in middle-aged men and postmenopausal women: a diet-controlled intervention study. Alcohol Clin Exp Res 28, 780-785.

Sikora, M.J., Johnson, M.D., Lee, A.V., and Oesterreich, S. (2016). Endocrine Response Phenotypes Are Altered by Charcoal-Stripped Serum Variability. Endocrinology 157, 3760-3766.

Singletary, K.W., Frey, R.S., and Yan, W. (2001). Effect of ethanol on proliferation and estrogen receptor-alpha expression in human breast cancer cells. Cancer Lett 165, 131-137.

Sladek, R., Beatty, B., Squire, J., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., and Giguere, V. (1997). Chromosomal mapping of the human and murine orphan receptors ERRalpha (ESRRA) and ERRbeta (ESRRB) and identification of a novel human ERRalpha-related pseudogene. Genomics 45, 320-326.

Smith, A.R., Marquez, R.T., Tsao, W.C., Pathak, S., Roy, A., Ping, J., Wilkerson, B., Lan, L., Meng, W., Neufeld, K.L., et al. (2015). Tumor suppressive microRNA-137 negatively regulates Musashi-1 and colorectal cancer progression. Oncotarget 6, 12558-12573.

Smollich, M., Gotte, M., Fischgrabe, J., Radke, I., Kiesel, L., and Wulfing, P. (2009). Differential effects of aromatase inhibitors and antiestrogens on estrogen receptor expression in breast cancer cells. Anticancer Res 29, 2167-2171.

Soria, C., Estermann, F.E., Espantman, K.C., and O'Shea, C.C. (2010). Heterochromatin silencing of p53 target genes by a small viral protein. Nature 466, 1076-1081.

Sorm, F., and Vesely, J. (1968). Effect of 5-aza-2'-deoxycytidine against leukemic and hemopoietic tissues in AKR mice. Neoplasma 15, 339-343.

Stein, R.A., Chang, C.Y., Kazmin, D.A., Way, J., Schroeder, T., Wergin, M., Dewhirst, M.W., and McDonnell, D.P. (2008). Estrogen-related receptor alpha is critical for the growth of estrogen receptor-negative breast cancer. Cancer Res 68, 8805-8812.

Stevens, R.G., Davis, S., Mirick, D.K., Kheifets, L., and Kaune, W. (2000). Alcohol consumption and urinary concentration of 6-sulfatoxymelatonin in healthy women. Epidemiology 11, 660-665.

Stingl, J. (2011). Estrogen and progesterone in normal mammary gland development and in cancer. Horm Cancer 2, 85-90.

Su, Z., Yang, Z., Xu, Y., Chen, Y., and Yu, Q. (2015). MicroRNAs in apoptosis, autophagy and necroptosis. Oncotarget 6, 8474-8490.

Sun, P., Wei, L., Denkert, C., Lichtenegger, W., and Sehouli, J. (2006). The orphan nuclear receptors, estrogen receptor-related receptors: their role as new biomarkers in gynecological cancer. Anticancer Res 26, 1699-1706.

Surakasula, A., Nagarjunapu, G.C., and Raghavaiah, K.V. (2014). A comparative study of pre- and post-menopausal breast cancer: Risk factors, presentation, characteristics and management. J Res Pharm Pract 3, 12-18.

Suzuki, T., Miki, Y., Moriya, T., Shimada, N., Ishida, T., Hirakawa, H., Ohuchi, N., and Sasano, H. (2004). Estrogen-related receptor alpha in human breast carcinoma as a potent prognostic factor. Cancer Res 64, 4670-4676.

Tai, Y.C., Domchek, S., Parmigiani, G., and Chen, S. (2007). Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. J Natl Cancer Inst 99, 1811-1814.

Tamimi, R.M., Hankinson, S.E., Ding, S., Gagalang, V., Larson, G.P., Spiegelman, D., Colditz, G.A., Krontiris, T.G., and Hunter, D.J. (2003). The HRAS1 variable number of tandem repeats and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 12, 1528-1530.

Tiraby, C., Hazen, B.C., Gantner, M.L., and Kralli, A. (2011). Estrogen-related receptor gamma promotes mesenchymal-to-epithelial transition and suppresses breast tumor growth. Cancer Res 71, 2518-2528.

Tremblay, G.B., Kunath, T., Bergeron, D., Lapointe, L., Champigny, C., Bader, J.A., Rossant, J., and Giguere, V. (2001). Diethylstilbestrol regulates trophoblast stem cell differentiation as a ligand of orphan nuclear receptor ERR beta. Genes Dev 15, 833-838.

Turnbull, C., Ahmed, S., Morrison, J., Pernet, D., Renwick, A., Maranian, M., Seal, S., Ghoussaini, M., Hines, S., Healey, C.S., et al. (2010). Genome-wide association study identifies five new breast cancer susceptibility loci. Nat Genet 42, 504-507.

Ursin, G., Lillie, E.O., Lee, E., Cockburn, M., Schork, N.J., Cozen, W., Parisky, Y.R., Hamilton, A.S., Astrahan, M.A., and Mack, T. (2009). The relative importance of genetics and environment on mammographic density. Cancer Epidemiol Biomarkers Prev 18, 102-112.

Vachon, C.M., Pankratz, V.S., Scott, C.G., Haeberle, L., Ziv, E., Jensen, M.R., Brandt, K.R., Whaley, D.H., Olson, J.E., Heusinger, K., et al. (2015). The contributions of breast density and common genetic variation to breast cancer risk. J Natl Cancer Inst 107.

Veigl, M.L., Kasturi, L., Olechnowicz, J., Ma, A.H., Lutterbaugh, J.D., Periyasamy, S., Li, G.M., Drummond, J., Modrich, P.L., Sedwick, W.D., et al. (1998). Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. Proc Natl Acad Sci U S A 95, 8698-8702.

Verma, M., and Banerjee, H.N. (2015). Epigenetic inhibitors. Methods in molecular biology 1238, 469-485.

Vrba, L., Munoz-Rodriguez, J.L., Stampfer, M.R., and Futscher, B.W. (2013). miRNA gene promoters are frequent targets of aberrant DNA methylation in human breast cancer. PLoS One 8, e54398.

Wagoner, M.P., Gunsalus, K.T., Schoenike, B., Richardson, A.L., Friedl, A., and Roopra, A. (2010). The transcription factor REST is lost in aggressive breast cancer. PLoS Genet 6, e1000979.

Walsh, T., Mandell, J.B., Norquist, B.M., Casadei, S., Gulsuner, S., Lee, M.K., and King, M.C. (2017). Genetic Predisposition to Breast Cancer Due to Mutations Other Than BRCA1 and BRCA2 Founder Alleles Among Ashkenazi Jewish Women. JAMA Oncol 3, 1647-1653.

Wang, L., Aasly, J.O., Annesi, G., Bardien, S., Bozi, M., Brice, A., Carr, J., Chung, S.J., Clarke, C., Crosiers, D., et al. (2015). Large-scale assessment of polyglutamine repeat expansions in Parkinson disease. Neurology 85, 1283-1292.

Warburton, A., Breen, G., Bubb, V.J., and Quinn, J.P. (2016a). A GWAS SNP for Schizophrenia Is Linked to the Internal MIR137 Promoter and Supports Differential Allele-Specific Expression. Schizophr Bull 42, 1003-1008.

Warburton, A., Breen, G., Rujescu, D., Bubb, V.J., and Quinn, J.P. (2015). Characterization of a REST-Regulated Internal Promoter in the Schizophrenia Genome-Wide Associated Gene MIR137. Schizophr Bull 41, 698-707.

Warburton, A., Miyajima, F., Shazadi, K., Crossley, J., Johnson, M.R., Marson, A.G., Baker, G.A., Quinn, J.P., and Sills, G.J. (2016b). NRSF and BDNF polymorphisms as biomarkers of cognitive dysfunction in adults with newly diagnosed epilepsy. Epilepsy Behav 54, 117-127.

Warburton, A.a. (2015). The REST/NRSF pathway as a central mechanism in CNS dysfunction (University of Liverpool).

Weischer, M., Bojesen, S.E., Ellervik, C., Tybjaerg-Hansen, A., and Nordestgaard, B.G. (2008). CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. J Clin Oncol 26, 542-548.

Weitzel, J.N., Ding, S., Larson, G.P., Nelson, R.A., Goodman, A., Grendys, E.C., Ball, H.G., and Krontiris, T.G. (2000). The HRAS1 minisatellite locus and risk of ovarian cancer. Cancer Res 60, 259-261.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42, D1001-1006.

West, C.N. (2004). Bilateral prophylactic mastectomy: Complications and subsequent procedures Journal of Clinical Oncology 22, 6041.

Wong, E.M., Southey, M.C., Fox, S.B., Brown, M.A., Dowty, J.G., Jenkins, M.A., Giles, G.G., Hopper, J.L., and Dobrovic, A. (2011). Constitutional methylation of the BRCA1 promoter is specifically associated with BRCA1 mutation-associated pathology in early-onset breast cancer. Cancer Prev Res (Phila) 4, 23-33.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., and Micklem, G. (1995). Identification of the breast cancer susceptibility gene BRCA2. Nature 378, 789-792.

Wright, C., Turner, J.A., Calhoun, V.D., and Perrone-Bizzozero, N. (2013). Potential Impact of miR-137 and Its Targets in Schizophrenia. Front Genet 4, 58.

Wu, H., Xu, H., Miraglia, L.J., and Crooke, S.T. (2000). Human RNase III is a 160-kDa protein involved in preribosomal RNA processing. J Biol Chem 275, 36957-36965.

Wu, Y., Zhang, D., and Kang, S. (2013). Physical activity and risk of breast cancer: a meta-analysis of prospective studies. Breast cancer research and treatment 137, 869-882.

Wu, Y.M., Chen, Z.J., Liu, H., Wei, W.D., Lu, L.L., Yang, X.L., Liang, W.T., Liu, T., Liu, H.L., Du, J., et al. (2015). Inhibition of ERRalpha suppresses epithelial mesenchymal transition of triple negative breast cancer cells by directly targeting fibronectin. Oncotarget 6, 25588-25601.

Xiang, C., Gao, H., Meng, L., Qin, Z., Ma, R., Liu, Y., Jiang, Y., Dang, C., Jin, L., He, F., et al. (2012). Functional variable number of tandem repeats

variation in the promoter of proto-oncogene PTTG1IP is associated with risk of estrogen receptor-positive breast cancer. Cancer Sci 103, 1121-1128.

Xiao, J., Peng, F., Yu, C., Wang, M., Li, X., Li, Z., Jiang, J., and Sun, C. (2014). microRNA-137 modulates pancreatic cancer cells tumor growth, invasion and sensitivity to chemotherapy. Int J Clin Exp Pathol 7, 7442-7450.

Xie, C., Shao, N., He, L., Li, X., Li, C., Li, J., and Su, D. (2018). Exploration of Racial Differences in Reproductive Factors for Breast Cancer among Women aged 55-74. J Cancer 9, 469-478.

Yarchoan, M., Hopkins, A., and Jaffee, E.M. (2017). Tumor Mutational Burden and Response Rate to PD-1 Inhibition. N Engl J Med 377, 2500-2501.

Yoon, S.L., Kim, D.C., Cho, S.H., Lee, S.Y., Chu, I.S., Heo, J., and Leem, S.H. (2010). Susceptibility for breast cancer in young patients with short rare minisatellite alleles of BORIS. BMB Rep 43, 698-703.

Zhang, C., Lv, G.Q., Yu, X.M., Gu, Y.L., Li, J.P., Du, L.F., and Zhou, P. (2011). Current evidence on the relationship between HRAS1 polymorphism and breast cancer risk: a meta-analysis. Breast cancer research and treatment 128, 467-472.

Zhang, L., Shen, Y.P., Hu, W.Z., Ran, S., Lin, Y., Lei, S.F., Zhang, Y.H., Papasian, C.J., Yi, N., and Pei, Y.F. (2016). A new method for estimating effect size distribution and heritability from genome-wide association summary results. Hum Genet 135, 171-184.

Zhang, N., Wang, X., Huo, Q., Sun, M., Cai, C., Liu, Z., Hu, G., and Yang, Q. (2014). MicroRNA-30a suppresses breast tumor growth and metastasis by targeting metadherin. Oncogene 33, 3119-3128.

Zhang, W., Chen, J.H., Shan, T., Aguilera-Barrantes, I., Wang, L.S., Huang, T.H., Rader, J.S., Sheng, X., and Huang, Y.W. (2018). miR-137 is a tumor suppressor in endometrial cancer and is repressed by DNA hypermethylation. Lab Invest.

Zhao, Y., Li, Y., Lou, G., Zhao, L., Xu, Z., Zhang, Y., and He, F. (2012). MiR-137 targets estrogen-related receptor alpha and impairs the proliferative and migratory capacity of breast cancer cells. PLoS One 7, e39102.

Zhu, Q., Pao, G.M., Huynh, A.M., Suh, H., Tonnu, N., Nederlof, P.M., Gage, F.H., and Verma, I.M. (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature 477, 179-184.

Zhu, X., Li, Y., Shen, H., Li, H., Long, L., Hui, L., and Xu, W. (2013). miR-137 inhibits the proliferation of lung cancer cells by targeting Cdc42 and Cdk6. FEBS Lett 587, 73-81.