Association of the SNP rs2623047 in the HSPG modification enzyme SULF1 with an Australian Caucasian Breast Cancer Cohort

*Rachel K Okolicsanyi[1], *Marion Faure[1,2], Jose ME Jacinto[3], Diego Chacon-Cortes[1], Suzanne Chambers[3-6], Philippa H Youl [3,4,7], Larisa M Haupt[1], Lyn R Griffiths[1]

1. Genomics Research Centre, Institute for Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia

2. AgroSup Dijon, 26 Boulevard Docteur Petitjean, 21000 Dijon, France

3. Griffith Health Institute, Griffith University, Southport, QLD, Australia

4. Cancer Council Queensland

5. Health and Wellness Institute, Edith Cowan University, Australia

6. Centre for Clinical Research, University of Queensland, Australia

7. School of Public Health, Queensland University of Technology

*RO and MF co-first authors

Address Correspondence to:

Professor Lyn Griffiths

Genomics Research Centre

Institute for Health and Biomedical Innovation

Queensland University of Technology

Musk Avenue, Kelvin Grove

Queensland, 4059

Australia

Email: lyn.griffiths@qut.edu.au

Phone: +61 7 3138 6102

Fax: +61 7 3138 6039

Financial support:

This research was supported by Cancer Council Queensland, Cancer Australia Grant (100639) and the Griffith Health Institute, Griffith University. SKC is an Australian Research Council Future Fellow. PHY is an NHMRC Early Career Fellow.

Abstract

Breast cancer is the second most common cancer worldwide and the most common cancer reported in women. This malignant tumour is characterised by a number of specific features including uncontrolled cell proliferation. It ranks fifth in the world as a cause of cancer death overall in developed countries and is the second most frequent cause of cancer death in women. Early diagnosis increases 5-year survival rates up to 95%. Heparan sulfate proteoglycans (HSPGs) are complex proteins composed of a core protein to which a number of highly sulfated side chains attach, ubiquitous to the cell surface and within the extracellular matrix. HSPG side chains are synthesised by a highly co-ordinated process resulting in distinct sulfation patterns, which determine specific interactions with cell-signalling partners including growth factors, their receptors, ligands and morphogens. The enzymes responsible for chain initiation, elongation and sulfation are critical for creating HS chain variability conferring biological functionality. This study investigated a single nucleotide polymorphism in *SULF1*, the enzyme responsible for the 6-*O* desulfation of heparan sulfate side chains. We investigated this SNP in an Australian Caucasian case-control breast cancer population and found a significant association between *SULF1* and breast cancer at both the allelic and genotypic level (allele, p=0.016; genotype, p=0.032). Our results suggest the rs2623047 SNP in *SULF1* may impact breast cancer susceptibility. Specifically, the T allele of rs2623047 in *SULF1* is associated with a increased risk of developing breast cancer in our cohort. The identification of markers including *SULF1* may improve detection of this disease at its earliest stages improving patient treatment and prognosis.

Keywords*:* Breast Cancer, PCR-RFLP, SNP rs2623047, SULF-1, HSPG

Introduction

Breast cancer is the second most common cancer worldwide (1.4 million cases, 10.9%) and is the most common form of cancer in women in all major regions of the world (ABS 2011) with an estimated 1.67 million new cancer cases diagnosed (25% of all cancers) in 2012 (Ferlay J, Soerjomataram I et al. 2013). It ranks fifth as cause of death (522,000, 6.1%) in the world, and it is the most frequent cause of cancer death in women (324,000 deaths, 14.3% of total) (Ferlay, Shin et al. 2010) in less developed regions and the second most common cause of cancer death in more developed regions (198,000, 15.4%). Non-modifiable risk factors increase the risk of developing breast cancer including: gender, age (50% of women diagnosed are between the ages of 50-69) (2006); breast cellular changes, including increased volume; personal history (previous diagnosis), family history of breast cancer (first-degree relative) and genetic factors (2010).

Breast volume (i.e 80%) is mainly attributed to the stroma comprised of collagen, fibroblasts, endothelial cells, adipocytes and a molecular network of proteoglycans. Stromal cells are embedded within the extracellular matrix (ECM) and provide a scaffold for cancer cells as well as producing ECM constituents for use by these cells. The two current models for tumour heterogeneity, the cancer stem cell hypothesis and the clonal evolution model, allow for a contribution from the acquisition of genetic events, epigenetic and microenvironmental changes in the metastasis and progression of cancer (Haupt and Griffiths 2009). Increasing evidence suggests there is extensive interaction between the tumour cells and the surrounding stromal compartment with both cells contributing to factors necessary for tumour survival (Haupt and Griffiths 2009). Key constituents of this microenvironment, proteoglycans (PGs) are composed of a core protein to which a number of glycosaminoglycan (GAGs) side chains are attached (Bernfield, Gotte et al. 1999) and include the heparan sulfate proteoglycans (HSPGs), a family of PGs predominantly decorated with heparan sulfate (HS) chains. HSPGs are ubiquitous to the cell niche and interact with a large number of ligands including growth factors, their receptors and ECM structural components (Haupt and Griffiths 2009). Localised to both the cell surface and the extracellular matrix (ECM), HSPGs are composed of a core protein to which a side chain of varying length and sulfation pattern is attached (Blackhall, Merry et al. 2001, Fernandez-Vega, Garcia et al. 2013). The heparan sulfate (HS) chains are synthesised by the addition of repeating units of unbranched disaccharides composed of alternate residues of N-acetylglucosamine and glucuronic acid (Gallagher 2001, Sugahara and Kitagawa 2002). The highly sulfated regions of HS are

responsible for most of the biological activity due to their charged interactions with basic amino acid clusters in proteins with the pattern of sulfation the significant contributor to their diverse biological activity (Gallagher 2001).

HSPGs have important roles in key biological functions in tissues, in both normal and pathological conditions, dependent on chain structure. HSPGs have been demonstrated to play a role in cell adhesion and migration, organisation of the extra-cellular matrix, differentiation and morphogenesis, along with cancer metastasis and the regulation of proliferation (Gallagher 2001). In breast tissue, the intralobular stroma rich in PGs mediates hormonally induced changes in breast volume (Wiseman and Werb 2002). However, these functions can be altered and regulated in several pathophysiological processes, such as cancer (Blackhall, Merry et al. 2001), with genes involved in the biosynthesis of these elements up- or down- regulated.

Many studies have associated HSPG core proteins and their modification enzymes with cancer and cancer-like diseases, most likely due to alterations in HSPG function and regulation of cell behaviour (Blackhall, Merry et al. 2001, Gallagher 2001). Core proteins carrying HS chains have previously been implicated in breast cancer development with increased expression of the core protein syndecan-1 (*SDC1*) associated with more severe forms of the disease (Lendorf, Manon-Jensen et al. 2011). More recently, the gene expression profile of HSPG chain initiation and modification enzymes as well as HSPG core proteins was examined following heparin treatment *in vitro*. Changes in gene expression was observed for *O*-sulfation enzymes (2-*O* and 6-*O*) as well as core syndecan proteins (*SDC2* and *SDC4*) along with altered proliferation, viability and tumourigeneity of these cells (Okolicsanyi, van Wijnen et al. 2013). Specifically, decreased expression of *HS6ST1*, an enzyme responsible for the addition of 6-*O* sulfation was observed in the lowly invasive, poorly metastatic MCF-7 cells following heparin treatment, while the same treatment produced an increase in expression in the highly invasive, highly metastatic MDA-MB-231 cells (Okolicsanyi, van Wijnen et al. 2013).

The heparan sulfatases are a family of HSPG enzymes that modulate HSPG/growth factor interactions and subsequent downstream signalling through modification of the HS side chain and includes HS 6-*O*-endosulfatase 1 (*SULF1*) (Morimoto-Tomita, Uchimura et al. 2002, Ai, Do et al. 2003, Isidor, Pichon et al. 2010). *SULF1* removes the 6-*O*-sulfate group from heparan sulfate chains, modulating HSPG function by altering binding through catalysing HSPG 6-*O* desulfation (Morimoto-Tomita, Uchimura et al. 2002, Ai, Do et al. 2003). *SULF1* and cancer risk have been

correlated in several gene expression studies (Han, Huang et al. 2011) with down regulation of *SULF1* described in malignant breast cancer cells. Similarly, i*n vitro* overexpression of *SULF*-1 in hepatocarcinoma (HCC) cells decreased sulfation of cell-surface HSPGs and reduced growth signalling (Lai, Chien et al. 2003). In addition, increased *SULF1* expression has been associated with increased overall survival from breast cancer, and poorly invasive tumours such as lobular carcinomas (Khurana, Beleford et al. 2013).

The development of targeted therapeutics is dependant on the identification of genetic and microenvironmental changes involved in the initiation, progression and malignant conversion of cancers (Haupt and Griffiths 2009). The ability of cancers to exploit HSPG function within their cells makes SNPs within HSPG genes potential markers of cancer disease susceptibility. Here, we examined the SULF1 SNP rs2623047 in Australian Caucasian breast cancer cohorts using two independent breast cancer case/control populations. Initial genotyping was conducted on the Genomics Research Centre Breast Cancer population with results replicated in the Griffith University-Cancer Council Queensland Breast Cancer Biobank population. Genotyping was performed using PCR-RFLP analysis to examine the potential of this SNP as a marker for breast cancer susceptibility.

Materials and Methods

Populations

All individuals comprising the two populations are of Caucasian (Northern European) origin. The initial Genomics Research Centre breast cancer (GRC-BC) population consisted of 243 breast cancer patient samples and 201 age and sex matched control samples. A subset of the Griffith University-Cancer Council Queensland Breast Cancer Biobank (GU-CCQ BB) population was used as a replication population and consisted of 443 case samples and 91 age and sex matched controls (Youl, Baade et al. 2011).

In collaboration with the Cancer Council Queensland, the Genomics Research Centre has collected samples for the GU-CCQ BB population as part of a 5-year population-based longitudinal study of women newly diagnosed with breast cancer. Recruitment commenced in January 2011 with 920 women aged 33 to 80 years (average age 60.2 years) available for this study. Study participants are

residents of Queensland with a histologically confirmed diagnosis of invasive breast cancer. Clinical and demographic information was obtained from the Queensland Cancer Registry and diagnostic and treatment information was obtained through telephone interviews with participants and medical record extraction. The matching control population includes women with no personal or familial history of cancer aged 32 to 88 years, with an average age of 60.2 years. These women were recruited through the Genomics Research Centre from January 2000.

Preparation of DNA samples from Blood

DNA was extracted from blood samples using a modified salting out method (Nasiri, Forouzandeh et al. 2005, Chacon-Cortes, Haupt et al. 2012). Quality and quantity of isolated DNA was measured by spectrophotometry using a Nanodrop (Thermo Scientific, Australia). If required DNA samples were further purified by ethanol precipitation as described (Buckingham 2007).

SNP selection and primer design

The *SULF1* SNP rs2623047 was identified following consideration of a number of HSPG SNPs where a minor allele frequency (MAF) greater than 0.05 was considered during the selection and design process. This SNP is a 5' near gene polymorphism significantly associated with early onset age and longer progression free survival in ovarian cancers (Han et al., 2011). Chromosomal location and MAF for this SNP can be found in Table 1. Primers were designed using NCBI Primer Blast with the sequences F (5'-GGGATGCACAGAAACCCTAA-3') and R (5'-TGTGGCAAACAGTGAAGAGC-3') used to amplify a 291bp fragment.

PCR Amplification

PCR amplification of the region surrounding the *SULF1* SNP (rs2623047) was conducted under the following conditions: 40ng of DNA was amplified with 100nM each forward and reverse primers (IDT, USA), 200nM dNTPs (NEB, Australia), 1.75mM $MgCl_2$, 0.5U GoTaq® Flexi DNA polymerase (Promega, Australia), 1x PCR buffer in a 15 µL reaction. An initial 3 min denaturation step at 95^{o}C was followed by 35 cycles of denaturation at 95^{o}C for 45s, annealing at 58^{o}C for 45s and extension at 72^{o}C for 45s. A final extension step of 7 min at 72^{o}C completed the cycling. These conditions produced a single 291bp fragment. Following amplification, the PCR product was held at 4^{o}C until genotyping analysis.

Restriction Fragment Length Polymorphism (RFLP) analysis

Genotyping was conducted using restriction fragment length polymorphism (RFLP) analysis. Following amplification, approximately 1µg (7µL) PCR product was digested with 1U *Psp*GI enzyme for 4hr at 75^{o}C with 1x reaction buffer in a 15µL reaction. The *Psp*GI enzyme recognises the sequence CCWGG and cuts the amplicon when the wild type (C) allele is present creating bands of 212bp and 78bp. The enzyme is unable to cut the fragment when the mutant (T) allele is present.

Agarose Gel Electrophoresis

To confirm amplification of the fragment of interest, the PCR product was run on 3% agarose gels in 1x Tris-acetate-EDTA (TAE) buffer at 90V for 45 min. For genotyping analysis following RFLP, the digested PCR product was run on a 4% agarose gel in 1x TAE at 70V for 60min for increased resolution. A 100bp DNA ladder was included for sizing purposes with DNA fragments visualised following the addition of ethidium bromide and excitation under UV light.

Statistical tests

Allele frequencies in case and control populations were determined. Hardy-Weinberg Equilibrium (HWE) (Kalmes R February, 2001) was used to test for deviation between observed and expected frequencies. A Chi-squared analysis (Fisher and Yates 1963) was conducted to test for significant differences between case and control populations and to determine if the alleles or genotypes were significantly associated with breast cancer (α=0.05). The odds ratio at a confidence interval of 95% was calculated to indicate disease risk.

Results

The *SULF1* SNP rs2623047 was initially analysed the GRC breast cancer cohort. Both case and control populations were determined to be in HWE (case, p=0.17; control, p=0.38). Our observed frequencies in the control population closely matched those found on HapMap for a Caucasian population with a calculated Odds Ratio (OR) of 0.72. When Chi-squared analysis was conducted to determine association, a significant difference between the case and control populations at both the genotypic and allelic level was observed (summarised in Table 2).

We then examined the SNP in an independent replication population, the GU-CCQ BB cohort. Once again, both case and control samples were determined to be in HWE (case, p=0.53; control,

p=0.46). Chi-squared analysis determined borderline significance at the allelic level (*p*=0.057) and no significant difference in genotype frequencies (*p*=0.15) with an OR of 0.72. These results are summarised in Table 3.

Due to the small number of control samples in the GU-CCQ BB cohort, we combined the data from the two independent populations and analysed them together, increasing the power of the study. When analysed together, the combined population once again reached significance. Both combined cases and combined controls were demonstrated to be in HWE (case, *p*=0.77; control, *p*=0.27) with chi-square analysis determining significant differences between case and control samples at both the allelic and genotypic level including an odds ration of 0.77. The calculated odds ratios obtained suggest the presence of the T allele within the rs2623047 SNP is associated with an increased risk of developing breast cancer. These combined results are summarised in Table 4.

Discussion

In this study we examined the potential association of the *SULF1* SNP rs2623047 in breast cancer susceptibility. Our results demonstrated a significant difference in allele (*p*=0.03) and genotype (*p*=0.03) frequencies in an Australian Caucasian population. Results of a replication study found a similar trend, although this study failed to reach levels for significance. There was no association at the allelic level, however genotype frequencies showed borderline significance (p=0.057). However, when the two populations were combined to increase the power of the population, a significant difference in allele and genotype frequencies was found (allele, *p*=0.016; genotype, *p*=0.032). The lack of significance in the replication population could be due to the low number of controls (n=80). However when data from the two populations was combined for analysis, association of this SNP and breast cancer susceptibility was identified.

Biosynthesis of HSPGs is a complex process with mutations and alterations to expression of a number of genes at various stages of this complex process previously associated with disease. *SULF1* encodes heparan sulfate 6-*O*-endosulfatase 1, responsible for 6-*O* desulfation of HS chains (Isidor, Pichon et al. 2010). A number of genetic modifications have been reported in the *SULF1* gene with SNPs within this gene associated with ovarian cancer, particularly with age of onset, suggesting its variations may have roles in prognosis and onset of the disease (Han, Huang et al.

2011). *SULF1* has also been shown to inhibit tumour growth in hepatocellular carcinoma (HCC) through desulfation of cell surface of HSPGs resulting in the downregultion of HCC cell growth (Lai, Yu et al. 2006). In addition, *SULF1* overexpression in gastric cancer has been suggested to correlate with the oncogene *MYC* amplification in HCCs, as both are located in the chromosomal region 8q, frequently amplified in gastric cancers (Junnila, Kokkola et al. 2010).

Members of the two major HSPG core protein families, the glypicans (attached through a GPI anchor to the cell membrane) and the syndecans (transmembrane proteins found in the cell surface and the extracellular matrix), have also been associated with disease. Glypican-3 (*GPC3*) acts as a cell proliferation inhibitor and apoptosis inducer in several tumour cell types, with its gene expression down-regulated in various types of tumours, including mesotheliomas and ovarian cancer as well as Simpson-Golabi-Behmel syndrome (Gonzalez, Kaya et al. 1998, Cano-Gauci, Song et al. 1999, Filmus 2001). Up-regulation of *GPC3* has also been associated with cancer with enhanced expression observed in thyroid cancer, indicating a tumour suppressive role, while silencing *GPC3* in breast cancer demonstrated a negative regulatory role on cell growth (Xiang, Ladeda et al. 2001, Yamanaka, Ito et al. 2007). These observed roles of *GPC3* appears to be through its interactions with the Wnt signalling where it is able to inhibit both the canonical and non-canonical pathways (Schambony, Kunz et al. 2004, Stigliano, Puricelli et al. 2009). This has also been demonstrated *in vitro*, where HSPG mediated human breast cancer cell line proliferation and migration were shown to be mediated through interactions with specific members of the Wnt pathway (Okolicsanyi, van Wijnen et al. 2013), interactions that may be modified as a result of changes to HS chain sulfation.

SDC1 encodes the HSPG syndecan 1 (Zhang, McKown et al. 2011), the most studied member of the syndecan family of HSPGs (Gallagher 2001) thought to have an important role in cancer progression (Zhang, McKown et al. 2011). Examination of dense breast tissue has demonstrated higher expression of *SDC1*, suggesting overexpression is related to breast cancer with tissue density a risk factor in breast cancer development (Lundstrom, Sahlin et al. 2006). In addition, increased *SDC1* expression has been associated with poorer prognosis for breast cancer patients suggesting a role for this HSPG in more malignant and higher-grade breast cancer tumours (Lendorf, Manon-Jensen et al. 2011). Interestingly, *SDC1* mediated endocytosis has been shown to be dependent on the presence of *N*- and 6-*O* sulfation of *SDC1* HS chains (Makkonen, Turkki et al.

2013). In addition, this was shown to be specific only to the *SDC1* core HSPG protein (Makkonen, Turkki et al. 2013).

Importantly, both these HSPG core protein families, the syndecans and glypicans, are reliant on the fine structure of their HS chains for their molecular interactions and biological functions. In Wnt-signalling, glypicans stabilise the interaction of Wnt with its receptor Frizzled (Filmus, Capurro et al. 2008). The syndecans often act through interactions with the fibroblast growth factor (FGF) family of growth factors, which require specific sulfation sites on HS chains to enable binding to their signalling partners (Guimond, Maccarana et al. 1993). Important roles for 6-*O* sulfation, and therefore *SULF1* have also been identified in FGF signalling. For example, FGF2-FGFR complex binding requires 2-*O* sulfation, while interactions with PGDF require 6-*O* sulfation (Lindahl, Kusche-Gullberg et al. 1998). Interactions of FGF1-FGFR2 require 6-*O* sulfation and FGF2-FGFR1 requires both 6-*O* and 2-*O* sulfation (Pellegrini, Burke et al. 2000). The ability of FGF-2 to bind its 'high affinity' receptor (FGFR-1) to stimulate growth is greatly decreased in the absence of appropriately sulfated HSPGs (Rapraeger, Krufka et al. 1991, Ornitz, Yayon et al. 1992). *QSulf1*, the avian homologue of mammalian *SULF1*, has been shown to promote Wnt signalling by modulating the binding affinity of Wnts to HS chains. This promotes HS-mediated initiation of signalling through presentation of Wnt to its receptor, Frizzled (Ai, Do et al. 2003). *SULF1* was also shown to inhibit FGF signalling activity in both Xenopus and chicken embryos (Wang, Ai et al. 2004). These studies suggest a role for *SULF1* as a positive regulator of Wnt signalling and a negative regulator of FGF signalling (Lin 2004).

These examples also demonstrate the importance of the sulfation pattern of HS side chains for their biological function. As such, the enzymes regulating both sulfation and desulfation of HSPG side chains have a critical impact on the regulation of a number of cellular processes. The mutation of the allele C to T in *SULF1* in the rs2623047 SNP results in changed heparan sulfate endosulfatase function and the removal of 6-*O*-sulfate groups from heparan sulfate chains. As such, modifications to the HS chain through the action of enzymes such as *SULF1* influence cell-cell and cell-matrix interactions in both healthy and disease tissues. Further examination of the protein levels of SULF1 in tumour tissue would add to our understanding of the involvement of this protein in breast cancer progression. In addition, studies examining the effect of modification of HS sulfation through the addition of sulfation inhibitors such as sodium chlorate to *in vitro* models may provide a better insight into the role of SULF1 and other HSPGs in breast cancers,

including their interaction with specific growth factors such as the FGFs in terms of downstream signalling affecting cell proliferation, differentiation and migration of tumour cells.

Conclusion

Breast cancer is an often-fatal disease affecting a significant number of women worldwide. With genetic susceptibility one of the numerous factors contributing to the development of this disease we examined the SNP (rs2623047) in *SULF1* as a potential marker of genetic susceptibility in this disease. In the first study of its kind investigating SNPs in the gene encoding the HSPG modification enzyme, our results demonstrate that the T allele of rs2623047 in *SULF1* is associated with an increased risk of developing breast cancer. Identification of markers including those within central roles in the stroma and matrix surrounding breast tumour cells may enable improved detection of this disease at an earlier stage to improve treatment regimes and patient prognosis.

References

ABS. (2011, 14/03/2013). "Leading Causes of Death by Gender." 3303.0 Causes of Death, Australia, 2011 Retrieved 04/03/2014, 2013, from **http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/F25E446E1BE6C931CA257B2E000D729C?opendocument**.

Ai, X., A. T. Do, O. Lozynska, M. Kusche-Gullberg, U. Lindahl and C. P. Emerson, Jr. (2003). "QSulf1 remodels the 6-O sulfation states of cell surface heparan sulfate proteoglycans to promote Wnt signaling." J Cell Biol **162**(2): 341-351.

Bernfield, M., M. Gotte, P. W. Park, O. Reizes, M. L. Fitzgerald, J. Lincecum and M. Zako (1999). "Functions of cell surface heparan sulfate proteoglycans." Annu Rev Biochem **68**: 729-777.

Blackhall, F. H., C. L. Merry, E. J. Davies and G. C. Jayson (2001). "Heparan sulfate proteoglycans and cancer." Br J Cancer **85**(8): 1094-1098.

BreastCancer.org. (2010, Sept 26, 2013). "Breast Cancer Risk Factors." 2013, from **http://www.breastcancer.org/symptoms/understand_bc/risk/factors**.

Buckingham, L. (2007). Common Techniques in Molecular Biology. Nucleic Acid Extraction Methods**:** 65-79.

CancerAustralia. (2006, 8 October, 2013). "Breast Cancer - CancerAustralia." 2013, from **http://www.canceraustralia.gov.au/affected-cancer/cancer-types/breast-cancer**.

Cano-Gauci, D. F., H. H. Song, H. Yang, C. McKerlie, B. Choo, W. Shi, R. Pullano, T. D. Piscione, S. Grisaru, S. Soon, L. Sedlackova, A. K. Tanswell, T. W. Mak, H. Yeger, G. A. Lockwood, N. D. Rosenblum and J. Filmus (1999). "Glypican-3-deficient mice exhibit developmental overgrowth and some of the abnormalities typical of Simpson-Golabi-Behmel syndrome." J Cell Biol **146**(1): 255-264.

Chacon-Cortes, D., L. M. Haupt, R. A. Lea and L. R. Griffiths (2012). "Comparison of genomic DNA extraction techniques from whole blood samples: a time, cost and quality evaluation study." Mol Biol Rep **39**(5): 5961-5966.

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D and F. Bray. (2013). "GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet].", from Available from: **http://globocan.iarc.fr**.

Ferlay, J., H. R. Shin, F. Bray, D. Forman, C. Mathers and D. M. Parkin (2010). "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008." Int J Cancer **127**(12): 2893-2917.

Fernandez-Vega, I., O. Garcia, A. Crespo, S. Castanon, P. Menendez, A. Astudillo and L. M. Quiros (2013). "Specific genes involved in synthesis and editing of heparan sulfate proteoglycans show altered expression patterns in breast cancer." BMC Cancer **13**: 24.

Filmus, J. (2001). "Glypicans in growth control and cancer." Glycobiology **11**(3): 19R-23R.

Filmus, J., M. Capurro and J. Rast (2008). "Glypicans." Genome Biol **9**(5): 224.

Fisher, R. and F. Yates (1963). Statistical tables for biological, agricultural and medical research. Edinburgh, UK, Oliver and Boyd.

Gallagher, J. T. (2001). "Heparan sulfate: growth control with a restricted sequence menu." J Clin Invest **108**(3): 357-361.

Gonzalez, A. D., M. Kaya, W. Shi, H. Song, J. R. Testa, L. Z. Penn and J. Filmus (1998). "OCI-5/GPC3, a glypican encoded by a gene that is mutated in the Simpson-Golabi-Behmel overgrowth syndrome, induces apoptosis in a cell line-specific manner." J Cell Biol **141**(6): 1407-1414.

Guimond, S., M. Maccarana, B. B. Olwin, U. Lindahl and A. C. Rapraeger (1993). "Activating and inhibitory heparin sequences for FGF-2 (basic FGF). Distinct requirements for FGF-1, FGF-2, and FGF-4." J Biol Chem **268**(32): 23906-23914.

Han, C. H., Y. J. Huang, K. H. Lu, Z. Liu, G. B. Mills, Q. Wei and L. E. Wang (2011). "Polymorphisms in the SULF1 gene are associated with early age of onset and survival of ovarian cancer." J Exp Clin Cancer Res **30**: 5.

Haupt, L. M. and L. R. Griffiths (2009). "Heparan Sulfate Proteoglycans, Tumour Progression and the Cancer Stem Cell Niche." Current Cancer Therapy Reviews **5**(4): 256-260.

Isidor, B., O. Pichon, R. Redon, D. Day-Salvatore, A. Hamel, K. A. Siwicka, M. Bitner-Glindzicz, D. Heymann, L. Kjellen, C. Kraus, J. G. Leroy, G. R. Mortier, A. Rauch, A. Verloes, A. David and C. Le Caignec (2010). "Mesomelia-synostoses syndrome results from deletion of SULF1 and SLCO5A1 genes at 8q13." Am J Hum Genet **87**(1): 95-100.

Junnila, S., A. Kokkola, T. Mizuguchi, K. Hirata, M. L. Karjalainen-Lindsberg, P. Puolakkainen and O. Monni (2010). "Gene expression analysis identifies over-expression of CXCL1, SPARC, SPP1, and SULF1 in gastric cancer." Genes Chromosomes Cancer **49**(1): 28-39.

Kalmes R, H. J. (February, 2001). "Modèle de Hardy-Weinbe." Atlas Genet Cytogenet Oncol Haematol. .

Khurana, A., D. Beleford, X. He, J. Chien and V. Shridhar (2013). "Role of heparan sulfatases in ovarian and breast cancer." Am J Cancer Res **3**(1): 34-45.

Lai, J., J. Chien, J. Staub, R. Avula, E. L. Greene, T. A. Matthews, D. I. Smith, S. H. Kaufmann, L. R. Roberts and V. Shridhar (2003). "Loss of HSulf-1 up-regulates heparin-binding growth factor signaling in cancer." J Biol Chem **278**(25): 23107-23117.

Lai, J. P., C. Yu, C. D. Moser, I. Aderca, T. Han, T. D. Garvey, L. M. Murphy, M. M. Garrity-Park, V. Shridhar, A. A. Adjei and L. R. Roberts (2006). "SULF1 inhibits tumor growth and potentiates the effects of histone deacetylase inhibitors in hepatocellular carcinoma." Gastroenterology **130**(7): 2130-2144.

Lendorf, M. E., T. Manon-Jensen, P. Kronqvist, H. A. Multhaupt and J. R. Couchman (2011). "Syndecan-1 and syndecan-4 are independent indicators in breast carcinoma." J Histochem Cytochem **59**(6): 615-629.

Lin, X. (2004). "Functions of heparan sulfate proteoglycans in cell signaling during development." Development **131**(24): 6009-6021.

Lindahl, U., M. Kusche-Gullberg and L. Kjellen (1998). "Regulated diversity of heparan sulfate." J Biol Chem **273**(39): 24979-24982.

Lundstrom, E., L. Sahlin, L. Skoog, T. Hagerstrom, G. Svane, E. Azavedo, K. Sandelin and B. von Schoultz (2006). "Expression of Syndecan-1 in histologically normal breast tissue from postmenopausal women with breast cancer according to mammographic density." Climacteric **9**(4): 277-282.

Makkonen, K. E., P. Turkki, J. P. Laakkonen, S. Yla-Herttuala, V. Marjomaki and K. J. Airenne (2013). "6-o- and N-sulfated syndecan-1 promotes baculovirus binding and entry into Mammalian cells." J Virol **87**(20): 11148-11159.

Morimoto-Tomita, M., K. Uchimura, Z. Werb, S. Hemmerich and S. D. Rosen (2002). "Cloning and characterization of two extracellular heparin-degrading endosulfatases in mice and humans." J Biol Chem **277**(51): 49175-49185.

Nasiri, H., M. Forouzandeh, M. J. Rasaee and F. Rahbarizadeh (2005). "Modified salting-out method: high-yield, high-quality genomic DNA extraction from whole blood using laundry detergent." J Clin Lab Anal **19**(6): 229-232.

Okolicsanyi, R. K., A. J. van Wijnen, S. M. Cool, G. S. Stein, L. R. Griffiths and L. M. Haupt (2013). "Heparan Sulfate Proteoglycans and Human Breast Cancer Epithelial Cell Tumorigenicity." J Cell Biochem.

Ornitz, D. M., A. Yayon, J. G. Flanagan, C. M. Svahn, E. Levi and P. Leder (1992). "Heparin is required for cell-free binding of basic fibroblast growth factor to a soluble receptor and for mitogenesis in whole cells." Mol Cell Biol **12**(1): 240-247.

Pellegrini, L., D. F. Burke, F. von Delft, B. Mulloy and T. L. Blundell (2000). "Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin." Nature **407**(6807): 1029-1034.

Rapraeger, A. C., A. Krufka and B. B. Olwin (1991). "Requirement of heparan sulfate for bFGF-mediated fibroblast growth and myoblast differentiation." Science **252**(5013): 1705-1708.

Schambony, A., M. Kunz and D. Gradl (2004). "Cross-regulation of Wnt signaling and cell adhesion." Differentiation **72**(7): 307-318.

Stigliano, I., L. Puricelli, J. Filmus, M. C. Sogayar, E. Bal de Kier Joffe and M. G. Peters (2009). "Glypican-3 regulates migration, adhesion and actin cytoskeleton organization in mammary tumor cells through Wnt signaling modulation." Breast Cancer Res Treat **114**(2): 251-262.

Sugahara, K. and H. Kitagawa (2002). "Heparin and heparan sulfate biosynthesis." IUBMB Life **54**(4): 163-175.

Wang, S., X. Ai, S. D. Freeman, M. E. Pownall, Q. Lu, D. S. Kessler and C. P. Emerson, Jr. (2004). "QSulf1, a heparan sulfate 6-O-endosulfatase, inhibits fibroblast growth factor signaling in mesoderm induction and angiogenesis." Proc Natl Acad Sci U S A **101**(14): 4833-4838.

Wiseman, B. S. and Z. Werb (2002). "Stromal effects on mammary gland development and breast cancer." Science **296**(5570): 1046-1049.

Xiang, Y. Y., V. Ladeda and J. Filmus (2001). "Glypican-3 expression is silenced in human breast cancer." Oncogene **20**(50): 7408-7412.

Yamanaka, K., Y. Ito, N. Okuyama, K. Noda, H. Matsumoto, H. Yoshida, A. Miyauchi, M. Capurro, J. Filmus and E. Miyoshi (2007). "Immunohistochemical study of glypican 3 in thyroid cancer." Oncology **73**(5-6): 389-394.

Youl, P. H., P. D. Baade, J. F. Aitken, S. K. Chambers, G. Turrell, C. Pyke and J. Dunn (2011). "A multilevel investigation of inequalities in clinical and psychosocial outcomes for women after breast cancer." BMC Cancer **11**: 415.

Zhang, Y., R. L. McKown, R. W. Raab, A. C. Rapraeger and G. W. Laurie (2011). "Focus on molecules: syndecan-1." Exp Eye Res **93**(4): 329-330.

Tables and Figures

Table 1: Chromosomal location and allele information for SULF1- rs2623047

SNP	Gene	Gene location	Wild Type Allele	Mutant Allele	Chromosomal Position	MAF
rs2623047	SULF1	8q13.1	C	T	70378496 Chromosome 8	0.474

*Table 2: Genotypes for SULF-1 (*rs2623047*) obtained from the GRC Breast cancer population*

	Allele			***Genotype***			
	T (%)	*C (%)*	*p-value*	*TT (%)*	*CT (%)*	*CC (%)*	*p-value*
Control	208 (59.4)	142 (42.6)	**0.028**	59 (31.1)	90 (56.6)	26 (16.4)	***0.027***
Case	270 (63.2)	132 (40.6)		95 (46.6)	80 (39.2)	26 (12.7)	
HapMap	58.4	41.6		32.7	51.3	15.9	

*Table 3: Genotypes for SULF-1 (*rs2623047*) obtained from the GHI Biobank cohort*

	Allele			***Genotype***			
	T (%)	*C (%)*	*p-value*	*TT (%)*	*CT (%)*	*CC (%)*	*p-value*
Control	87 (54.4)	73 (45.6)	**0.057**	22 (27.5)	43 (53.8)	15 (18.8)	*0.15*
Case	503 (62.4)	303 (37.6)		154 (38.2)	195 (48.4)	54 (62.4)	
HapMap	58.4	41.6		32.7	51.3	15.9	

*Table 4: Genotypes for SULF-1 (*rs2623047*) obtained from the combined populations*

	Allele			***Genotype***			
	T (%)	*C (%)*	*p-value*	*TT (%)*	*CT (%)*	*CC (%)*	*p-value*
Control	295 (57.8)	*215 (42.2)*	***0.016***	*81 (31.8)*	*133 (52.2)*	*41 (16)*	***0.032***
Case	773 (64.0)	*435 (36.0)*		249 (41.2)	275 (45.5)	*80 (13.3)*	
HapMap	58.4	41.6		32.7	51.3	15.9	