The Chr X-linked *Gct6* locus: A granulosa cell tumor suppressor in mice

by

© Zoha Rabie, B.Sc.

A thesis submitted to the
School of Graduate Studies
in partial fulfillment of the requirements
for the degree of

Master of Science in Medicine

Faculty of Medicine

Memorial University of Newfoundland

May 2015

St. John's, NL

Abstract

Juvenile-type granulosa cell tumors (GCT) originate in the somatic tissues that surround the germ cells of the ovarian follicle in children and young women. Unlike Adult-type GCT that share a common, acquired mutation in the *FOXL2* gene, the genetic determinants for juvenile-type GCT susceptibility are not so well defined. A spontaneous, early-onset GCT phenotype in SWR inbred female mice has revealed multiple *Gct* loci: *Gct1* on Chromosome (Chr) 4 initiates the tumorigenic program; *Gct4* on Chr X modifies trait penetrance and *Gct6* on Chr X is a suppressor of GCT initiation. The *Gct6* locus has been mapped to a 1.02 million base pair region with over 20 annotated genes. Two complementary approaches were taken to help prioritize *Gct6* gene candidates. First, a congenic strain approach was taken to determine the tumor suppressor phenotype of the $Gct6^{C57}$ allele in SWR.C57-X females. Second, a whole-locus capture and Next Generation Sequencing (NGS) protocol was applied to identify all nucleotide variations between SWR (tumor-susceptible SW) and Castaneus (CAST; tumor-resistant CA) alleles vs. the C57BL/6J (C57) genome. The congenic strain data supported the hypothesis that a common *Gct6* tumor suppressor allele in C57 and CAST genomes was distinct from the permissive $Gct6^{SW}$ allele, which reduced the prioritized list of NGS variants (SNPs, INDELs) to 746. Priority was given to non-synonymous variants identified in the coding region of 5 candidate genes: *BC065397, Esx1, Slc25a53, Tmsb15b1 and Tmsb15l*. Sanger sequencing confirmed the 5 non-synonymous variants between C57 and SWR strains and identified a novel 28 bp deletion in the coding region of the *Slc25a53* gene. Allele association analysis between 5 *Gct6* permissive and 2 *Gct6* suppressive strains demonstrated that the 5 non-synonymous variants were not conserved between permissive strains. However, all 5 permissive strains tested (SWR, SJL/J, PL/J, BUB/BnJ, ST/bJ) had deletion mutations with inferred deleterious frameshift effect in the *Slc25a53* gene when compared to the GC tumor suppressive strains (CAST and C57), making *Slc25a53* a promising candidate for shared identity with *Gct6*.

Acknowledgments

I sincerely thank my supervisor, Dr. Ann Dorward for accepting me as a graduate student and providing me with the means and opportunity to present this work at the 7th Canadian Conference on Ovarian Cancer Research (CCOCR 2014). I am truly grateful for her guidance, support, and mentorship throughout all stages of my program.

The support and coordination of this program by Dr. Tabrizchi, Associate Dean of Research and Graduate Studies – Faculty of Medicine, is greatly appreciated. I am thankful to Dr. Yaskowiak for the training I received and deeply appreciative of the cooperative work he did towards the progress of this research. The financial support from Research & Development Corporation of Newfoundland and Labrador (RDC), Canadian Institutes of Health Research (CIHR), and The Medical Research Endowment Fund (MRF) is gratefully acknowledged. The support and resources provided by The Jackson Laboratory (Bar Harbor, Maine) and Genomics and Proteomics Facility, Core Research and Equipment and Instrument Training Network, Memorial University of Newfoundland is also greatly appreciated.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ºC	Degrees Celsius
µL	Microliter
µM	Micromolar
17β-HSD1	hydroxysteroid (17- β) dehydrogenase 1
AA	Amino Acid
A	Adenine
AF	Allele Frequency
AGCT	Adult-type Granulosa Cell Tumor
AGCT	Adult-type Granulosa Cell Tumor
AIS	Androgen Insensitivity Syndrome
Ala	Alanine
Ar	Androgen receptor gene
Arg	Arginine
BM	Basement membrane
bp	Base pair
BRCA1	Human breast cancer 1, early onset gene
BRCA2	Human breast cancer 2, early onset gene
C	Cytosine
C57	C57BL/6 inbred mouse strain
C57	C57 strain alleles
CA	CAST strain alleles
CAST	Castaneus inbred mouse strain
Chr	Chromosome
cm	Centimeter(s)
CO_2	Carbon dioxide
CYP19A1	Cytochrome P450 Family 19 Subfamily A Polypeptide 1 gene
CYP19A1	Cytochrome P450 Family 19 Subfamily A Polypeptide 1 protein
Cys	Cysteine
dbSNP	Single Nucleotide Polymorphism Database
DHEA	Dehydroepiandrosterone
DHT	Dihydrotestostrone
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dpc	Days post coitum
E	Embryonic day
E_2	17β-estradiol
EDTA	Ethylenediaminetetraacetic acid
F	Filial generation
FOXL2	Human forkhead box L2 protein
FOXL2	Human forkhead box L2 gene
FSH	Follicle Stimulating Hormone
g	Gravity

g	Grams
GC	Granulosa cell
GCT	Granulosa Cell Tumor
Gct	Granulosa cell tumor susceptibility locus
Gct1	Granulosa cell tumor susceptibility locus 1
Gct4	Granulosa cell tumor susceptibility locus 4
Gct6	Granulosa cell tumor susceptibility locus 6
GnRH	Gonadotropin-Releasing Hormone
h	Hour
H	Height
HBOC	Hereditary Breast Ovarian Cancer syndrome
HCl	Hydrochloric acid
HPG	hypothalamic- anterior pituitary-gonadal
INDEL	Insertion or deletion of DNA bases
J	SJL strain alleles
JGCT	Juvenile-type Granulosa Cell Tumor
kDa	Kilo Dalton, atomic mass unit
L	Length
Leu	Leucine
LH	Luteinizing Hormone
M	Molar
Mbp	Million base pairs
MCs	Mitochondrial carriers
mg	Milligrams
min	Minute
mL	Milliliter
mm	Millimeter
mM	Millimolar
n	Sample size
N	Backcross generation
NCBI	National Center for Biotechnology Information
ng	Nanograms
NGS	Next Generation Sequencing
nm	Nanometer
OMIM	Online Mendelian Inheritance in Man
p	p-value
PCR	Polymerase chain reaction
PGC	Primordial Germ Cell
Phe	Phenylalanine
pmol	Picomole
Pro	Proline
RSPO1	R-spondin 1
RT	Room temperature
s	Second
SCST	Sex cord-stromal cell tumor
Ser	Serine

SJL	SJL inbred strain
SNP	Single Nucleotide Polymorphism
SOX9	Sex-determining region Y-box gene 9 transcription gene
SRY	Sex-determining region Y gene
SSLP	Simple Sequence Length Polymorphism
SW	SWR strain alleles
SWR	SWR inbred mouse strain
Taq	*Thermus aquaticus*
TBE	Tris Borate Ethylenediamineteraacetic acid
TC	Thecal Cell
TE	Tris Ethylenediamineteraacetic acid
TGFβ	Transforming Growth Factor β
Thr	Threonine
UTR	Untranslated region
V	Volt
Val	Valine
W	Width
W/V	Weight/Volume, mass concentration
wk	Week
Wnt4	Wingless-type MMTV integration site family, member protein

List of Appendices

1.0 Introduction

1.1 The Ovary

1.1.1 Ovarian Anatomy and Physiology

The ovary is a specialized female reproductive organ with two major functions, gametogenesis and steroidogenesis. Gametogenesis is the production and release of mature haploid gametes whereas steroidogenesis is the production and release of steroid hormones. To perform its specialised functions, the ovary is composed of germ cells (oocytes) and companion somatic cells. The germ cells help establish follicular organization and the lack of germ cells in the developing ovaries results in the absence of the formation of ovarian follicles (Piprek, 2009).

The ovary is composed of two general regions: the inner medulla, a highly vascularized region that contains lymphatic vessels, nerves, and connective tissue, and the outer cortex, where follicles proceed along a course of maturation under hormonal control (Figure 1.1). An ovarian follicle is made up of an oocyte, surrounded by granulosa cell layers (GCs), and an outer basement membrane (BM) surrounded by additional layers of thecal cells (TCs). Working together, TCs and GCs synthesize estrogen to control the maturation of additional follicles, promote the appearance of the secondary sex characteristics of females at puberty, maintain a woman's reproductive organs in their mature functional state, and influence the ovarian cycle. Upon cessation of woman's reproductive ability and exhaustion of the follicular pool menopause occurs.

During the female reproductive years, the normal functions of the ovary are controlled by a series of complex interactions between the hypothalamus, the anterior pituitary gland and

locally produced factors in the gonad such as hormones, growth factors, and cytokines (Figure 1.2).

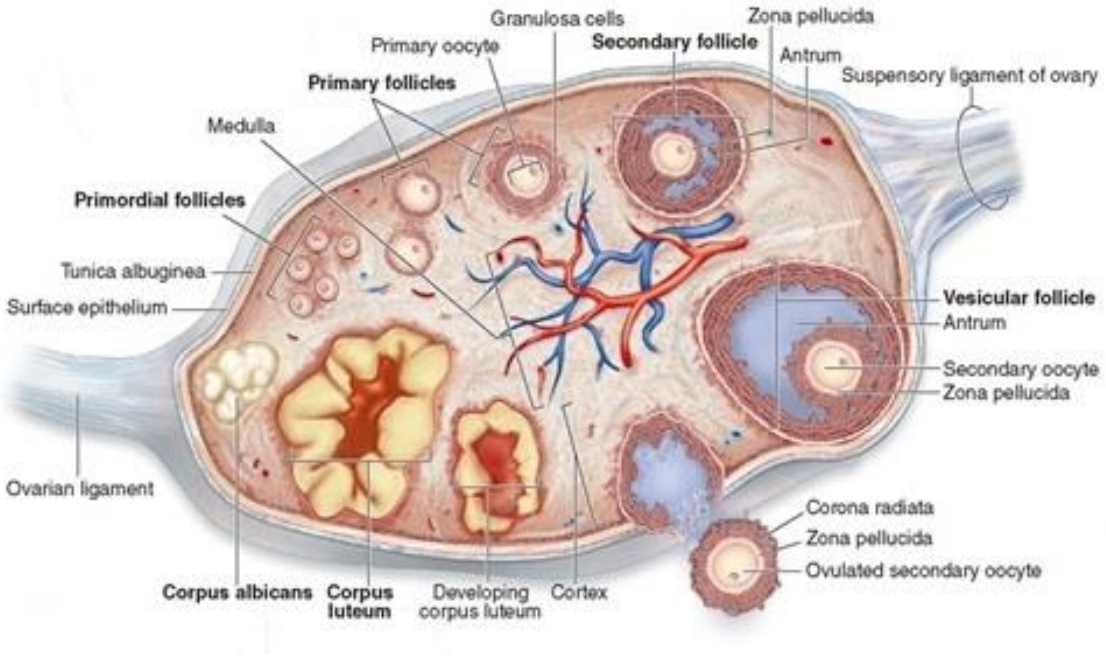


Figure 1.1 Human ovarian anatomy and cellular components

The ovary is held in place by the ovarian ligament and suspensory ligament. The cortex of the ovary is covered by a layer of epithelium, while the inner most region of the ovary is highly vascularised and contains maturing and resting follicles, lymphatic vessels, nerves, and connective tissue. The ovary is covered by surface epithelium; beneath the surface epithelium is tunica albuginea which is the connective tissue covering of the ovaries. The zona pellucida is a glycoprotein layer surrounding the plasma membrane of oocytes. During follicogenesis, an ovarian follicle passes through the following stages: primordial, primary, secondary (pre-antral), tertiary (antral) and finally the pre-ovulatory (Vesicular) follicle stage. The formation of a fluid-filled cavity adjacent to the oocyte called the antrum designates the follicle as an antral follicle. Upon ovulation and the release of the mature egg from the ovary, the remaining follicular GCs undergo luteinisation and switch from estrogen to progesterone production (corpus luteum) to maintain early pregnancy.

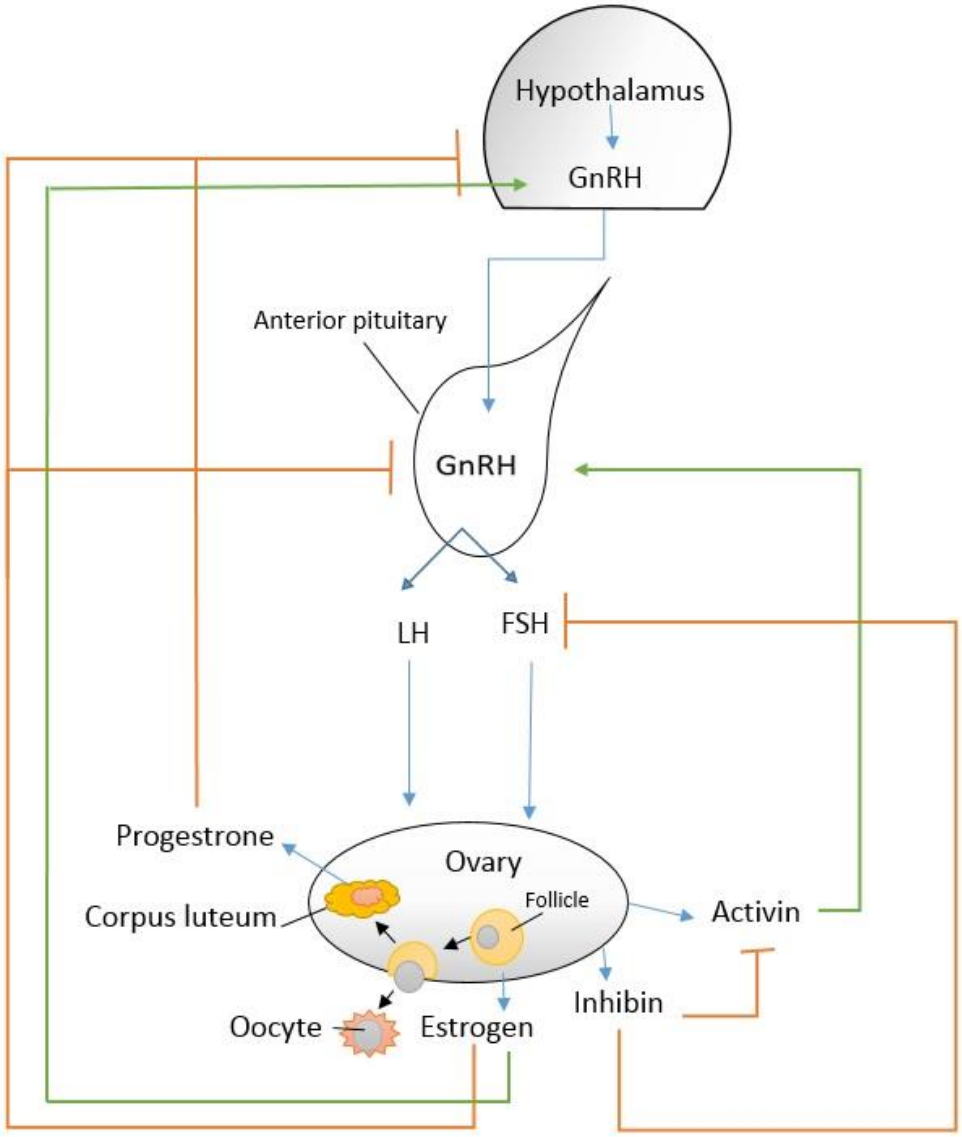


Figure 1.2 Hypothalamic-pituitary-gonadal (HPG) axis

The function of normal ovary is controlled by complex interactions between the hypothalamus, the anterior-pituitary and locally produced factors such as hormones, growth factors, and cytokines. GnRH released from hypothalamus travels to anterior pituitary which stimulates secretion of LH and FSH into the general circulation and the gonads produce estrogen and progesterone. The gonads, in response to FSH and LH, also produce inhibin and activin proteins. Activin enhances FSH biosynthesis and secretion, and participates in the regulation of menstrual cycle. Inhibin down regulates FSH synthesis and inhibits FSH secretion.

The regulatory endocrine loop between these organs is called the hypothalamic-pituitary-gonadal (HPG) axis. One important function of the hypothalamus is to release Gonadotropin-Releasing Hormone (GnRH), which binds to specific receptors on the gonadotrope cells within the anterior pituitary. The gonadotropes induce the synthesis and release of gonadotropins, Follicle Stimulating Hormone (FSH) and Luteinizing Hormone (LH), into the systemic blood circulation. Both FSH and LH are heterodimeric glycoproteins; the alpha subunits of both FSH and LH are composed of 92 *amino acids* (AA) while the β subunits vary. FSH has a β subunit of 111 AA (FSH β), which promotes its specific biologic action and is responsible for interaction with the *FSH-receptor* whereas LH has a β subunit of 120 AA (LH β) that confers its biologic action and is responsible for the specificity of the interaction with the LH receptor. After FSH and LH reach the ovary, they influence follicular maturation, ovulation, and corpus luteum formation through binding to their corresponding receptors expressed on the surface of target cells. Additionally, GCs of the ovary produce peptide-hormones such as inhibin and activin that are involved in ovarian hormone synthesis regulation (Drummond et al., 2012). Inhibin is composed of two partially homologous disulfide linked subunits; α and β_A or β_B thus creating inhibin A and B. Activin is made up of a dimer of disulfide-linked β subunits A or B creating activin A, B, and/or AB. Activins promote differentiation and proliferation of GCs and expression of FSH receptors in normal ovaries (Risbridger et al., 2001). Inhibin acts primarily to inhibit the secretion of FSH by the anterior pituitary gland. The relationship between inhibin and FSH represents a typical negative feedback mechanism since the major action of FSH is to stimulate the formation and function of GCs (Drummond et al., 2012).

1.1.2 Folliculogenesis

The maturation process of an ovarian follicle is called folliculogenesis. During follicular development, the GCs surrounding the oocyte proliferate and form several layers to nuture the the developing oocyte and serve the endocrine function of the ovary. TCs are arranged in layers that surround the follicular BM and engage in paracrine signalling with the GCs of the follicle. The process of follicular maturation begins when the single layer of flattened GCs surrounding the oocyte change into a layer of cuboidal GCs (Verlhac *et al.,* 2010; Figure 1.3). The plasma membrane of an oocyte in a primary follicle is surrounded by a glycoprotein membrane known as zona pellucida (ZP). After development of the primary follicle, the GCs continue to proliferate and stay connected by tight junctions to form a protective barrier around the oocyte (Baerwald *et al.,* 2012). The follicle develops into a secondary follicle when another layer of GCs is present outside of the ZP membrane (Verlhac *et al.,* 2010). The fluid filled patches enlarge in secondary follicles to form an antral space in which the oocyte and some GCs migrate to one side of the follicular cavity (Verlhac *et al.,* 2010). At this stage the follicle is called a pre-ovulatory follicle. Thereafter, connective tissue surrounding the GCs differentiates, forming a layer of TCs. As the follicle matures, it moves closer to the ovarian surface in preparation for ovulation. The GCs that surround the oocyte during ovulation are called cumulus cells and those remaining behind mural cells. The mural population undergoes luteinisation post ovulation, developing into the corpus luteum, which acts as a temporary endocrine gland secreting estrogen and progesterone to support the earliest steps of gestation (Baerwald *et al.,* 2012).

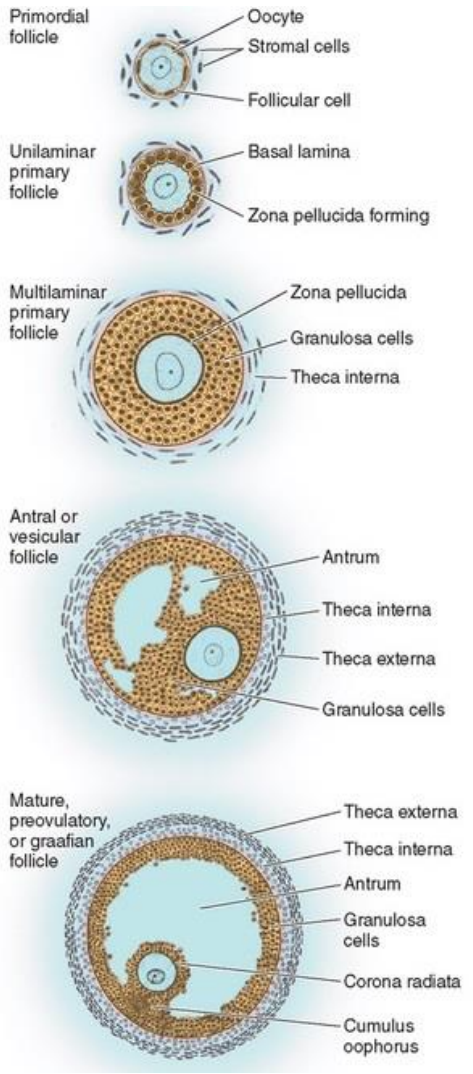


Figure 1.3 Ovarian folliculogenesis

The maturation of the ovarian follicle begins with the recruitment of non-growing primordial follicles. The GCs of primordial follicles change from a flat (follicular cells) to a cuboidal structure (basal lamina), the follicle is considered a primary follicle. The plasma membrane of an oocyte in a primary follicle is surrounded by a glycoprotein membrane known as zona pellucida (ZP). The formation of zona pellucida membrane begins at early unilaminar primary follicle stage. At multilaminar stage, multiple layers of GCs and TCs proliferate surrounding the oocyte. The TCs are divided into two layers, the theca interna and the theca externa. In response to follicle-stimulating hormone, follicular fluid accumulates in the antrum forming antral or vesticular follicle. After cavitation, the mature (preovulatory or graafian) follicle develops a polarity and populations of granulosa cells of the tertiary follicle undergo differentiation into four distinct subtypes: corona radiata, surrounding the zona pellucida; membrana, interior to the basal lamina (not shown in the figure); periantral, adjacent to the antrum (not shown in the figure) and cumulus oophorous, which connects the membrana and corona radiata granulosa cells together.

The first phases of follicular development are controlled by groups of growth factors, including Transforming Growth Factor *β* family members (TGF-β; Myers *et al.,* 2010). Activin and Inhibin are subunits which form dimeric proteins that are members of the TGF-β superfamily of growth and differentiation factors (O'Shea, 1981). Under the regulation of LH, TCs synthesize androgens, but cannot convert them into estrogen without the help of Cytochrome P450, Family 19, Subfamily A Polypeptide 1 (CYP19A1 or aromatase; see Figure 1.4). Similarly, GCs under the action of FSH (also by LH at the late stage) synthesize progesterone, although they are unable to convert it to androgen due to lack of the necessary converting enzymes. Thus, GCs and TCs work collaboratively to synthesize androgens into estrogen. This collaboration starts when TCs convert cholesterol into androstenedione, which is then transported into the GCs and converted into estrogen. Besides this obligatory role of androgens as estrogen precursors in steroidogenesis, very little is known about their direct involvement in the female body. Vendola *et al.* in 1999 used female monkeys to investigate the interactions between FSH and androgens in follicular development through *in situ* hybridization. Vendola *et al.* proved that androgens promote follicular growth and estrogen biosynthesis indirectly, by amplifying the effects of FSH. Additionally, Cárdenas *et al.* in 2007 examined the expression of the androgen receptor (Ar) and FSH receptor (FSHR) in late developing follicles in pigs and concluded that androgens could directly promote follicle growth (Cárdenas *et al.,* 2007). The Androgen- Receptor Knockout (ARKO) mice have significant reproductive defects and reduced fertility suggesting that androgen signaling is important for normal female reproductive health. Sen *et al.* (2014) used GC-specific ARKO mouse model to demonstrate that androgens regulate follicular development and female fertility by reducing rates of follicular atresia while simultaneously promoting preantral follicle growth and development. However, the specific mechanisms by which

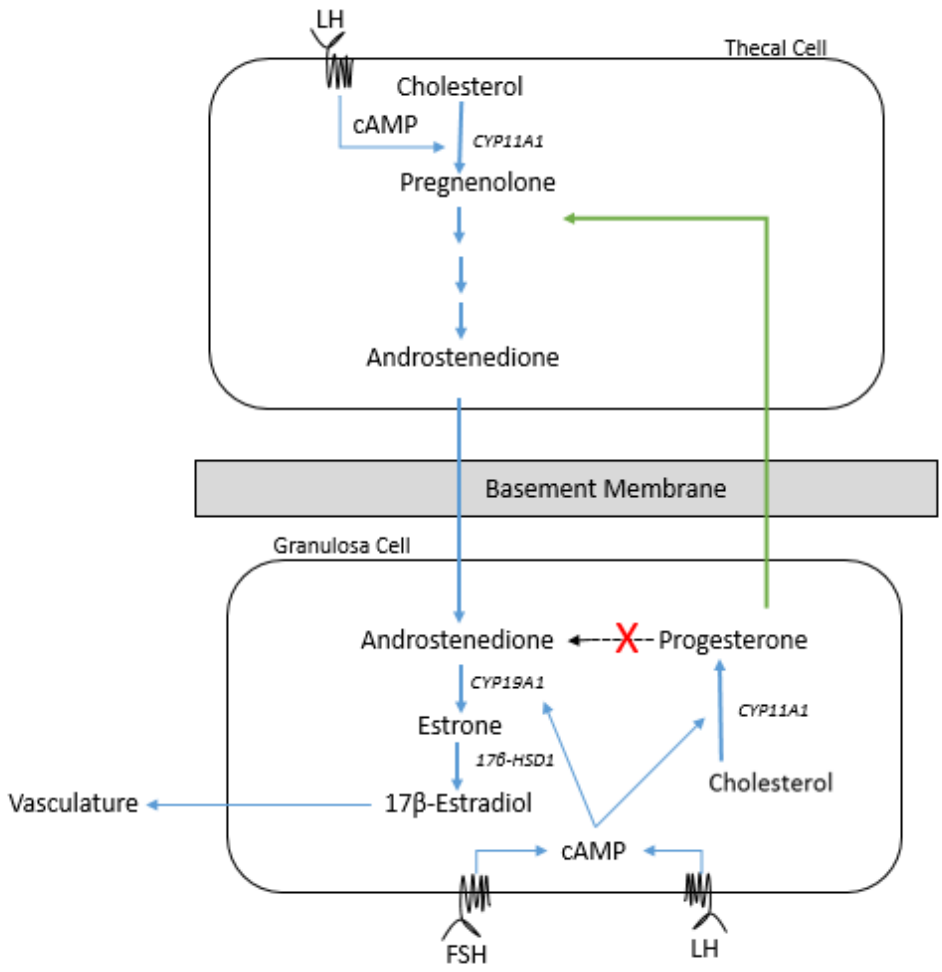


Figure 1.4 Interactions between GC and TC in the follicle

The pre-ovulatory follicle produces estrogen through a paracrine interaction between TC and GC. Two critical steps in estrogen formation are the first entry of cholesterol into the TC which is converted to pregnenolone by Cyclic Adenosine MonoPhosphate (cAMP) and cytochrome P450 superfamily of family 11, subfamily A, polypeptide 1 (CYP11A1). cAMP stimulates protein kinase A family of enzymes which leads to increased concentrations of cholesterol in the TCs. Pregnenolone is then converted to androstenedione through multiple conversion reactions (multiple arrows) which is then transported into GCs. The second critical step occurs in GCs which is the conversion of androstenedione to Estrone by Cytochrome P450, family 19, subfamily A, polypeptide 1 (CYP19A1). Estrone is then converted to 17β-estradiol (E_2) catalyzed by hydroxysteroid (17- β) dehydrogenase 1 (17β-HSD1) and diffused into vasculature.

androgens mediate female fertility remain unclear and the roles for androgenic signalling through the Ar receptor have not been examined specifically during the distinct phases of embryonic follicle assembly and cyclic follicular maturation in the post-pubertal ovary.

Hirshfield was one of the first reproductive biologists to suggest the ovarian follicle complement could be divided into two categories based on their growth characteristics. Hirshfield *et al.* investigated the patterns of ovarian cell proliferation during the earliest stages of folliculogenesis in rats and concluded that the cortical primordial follicles are activated gradually to provide mature ova over the entire course of animal's reproductive life, while the medullary follicles had a shorter span of existence with different proliferation kinetics (Hirshfield *et al.,* 1995). A recent report by Zheng *et al.* using fluorescent transgene tracing methods in the mouse supported Hirshfield's findings, with convincing evidence there are two distinct waves of primordial follicle assembly that have different developmental dynamics (Zheng *et al.,* 2014). Generally, the most up-to-date model suggests the first wave of follicles to mature in the rodent ovary are the primordial follicles that reside in the medulla; these follicles are synchronously activated to provide mature eggs until approximately 3 months of age, they contribute to the onset of puberty and the establishment of cyclic HPG axis signalling, but they are less likely to contribute to fertility (Zheng *et al.,* 2014). The cortical primordial follicles, on the other hand, are formed from the recruited supporting cells in the cortical region of postnatal ovaries and contribute to ovulation and fertility post-puberty. Rodent models have been instrumental for our understanding of the major genetic and endocrine factors that influence mammalian ovarian development and function, but there is always the potential for species-specific differences. In humans, for instance, ovulation of one oocyte is most common, while in multi-ovulatory species (mice, rats, cats, dogs), several oocytes are ovulated from each ovary, indicating different criteria

or thresholds for dominant follicle selection leading to full maturation and ovulation. The establishment of the two-wave theory of follicle development in rodent species opens the door for the identification of molecular markers to further explore the unique properties of each wave, and to confirm the two-wave model of follicle recruitment in human ovaries.

1.1.3 Ovarian Development and the Maintenance of Female Gonad Specification

The complex process of ovarian development starts with differentiation and specification of Primordial Germ Cells (PGCs), the precursors to germ cells, in the early embryo (McGraw-Hill Companies, Inc. 2010). PGCs are cells that migrate to the bilateral genital ridges, adopt a sex-specific gamete fate under appropriate signals and subsequently orchestrate the somatic cell organization in the embryonic gonad. The differentiation and development of PGCs is crucial for assuring normal fertility of the individual and the correct transmission of the genome to the next generation. As such, the steps of female and male gonad specification remain active areas of reproductive research. Mice and humans have very similar gonad development pathways, which when combined with the reproductive and engineering potential of mice, makes them an excellent embryological model to study gonad specification or pathology (Quinn *et al.*, 1998).

In the early weeks of embryogenesis in humans, the embryonic disc undergoes a process of folding and PGCs are passively incorporated into the embryo together with the yolk sac wall. PGCs then migrate through the dorsal mesentery during early gastrulation to reach the developing gonads, the genital ridges, where they colonize and are rapidly surrounded by cords of somatic cells. Sexually undifferentiated mammalian embryos can develop into either female or male; this stage is called bi-potential or an indifferent stage, in which the appearance of the gonad is essentially the same in the two sexes (Gilbert, 2003). In humans, this indifferent stage is from 4 to 7 weeks (wks) of gestation. At the time of gonad colonization, Chr XY PGCs express

the Sex-determining region Y (Sry*)* protein, which triggers differentiation of the gonad towards testis fate through its target gene, the Sex-determining region Y-box 9 (Sox9) transcription factor (Figure 1.5). In contrast, Chr XX PGCs differentiate through an active process regulated by Wingless-type MMTV integration site family member 4 (Wnt4) and R-spondin 1 (Rspo1), which up-regulate the expression of β-catenin (Ctnnb1). Another female-sex-determining factor is Forkhead Box L2 (Foxl2), which is expressed as a nuclear protein in the granulosa cells of the embryonic and adult mouse and human ovary (Lima *et al*., 2012). The Wnt4/Rspo1 and Foxl2 signaling pathways act in a complementary manner to promote ovarian development and maintain somatic cell identity in the female ovary through suppression of the *Sox9* gene (Nef *et al.,* 2009). *FOXL2* mRNA expression was demonstrated by Duffin *et al.* to be between 8 and 19 wks of gestation in the human fetal ovary (Duffin *et al.,* 2009). Foxl2 protein expression in mice, however, begins at embryonic day (E) 12.5 in pre-granulosa somatic cells (Rafa, 2009) and is persistent through the reproductive lifespan. Evidence from human syndromes and other mammals has confirmed these pathways are not only important for the initial ovarian specification, but also for maintenance of female somatic cell identity and reproductive function. For instance, *Foxl2* mutations are involved in Chr XX sex reversal in goats and depletion of ovarian follicles in mice and humans (Garcia-Ortiz *et al.,* 2009). Ablation of the *Foxl2* gene causes partial secondary sex reversal in mice and partial loss of FOXL2 function leads premature ovarian failure in humans (OMIM® and Nef *et al.,* 2009). The concept that sex specification requires specific signals for maintenance as well as establishment represented a change of thinking for the reproductive biology community; it also raised the possibility that a signaling imbalance could lead to a potentially tumorgenic phenotype, as one outcome of uncontrolled transdifferentiation.

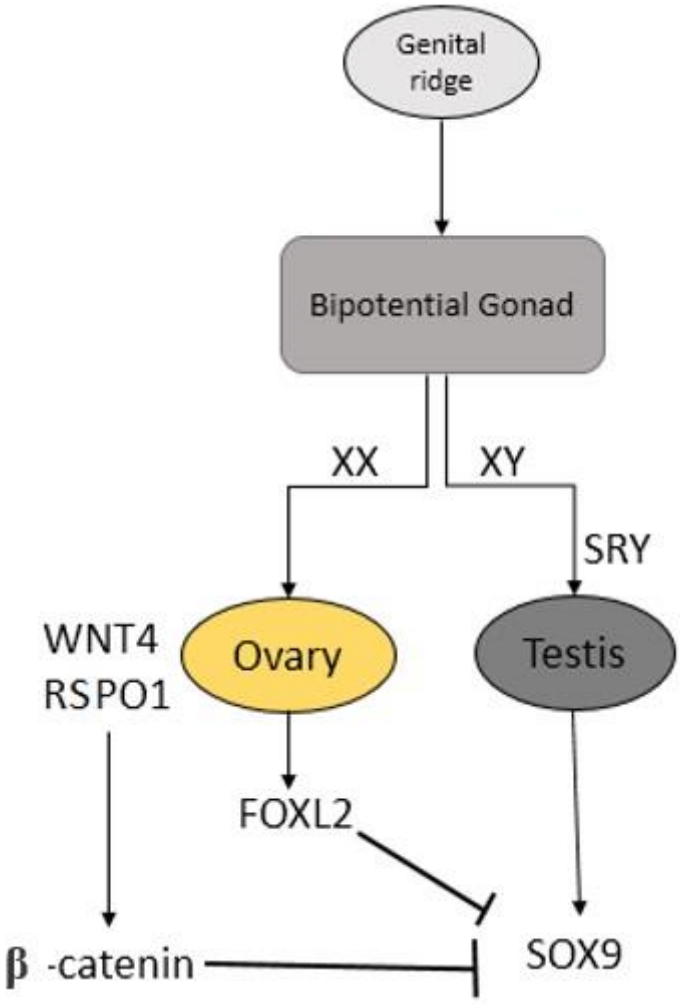


Figure 1.5 Sex determination in the bipotential mammalian gonad

The embryo with Chr XX genotype develops ovaries through Wnt4 (Wingless-type MMTV integration site family, member 4) and RSPO1 (R-spondin 1) stabilization of β-catenin which prevents the SRY induced expression of SOX9 and testis formation. The differentiation of ovarian somatic cells is actively maintained by FOXL2 (Forkhead box L2) while suppressing SOX9. β-catenin stabilized by WNT4 and RSPO1 signaling suppresses SOX9 expression in female gonads. (Adapted from Piprek, 2009)

1.2 Ovarian Cancer

In Canada, approximately 2,600 diagnoses of ovarian cancer are recorded annually, representing 2.9 %, or the eighth most common cause, of all new cancers in Canadian women (Canadian Cancer Statistics 2014). Even more disturbing, ovarian cancer ranks as the fifth leading cause (4.7 %) of cancer-related deaths in women. Ovarian cancer is considered to be the most serious of all gynecological cancers since the majority of women with ovarian cancer are not diagnosed until the disease is very advanced and has metastasized; this is because symptoms usually do not become apparent until the tumor compresses or invades adjacent structures, ascites develops, or metastases become clinically evident. The prognosis of ovarian cancer remains very poor with 5-year survival rate of only 45 % (Erickson *et al.,* 2013). Since ovarian cancer has been treated as one disease for decades, and that has led to no improvements in mortality, the most promising way continuing to battle ovarian cancer is to consider improvements in molecular genetic diagnoses, to identify the subtypes of ovarian cancer and treat them as distinct diseases using targeted therapies.

Ovarian cancer has three major subclasses based on cell populations in the ovary: epithelial, germ cell, and sex cord-stromal. Approximately 90 to 95 % of ovarian tumors diagnosed in women have an epithelial origin (Serov *et al*., 1973 and Tavassoli *et al.,* 2003). Epithelial ovarian tumors are further grouped into five histological types: serous, mucinous, endometrioid, clear cell and transitional cell tumors (Brenner tumors). Serous epithelial carcinomas of the ovaries are the most common type, making up about two thirds of the cases diagnosed in women. Epithelial ovarian cancers were all thought to originate from epithelial cells that cover the ovaries, but more recent research has shown that a high percentage of serous

epithelial tumors have migrated from primary cancers of the fallopian tube epithelium (Kuhn *et al.,* 2013 and Kurman *et.al.* 2010).

Sex cord-stromal cell tumors (SCST) are the second most common form of ovarian cancer, accounting for approximately 5-10 % of all cases in women (Leung *et al.,* 2004). This class of ovarian cancer encompasses tumors that may originate in the GCs, TCs, or a mixed population of stromal cells that are endocrinologically active; therefore, SCST ovarian cancers are often associated with abnormal production of sex steroid hormones, leading to clinical symptoms of precocious puberty in pre-pubertal girls, or excessive uterine bleeding and/or virilisation in post-pubertal women (Young *et al.,* 1984). The most common types of SCSTs are GC tumors, TC tumors (Thecomas) and Sertoli-Leydig cell tumors, wherein somatic cells of the female ovary take on the characteristics of Sertoli-Leydig cells, the somatic cell counterpart of the testis. Germ cell tumors are the rarest form of ovarian cancer, affecting approximately 1 % of women (Leung *et al.,* 2004). As a consequence of their rarity, SCSTs and germ cell tumors are less well understood than epithelial ovarian cancers, in terms of their genetic or environmental etiology. In contrast, epithelial ovarian cancer investigations have been spurred on by their association with heritable cancer syndromes and the identification of susceptibility genes and pathways useful for cancer subclassifications and targeted therapy development.

A family history is a risk factor for ovarian cancer; such that if a woman has two or more close relatives (mother, sister or daughter) with ovarian cancer, her risk to develop ovarian cancer is increased (Lynch *et al.,* 2009). There are three hereditary syndromes known to predispose a woman to epithelial ovarian cancer. Two are named as Hereditary Breast-Ovarian Cancer (HBOC) syndromes, as a result of mutations in the tumor suppressor genes *Breast Cancer 1, Early Onset (BRCA1) and Breast Cancer 2, Early Onset (BRCA2*), accounting for

approximately 90 % of ovarian cancer diagnoses due to a hereditary syndrome (Buller *et al.*, 2002 and Lynch *et al.*, 2009). Both BRCA1 and BRCA2 proteins play a role in DNA double strand break repair, such that deleterious mutations in these factors contribute to cancer risk as a result of acquired DNA damage (O'Donovan and Livingston, 2010). Female patients with Lynch Syndrome also exhibit increased risk for epithelial ovarian cancer, along with other cancer sites such as the colonic epithelium. Lynch syndrome is associated with mutations in genes of the DNA mismatch repair pathway, including: *MutL homolog 1 (MLH1), MutS protein Homolog 2 (MSH2), MutS Homolog 6 (MSH6), PostMeiotic Segregation increased 2 (PMS2)* or *Epithelial Cell Adhesion Molecule (EPCAM)* genes. Defective DNA mismatch repair results in microsatellite instability (MSI), which is a useful diagnostic indicator. Mutations in *BRCA1, BRCA2* and other tumor suppressor genes predisposes women primarily to epithelial ovarian cancer, but not to other classes of ovarian cancer, suggesting different pathways are important for sex cord-stromal and germ cell tumor initation (Lynch *et al.*, 2009)

1.2.1 Granulosa Cell Tumors of the Ovary

Sex cord-stromal GC tumors are broadly classified into two subtypes: Juvenile-type GC tumors (JGCTs) that occur in children or young women, and Adult-type GC tumors (AGCTs) that occur in peri- and postmenopausal women. AGCTs have histological proliferation of GCs which are often with a stromal component of fibroblasts, theca or luteinized cells (Tavassoli *et al*., 2003 and Wheater *et al.*, 1979). AGCTs may recur up to 40 years after diagnosis (Ud Din *et al.*, 2014 and Rebstock *et al.*, 2014). This recurrence feature of AGCTs is not fully understood, but is a significant cause of late mortality.

In 2009, high throughput sequencing analysis led to the identification of a common, somatic mutation (c.402C>G) in the *FOXL2* gene that is present in almost all morphologically

identified human AGCTs but not JGCTs (Shah *et al.*, 2009 and Köbel *et al.*, 2009). The *FOXL2* mutation changes a highly conserved cysteine residue to a tryptophan (p.C134W), implicating a key loss of protein function, although the mechanism is not yet fully elucidated (Kim *et al.*, 2011). There are several hypotheses under investigation to explain the consistent association of this unique mutation in *FOXL2* with AGCT development. Fleming *et al.* (2011) deduced that mutant FOXL2 protein increases transcription of the target gene aromatase (*CYP19A1*) but with a reduced capacity to induce apoptosis, thereby compromising GC death (Fleming *et al*., 2011). Mutation in *FOXL2* gene causes the GCs to lose the ability to modulate cell cycle which results in decreased apoptosis in GCs (Benayoun *et al.*, 2011 and Kim *et al.*, 2011).

JGCTs are the least common form of GC tumor, accounting for approximately 5 % of all the GCT diagnoses. Relative to AGCTs, the tumor cells of JGCTs have a high proliferation rate, a moderate-to-high degree of cellular atypia but are also hormonally active (Tavassoli *et al*., 2003). Patients frequently have elevated levels of androgen or E_2, leading to symptoms of virilisation or precocious puberty, respectively, with a concomitant decrease in circulating gonadotropins (Leyva-Carmona *et al*., 2009). Analysis of a cohort of 125 JGCTs showed that 44 % of the cases occurred in the first decade of life, 34 % in the second, 18 % in the third, and only 3 % after age 30 years (Young *et al*., 1984). The occurrence of JGCTs in infants and young girls in the absence of the somatic *FOXL2* mutation suggests that JGCTs have a different genetic or environmental etiology compared with AGCTs.

As a cancer of children and young adults, there is a strong rationale to search for JGCT susceptibility genes; however, the relative rarity of JGCTs has hindered human genetic linkage studies. In rodents, SCSTs are the most common spontaneous ovarian neoplasms (Greenacre, 2004 and Thayer *et al.*, 2007). GC tumors are observed in advanced aged (12-24 months) female

rodents, near the end of the reproductive life, akin to human AGCTs (Tavassoli, *et al*., 2003 and Beamer *et al.*, 1985). Beamer *et al*. reported one non-engineered, inbred mouse strain called SWR that exhibits spontaneous development of juvenile-onset GCT (Beamer *et al.*, 1985). Therefore, the SWR mouse model has presented a unique opportunity for researchers to study genetic and endocrine mechanisms related to initiation and support of tumor growth that are applicable to human JGCTs.

1.2.2 The SWR Mouse Model of GC Tumorigenesis

Beamer *et al*. (1985) first described GCTs that appeared spontaneously in a proportion (≤ 1 %) of young females of the SWR inbred strain (Figure 1.6). The same group confirmed that manipulation of the endocrine environment increased trait penetrance, such that 3 wk old SWR inbred females or recombinant inbred strain females derived from SWR both exhibited significantly increased GCT incidence (approximately 20 %), with dehydroepiandrosterone (DHEA) or testosterone exposure (Beamer *et al*., 1988; Beamer *et al*., 1993). This elevated GCT incidence was deemed sufficient to investigate the biology of the spontaneous GCTs and perform genetic analysis for the GC tumor susceptibility loci (Beamer *et al.*, 1988).

In young SWR female mice, GCT initiation takes place at puberty, age 3-4 weeks, and by 8-10 weeks the majority of females that were destined to develop GCTs had done so, while unaffected females retained normal fertility through adulthood (Beamer *et al*., in 1985). The GC tumors that develop in this mouse model have striking similarities to human JGCTs (Fox, 1985). In both species, GC tumors occur spontaneously and at a young age in the presence of normal ovarian follicles; furthermore, they display similar histopathology, are associated with disturbed endocrine activity and have malignant properties (Beamer *et al*., 1988).

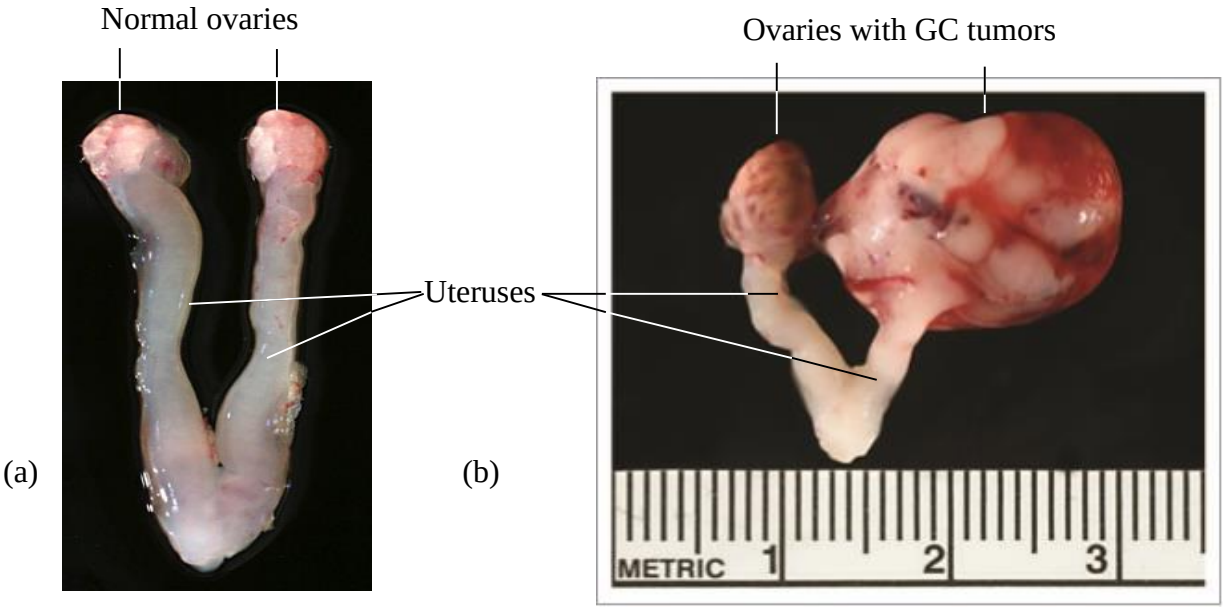


Figure 1.6 Normal and bilateral early-onset GC tumors derived from SWR female mouse

(a) Normal ovaries with uterus attached, taken from an SWR-derived female mouse at 8 wks of age (b) Isolated bilateral GC tumors with uterus attached, taken from an SWR-derived female mouse at 8 wks of age. At 8 wks, GCTs are large (~10 mm^3), easily visible, and highly vascularized (Used with permission from Dr. Dorward's Laboratory, 2012).

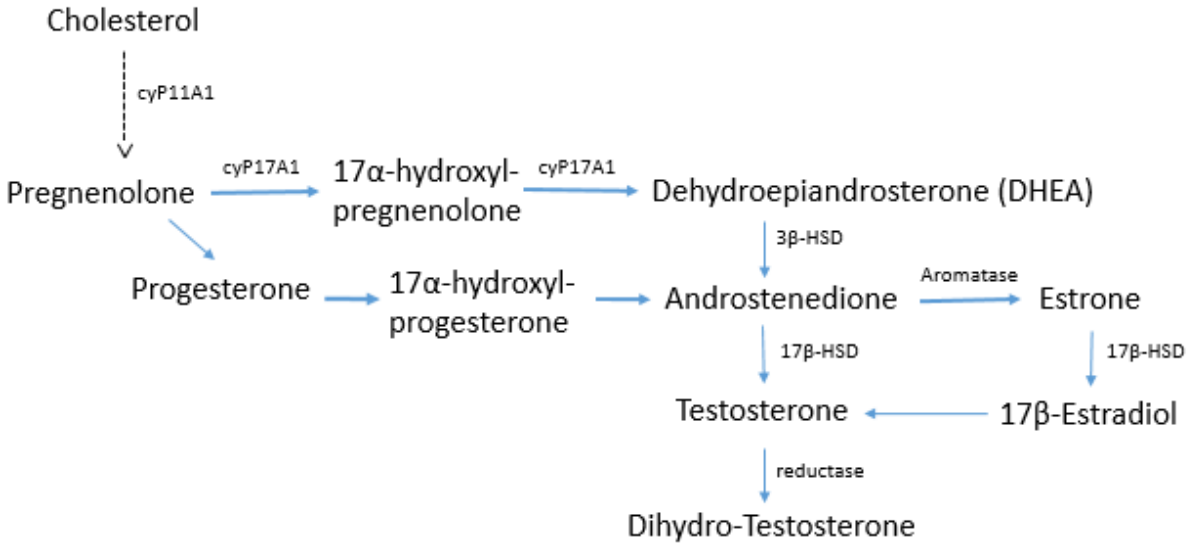


Figure 1.7 Steroid biosynthesis pathway

Schematic diagram of the steroid biosynthetic pathways leading to androgen production that takes place in the TCs and GCs in the ovaries. In humans, the main pathway to androgen production is through conversion of 17-hydroxypregnenolone to dehydroepiandrosterone (DHEA) rather than through conversion of 17-hydroxyprogesterone to androstenedione. Subsequent testosterone biosynthesis can occur through conversion of DHEA to androstenedione (by 3β-hydroxysteroid dehydrogenase type 2), followed by the actions of 17β-hydroxysteroid dehydrogenase type 3 to generate T, or via the intermediate metabolite androstenediol.

The SWR female mice present a very unique and valuable heritable GC tumor susceptibility phenotype compared to other inbred or engineered mouse models currently available to study GC tumorigenesis. One of the most distinct and unusual phenotypes of SWR mice is that GC tumors spontaneously develop without any genetic manipulation. Beamer *et al.* hypothesized there must be one or a group of low penetrance GCT susceptibility (*Gct*) genes controlling the initiation and development of GCTs. Another unique phenotype of SWR mice is

that the GCT initiation is restricted to puberty, suggesting that *Gct* susceptibility genes have only a very short window to influence GC fate and initiate tumorigenesis (Beamer *et al.,* 1988). Beamer and Tennent postulated, at the time of puberty in SWR female mice, that follicular maturation occurs abnormally in the first cohort of maturing follicles, leading to follicular breakdown and uncontrolled GC proliferation, instead of follicular atresia typically observed for the first wave of follicles post-puberty. However, follicular atresia in general does not correlate with GC tumor susceptibility in SWR females, since ongoing follicular atresia that occurs during the reproductive cycles of adulthood does not trigger GC tumorigenesis (Dorward *et al.,* 2003). The SWR model indicates there is a discrete developmental window of JGCT susceptibility, with tumor initiation risk limited to the peri-pubertal stage. In combination with the evidence for two distinct follicle populations in the mammalian ovary, it is a reasonable hypothesis to suggest that the earliest, medullary follicle cohort contributes to JGCT susceptibility in SWR female mice. The pattern of diminishing GCT susceptibility with post-pubertal age in human cohorts of JGCT cases mirrors what is observed in the SWR strain, indicating that GC tumor susceptibility genes identified in SWR mice may provide a unique opportunity to investigate human JGCT etiology at the molecular level

1.2.2.1 Genetics of GCT susceptibility in the SWR strain

Genetic mapping studies (Beamer *et al.,* 1988, Tennent *et al.,* 1993, and Dorward *et al.,* 2013) designed to correlate genotype with phenotype have clearly demonstrated that GC tumor susceptibility is a polygenic and heritable trait in SWR mice, involving multiple autosomal and Chr X-linked loci (Table 1.1.). Three independent *GC tumor susceptibility loci* - *Gct1* on Chr 4, *Gct4* and *Gct6* on Chr X – have a major influence over GCT susceptibility in SWR female mice, and so were investigated further using congenic strains that isolate the independent regions for

the purpose of high resolution genetic mapping and assignment of specific functions to the various *Gct* alleles.

1.2.2.2 Gct1: The driver mutation for GC tumorigenesis

A locus named *Gct1* on distal mouse Chr 4 derived from the SWR genome was strongly associated with GC tumor initiation in three independent mapping crosses (Beamer *et al.,* 1988; Beamer *et al.,* 1998; Dorward *et al.,* 2005). Consequently, the SWR allele at *Gct1* (*Gct1*^{SW}) was deemed necessary for susceptibility to GC tumor initiation in SWR mice and is considered to be the driver mutation for the JGCT phenotype (Beamer *et al*., 1988b; Beamer *et al.,* 1988a; Beamer *et al*,. 1993; Dorward *et al.,* 2005). The mapping studies revealed that the *Gct1*^{SW} locus had semi-dominant activity, with required inheritance of at least one copy to confer GCT susceptibility in young female mice, and therefore fitting with the description of an oncogenic locus (Beamer *et al*., 1998). The *Gct1*^{SW} locus also exhibits a distinct endocrine interaction with the androgenic metabolite dehydroepiandrosterone (DHEA), implying the genetic determinant within the *Gct1*^{SW} locus is more likely to initiate the GC tumorigenic program in the presence of DHEA (Beamer *et al*., 1988b). The *Gct1*^{SW} locus is unique in this regard, since JGCTs do not appear when young female mice of other inbred strains (Figure 1.8) - C57BL/6J (C57) or Castaneus/Ei (CAST) - are similarly treated with DHEA at puberty (Table 1.1; Beamer *et al.,* 1998, Dorward *et al.,* 2005 and Smith *et al.,* 2013).

Table 1.1 Chromosomal location and DNA markers for *Gct* loci identified in multiple mapping studies

Gct locus	Chr *	Position (Mbp)	Associated DNA Marker	Strain used for mapping	*Gct* allele activity
Gct1[1,2,4,6]	4	145,057	*D4Mit232*	SWR, SJL, CAST	GC tumor permissive
Gct2[1]	12	46.113	*D12Mit172*	SWR, SJL	GC tumor permissive
Gct3[1]	15	64.537	*D15Mit133*	SWR, SJL	GC tumor permissive
Gct4[1,2]	X	99.231	*DXMit96*	SWR, SJL, CAST	GC tumor permissive
Gct5[1]	9	115.121	*D9Mit17*	SWR, SJL	GC tumor permissive
Gct6[3,5]	X	137.618	*DXamd14*	SWR, SJL, CAST	GC tumor suppressor
Gct7[4]	1	78.206	*D1Mit215*	SWR, SJL, CAST	GC tumor permissive
Gct8[4]	2	166.394	*D2Mit145*	SWR, SJL, CAST	GC tumor permissive
Gct9[4]	13	111.900	*D13Mit292*	SWR, SJL, CAST	GC tumor permissive

* Chromosomal positions were obtained through the Ensembl website (GRCm38; Ensembl Genome Browser release 75, Flicek *et al.,* 2014).

[1] Beamer W.G. *et al.,* 1988. Cancer Research 48:2788-2792.

[2] Beamer W.G. *et al.,* 1998. Cancer Research 58:3694-3699.

[3] Dorward A.M., *et al.,* 2003 Cancer Research 63:8197-8202.

[4] Dorward A.M., *et al.,* 2005 Cancer Research 65:1259-1264.

[5] Dorward A.M., *et al.,* 2013. Epigenetics 8:184–191.

[6] Smith K.N., *et al.,* 2013. Mammalian Genome. 24: 63-71.

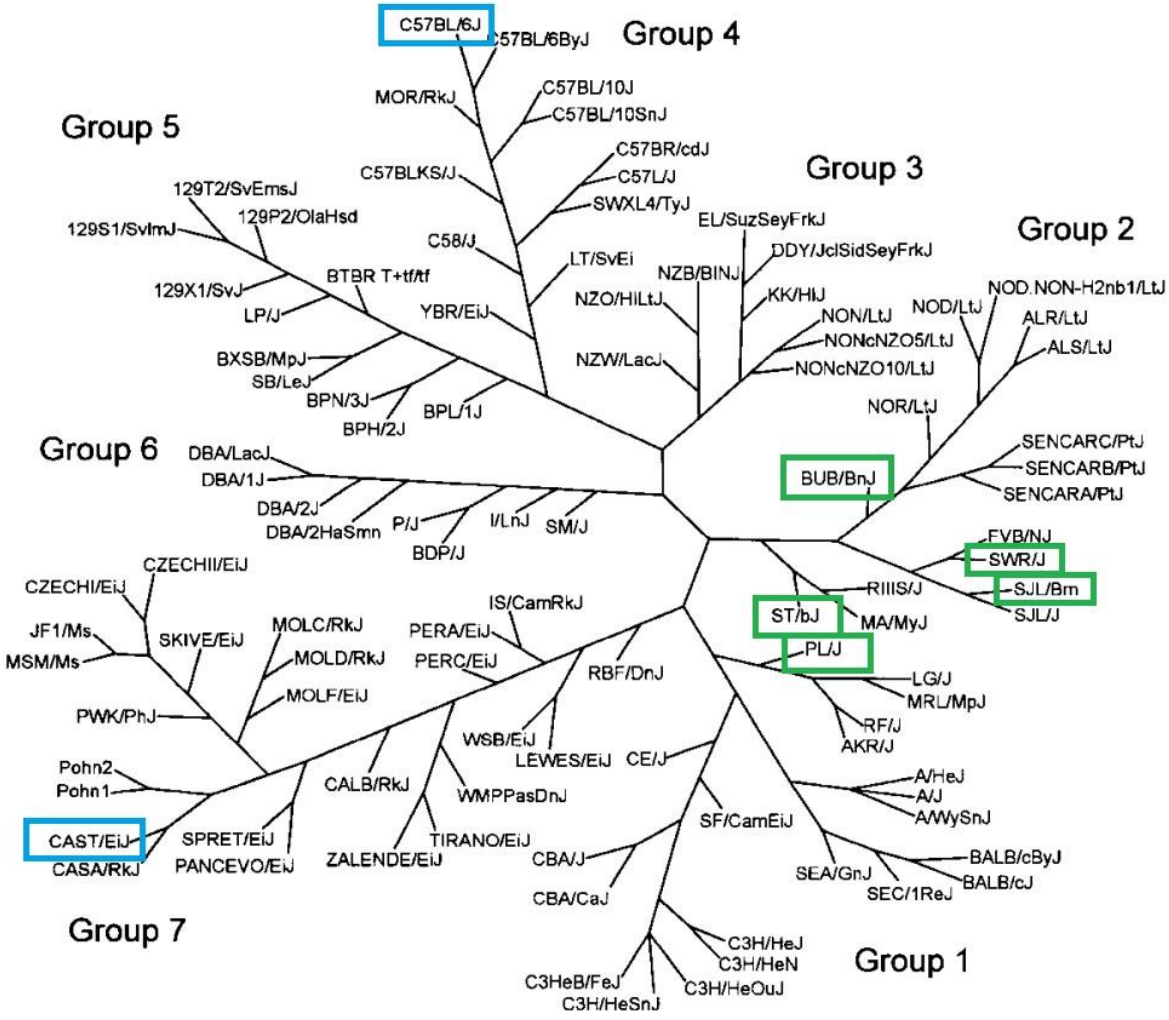


Figure 1.8 Family tree of 102 inbred strains based on SNP analysis

Ancestral relationships between 102 inbred mouse strains based on SNP genotyping. The indicated (boxed) strains were crossed with SWR female mice to generate heterozygous F1 generation female offspring that were treated with DHEA at puberty. Strains highlighted with a blue box (CAST/EiJ and C57BL/6J) rendered the F1 females resistant to GCT initiation, while the other strains highlighted in green were permissive for GCT initiation. Adapted and reprinted with Permission from Cold Spring Harbor Laboratory Press: Appendix D (b).

Blue: GC tumor suppressive strain. Green: GC tumor permissive strain.

Given the prominent role of *Gct1*^{SW} to drive the JGCT phenotype, the *Gct1* locus was selected for independent locus mapping by a congenic strain development strategy, beginning with the transfer of a large genomic segment from the GCT-resistant CAST (allelic designation "*CA"*) (Figure 1.8) strain through 10 backcross generations onto the SWR strain. The result of congenic mapping was complete suppression of GCT susceptibility when homozygous *Gct1*^{CA} alleles displaced homozygous *Gct1*^{SW} alleles, even in the presence of DHEA (Smith *et al.,* 2013)**.** The generation of subcongenic strains carrying smaller, overlapping segments of the donor CAST segment resolved the *Gct1* interval to 1.31 Mbp on mouse Chr 4 (Smith *et al.*, 2013)**.** A focussed investigation for *Gct1* gene candidates is underway by a doctoral trainee in the Dorward lab, with the potential for national collaborative translation of *Gct1* gene candidates identified in the mouse model to human JGCT cases.

1.2.2.3 Gct4: A modifier of GCT susceptibility

Females of the SWR inbred strain are homozygous for the *Gct1*^{SW} driver mutation, but the penetrance of spontaneous GC tumor susceptibility without exogenous hormonal intervention is surprisingly low, historically recorded as ≤ 1 % (Beamer *et al.*, 1985). Early-onset GCTs were first observed in female offspring (F1 generation) of a breeding cross between SWR females and SJL (allelic designation "*J*") inbred strain males for a different experimental purpose (personal communication with A. Dorward). Reciprocal crosses between SWR and SJL strains subsequently led to the identification of a modifier locus for spontaneous GCT susceptibility named *Gct4* on Chr X, based on the GCT stimulation activity of the *Gct4*^{J} allele (Beamer *et al.*, 1998). The phenotypic penetrance of spontaneous GCT incidence is increased from 1 % to about 20 % in SWR.SJL-X congenic females that carry homozygous *Gct1*^{SW} alleles along with *Gct4*^{J} alleles on Chr X. Given the significant influence of the *Gct4* locus to spontaneous GCT

initiation, locus mapping was undertaken for *Gct4* with an inbred strain (CAST) that is more distantly related to SWR than the SJL strain, offering greater mapping resolution (Figure 1.8). Following SWR.CAST-X congenic strain development, it was determined that the $Gct4^{CA}$ allele had a similar influence to $Gct4^{J}$, increasing spontaneous GCT initiation in the absence of androgenic support, whereas increased concentrations of circulating androgen (T or DHEA) was necessary to render GCT penetrance in SWR females equal to the SWR.JL-X or SWR.CAST-X females. Thus, $Gct4^{SW}$ was considered a low penetrance allele (1 %) whereas $Gct4^{CA}$ and $Gct4^{J}$ support moderate penetrance (20 %) (Dorward *et al.*, 2013). The genotype-phenotype association for the modifier locus mapped to the same region of Chr X in SWR.SJL-X and SWR.CAST-X congenic lines, inclusive of the Androgen receptor (*Ar*) gene. Overall, the genetic and functional evidence was strong for the role of androgen signaling during GCT initiation.

An unexpected but consistent observation was made during the mapping crosses and congenic line development, that paternal inheritance of $Gct4^{J}$ or $Gct4^{CA}$ alleles had the same effect to support increased GCT incidence as the homozygous condition, but this was not true for maternal inheritance (Dorward *et al.*, 2013). Although the epigenetic mechanism for the parent of origin bias is not yet explained, the fact that the paternal Chr X had a dominant influence on trait penetrance greatly accelerated the genetic mapping of X-linked *Gct4* using a recombinant Chr X, paternal progeny test. Males carrying an SWR autosomal complement but with unique combinations of SWR and CAST segments on Chr X were mated to SWR dams and cohorts of their female offspring examined for “low penetrance” or “moderate penetrance” rates of spontaneous GCT development. As a result of this investigation, there are three complementary lines of evidence that the *Androgen receptor* (*Ar*) gene shares identity with *Gct4*: 1) the 26 Mbp mapping resolution of the male progeny test placed the *Gct4* locus around the *Ar* gene on central

Chr X (Dorward *et al.*, 2003), 2) in terms of endocrinology, SWR females have increased GCT incidence in the presence of Testosterone (T) ligand for the *Ar* gene, which is distinct from the interaction of DHEA with the Chr 4 autosomal locus, and 3) although there are no mutations in the eight protein coding exons of the *Ar* gene that correlate with "low penetrance" vs "moderate penetrance" outcomes, quantitative gene expression analysis revealed significantly reduced expression of the *Ar* gene in SWR female ovaries relative to SWR-SJL-X congenic female ovaries that exhibit higher spontaneous GCT incidence, even without exogenous testosterone exposure (Dorward *et al*., 2003; Dorward *et al*., 2013). The weight of evidence for the *Ar* gene as a candidate modifier of GCT susceptibility in the SWR model, combined with knowledge that the *Ar* protein plays a regulatory role in ovarian follicle maturation in the reproductive lifespan of the mouse (Sen *et al.*, 2014) was sufficient to support the investigation of a conditional knock-out strategy for the *Ar* gene in the pre-pubertal GCs of SWR females. The goal of this ongoing research is to determine the contribution of Ar-mediated signalling to GCT susceptibility in the SWR model, and to identify the target of DHEA action at the *Gct1* locus in the absence of *Ar*.

1.2.2.4 Parent-of-Origin Effects

In the current era of genetics, it is understood that the mammalian genome is not expressed equally between the sexes, with mechanisms in place that create mild or extreme examples of parent-of-origin expression bias. Genomic imprinting is an example of extreme expression bias, whereby non-coding RNA and chromatin modifications ensure certain genes are expressed in a strict parent-of-origin-specific manner. These epigenetic processes set the imprints at the earliest stages of embryonic development and are maintained through mitotic cell divisions in all somatic cells of an adult organism (Merzouk *et al.*, 2014). Appropriate imprinting of certain genes is important for normal development, with severe consequences when genomic

imprints are defective, causing broad-spectrum abnormalities such as Angelman syndrome and Prader–Willi syndrome (Cassidy *et al.,* 201). An entire copy of Chr X is subject to epigenetic silencing in mammalian females in a process termed X-inactivation, to achieve dosage compensation with Chr XY males. Which Chr X to silence is generally thought to be a random choice in the soma of the embryo, although X-inactivation itself can be imprinted (as for the extraembryonic tissues) or biased towards one parental Chr X copy (Merzouk, *et al.,* 2014). As yet, it is uncertain what epigenetic mechanism drives the paternal effect for X-linked alleles that increase GC tumor susceptibility in the SWR model, although biased X-inactivation or allele-specific expression bias for candidate genes in the GC populations of the ovarian follicles are reasonable hypotheses to be explored.

1.2.2.5 Gct6: A suppressor of GCT initiation

The strong paternal effect of the Chr X-linked *Gct4^{J} and Gct4^{CA}* alleles to increase the spontaneous incidence of JGCTs in SWR-derived congenic female mice created a new research opportunity with the incidental finding of another Chr X allele that could functionally suppress GCT initation. Identification of a *Gct* suppressor locus arose by chance following the creation of the SWR.CAST-X congenic strain, which was originally developed to fine map the *Gct4* locus in the central region of mouse Chr X. Since the creation of congenic strains is labor intensive, costly and takes a relatively long time to complete, mouse geneticists are ultra-conservative at the outset, transferring large genomic segments from the donor strains to establish congenic founders' lines that don't "miss" the genetic locus of interest. In the case of the founder SWR.CAST-X congenic line, the CAST donor segment included 73.511 Mbp of Chr X, from *DXMit109* at 69.675 Mbp, to *DXMit35* at 143.186 Mbp, well beyond 98.1 Mbp, the physical location of the *Ar* gene (Dorward *et al*., 2003). The SWR.CAST-X congenic line underwent 10

generations of backcrossing to the SWR host strain, achieving 99 % return of SWR genome, with the exception of the CAST donor segment (Lee M. Silver, 2001). The goal for the congenic strain was to investigate the location and activity of the X-linked *Gct*4 locus independently from other *Gct* loci; therefore, the SWR.CAST-X line was confirmed to have the GCT driver $Gct1^{SW}$ alleles reconstituted. Despite the presence of homozygous $Gct1^{SW}$, SWR.CAST-X females were found to be completely resistant to GCT initiation (Dorward *et al.*, 2003; Dorward *et al.*, 2013), even in the presence of androgenic (DHEA or T) steroid stimulus (Table 1.2). A new locus named *Gct6* was hypothesized to exist within the CAST donor segment that serves a tumor suppressor role in the androgen-sensitive JGCT mouse model.

Once the GCT suppressor phenotype was confirmed in homozygous SWR.CAST-X females, a reciprocal mating test was run with SWR inbred strain partners, to determine if the paternal parent-of-origin effect applied to the GCT suppressor activity of $Gct6^{CA}$ locus. The test was positive for dominant paternal activity of $Gct6^{CA}$; therefore, the *Gct6* locus was mapped using the same paternal Chr X progeny test strategy as for the *Gct4* locus (Table 1.2). Males from the SWR.CAST-X congenic strain that carried the $Gct4^{CA}$ locus to support 20 % spontaneous GCT incidence, but with different combinations of permissive SWR genome and the suppressive CAST genome on distal Chr X near the *Gct6* locus were mated to SWR dams. Cohorts of their F1 daughter offspring were subsequently examined for spontaneous GCT incidence at 8 wks of age. This proved to be an effective strategy, since it was relatively easy to measure phenotype differences in the daughter offspring when comparing GCT resistant females (0 % incidence) with cohorts that exhibited 20 % GCT incidence. The mapping strategy resolved the *Gct6* locus to a 1.019 Mbp interval on distal Chr X (Dr. Dorward lab, 2013), between the markers *DXamd11* at 136.762 Mbp and *DXamd17* at

Table 1.2 GC tumor incidence with DHEA treatment in female progeny

GCT phenotype			
Permissive		Suppressive	
Strain Cross (Maternal x Paternal)	DHEA-induced GCT incidence % (n)*	Strain Cross (Maternal x Paternal)	DHEA-induced GCT incidence % (n)◊
SWR x SWR (inbred)	24.3 (107)	SWR x C57 (F1)	0.0 (155)◊
SWR x SJL/Bm (F1)	14.0 (107)	SWR x SWR.CAST-X (F1 congenic, *Gct6*^{CA})	0.0 (58)◊
SWR x BUB/BnJ (F1)	13.7 (95)		
SWR x PL/J (F1)	10.0 (80)	SWR x SWR.C57-X (F1 congenic, *Gct6*^{C57})	Unknown
SWR x ST/bJ (F1)	21.9 (32)		

* Dietary DHEA

◊ Capsule DHEA

Strains are grouped according to GC tumor phenotype; permissive or suppressive for DHEA-induced GC tumorigenesis. Mice were necropsied at 8 weeks of age and examined for GC tumors. This project will investigate whether SWR x SWR.C57-X (F1 congenic) line belongs under the suppressive category (Adapted from Petkov *et al.,* 2004, Beamer *et al.,* 1988 and Tennent *et al.,* 1993).

137.053 Mbp (Figure 1.9). The mapped interval includes 20 annotated candidate genes along with other genetic determinants that were candidates for *Gct6* identity (Figure 1.10 and Table 1.3). *Gct6*^{CA} action is fundamentally different from *Gct6*^{SW} *or Gct6*^{J} alleles contributed by the SWR or SJL strains, respectively, since it blocks the action of both DHEA and testosterone in female mice that carry confirmed, autosomal susceptibility alleles for GCT initiation. This suggests that the *Gct6*^{CA} allele interferes with the oncogenic mechanism triggered by the DHEA-responsive *Gct1*^{SW} locus on Chr 4 (Dorward *et al*., 2013), and/or suppresses androgenic signaling through *Gct4* (*Ar).* To verify the mechanism of tumor suppressor action in the GC population, research efforts are now focussed on the identity of *Gct6*.

The *Gct6* allele from the CAST strain gave the first indication that an independent X-linked locus had GCT suppressor activity, implying biologically-relevant allelic differences exist between the *Gct6* locus of CAST and the other strains involved in the search for GCT susceptibility genes. Since the activity of all the *Gct* loci are heritable across mouse generations, *Gct1*, *Gct4* and *Gct6* variation is expected in the germline DNA of the strains of interest. For *Gct6*, DNA sequence comparisons between the SWR and CAST genome would be a starting point for the identification of unique *Gct6*^{CA} alleles; however, the CAST strain is highly divergent from SWR and most common lab strains of mice, as illustrated by the ancestral relationship of inbred strains (Figure 1.8) (Petkov *et al.,* 2004). Rather than systematic candidate gene investigations for unique *Gct6*^{CA} alleles, it is feasible to capture all the genetic variation across the *Gct6* locus at one time, incorporating both protein coding, regulatory and intronic regions. With the refined *Gct6* interval already mapped to 1.019 Mbps, an objective sequencing strategy was implemented that combined whole *Gct6* locus capture of CAST and SWR strain DNA with a Next Generation Sequencing (NGS) approach, recognizing that a very high number

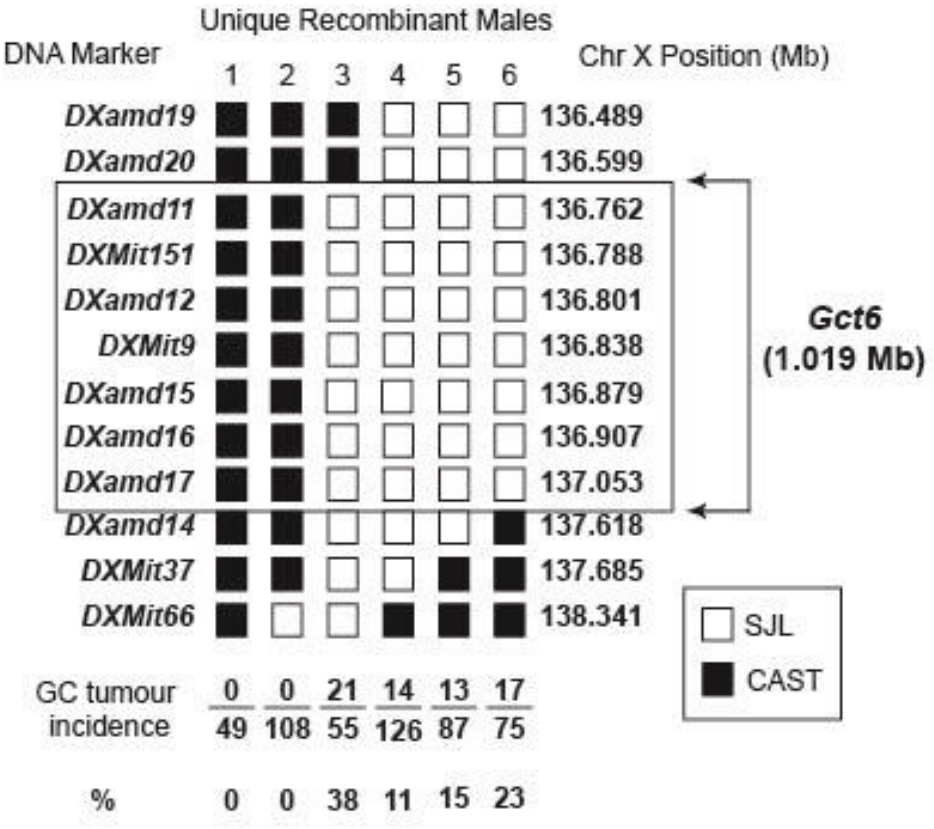


Figure 1.9 Haplotype map of the *Gct6* interval on distal mouse Chr X

The CAST strain donates a GC tumor resistant allele at *Gct6*, which greatly contrasts with $Gct6^{SW}$ and $Gct6^{J}$ alleles that are permissive for GC tumorigenesis (~20 %) when tested on an SWR autosomal inbred background in the presence of a strong $Gct4^{CA}$ allele. Based on the allelic difference between $Gct6^{J}$ and $Gct6^{CA}$, and the known paternal parent-of-origin effect contributed by the paternal Chr X, males carrying unique recombination of SJL (white boxes) and CAST genomic segments (black boxes) were derived from the SWR.CAST-X congenic strain, mated to SWR females, and the spontaneous GC tumor incidence of F1 female offspring was recorded. This map represents the genotype of only six of the most informative recombinant males tested in the region of *Gct6*, placing *Gct6* between two custom SSLP DNA markers (*DXamd20* and *DXamd14*) that are 1.019 Mbp apart on distal Chr X. The map does not show the *Gct4* locus on Chr X, which was fixed as a strong, permissive allele ($Gct4^{CA}$) for all tested males. Ratios at the base of each unique male genotype represent the total number of female mice with spontaneous GC tumors over the total number of females examined at 8 wks of age (Used with permission from Dr. Dorward's Laboratory, 2012).

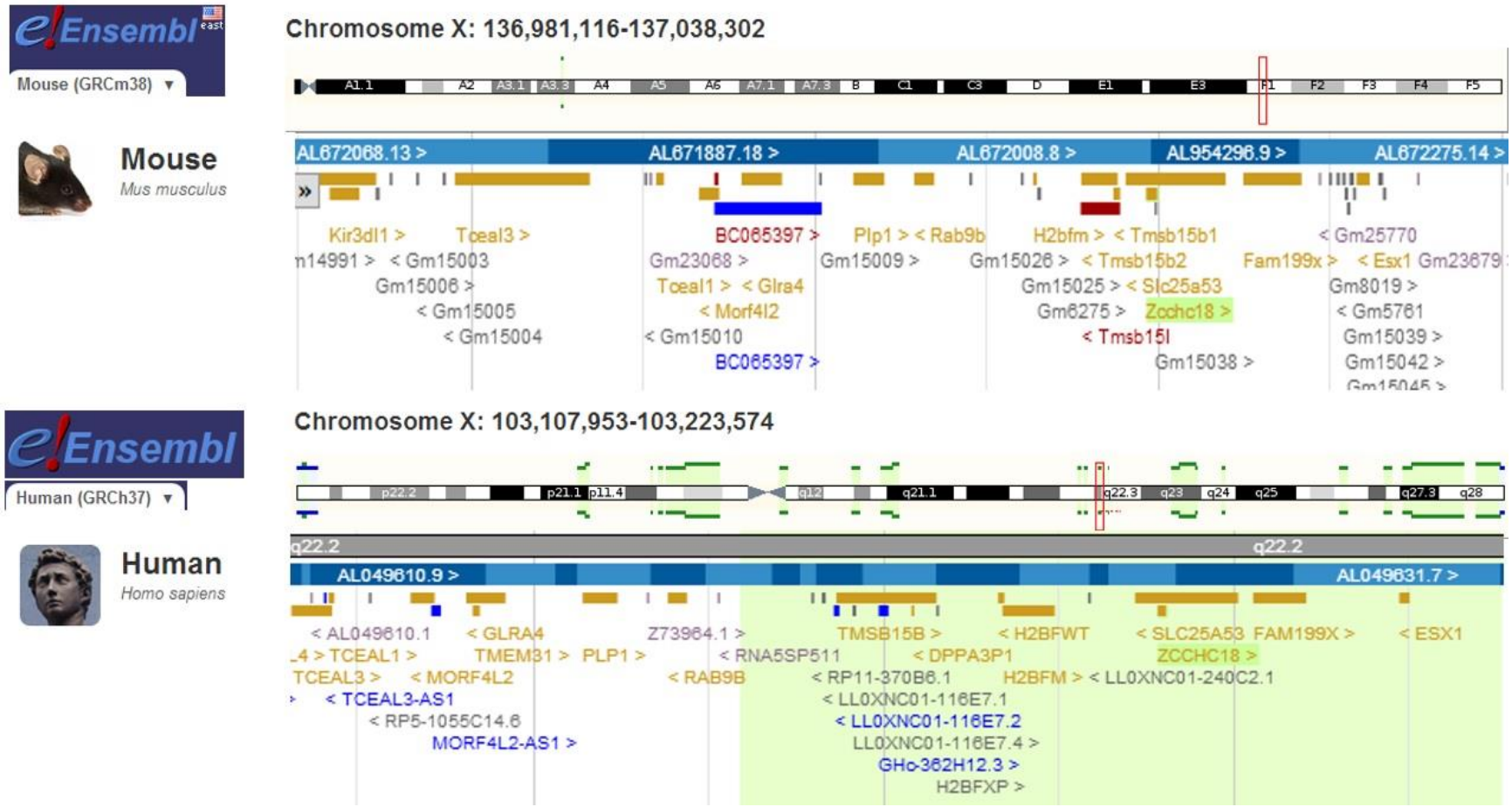


Figure 1.10 *Gct6* locus Ensembl screen shot comparison between mouse and human.

Gct6 locus identified by red box both in mouse and human. Ensembl Genome Browser release 75 (Flicek *et al.*, 2014) was used to capture both screen shots.

Table 1.3 Predicted genes within the *Gct6* interval

Gene Symbol	Position (38 Ensembl)	Ensembl ID (ENSMUSG000000)	Description	Function	Gene Type	Predicted number of Transcripts
Tceal3	X: 136590842-136668378	44550	*Transcription Elongation factor A (SII)-like 3*	Transcriptional regulation	protein-coding	7
7SK	X: 136704178-136704470	88864	*ncRNA*	---	Misc RNA	1
Tceal1	X:136707982-136711478	49536	*Transcription Elongation Factor A (SII)-Like 1*	Transcriptional regulation	protein-coding	1
Morf4l2	X:136732942-136743690	31422	*Mortality Factor 4 Like 2*	Transcriptional regulation	protein-coding	16
BC065397	X:136741858-136743143	72960	*cDNA BC065397*	---	protein-coding	1
BC065397	X:136741858-136743143	87368	*cDNA BC065397*	---	processed transcript	7
Glra4	X:136757674-136780141	18595	*Glycine Receptor Subunit Alpha-4*	Inhibition of neuronal firing	Protein-coding	4
Plp1	X:136822671-136839733	31425	*Proteo-lipid protein (myelin) 1*	Formation or maintenance myelin	Protein-coding	5
Rab9b	X:136858147-136868755	43463	*Rab9b member RAS oncogene family*	Protein transportation between the endosomes and the trans Golgi network	Protein-coding	1
H2bfm	X:136927325-136928373	48155	*H2B histone family, member M*	Core component of nucleosome	Protein-coding	1

Gene Symbol	Position (38 Ensembl)	Ensembl ID (ENSMUSG000000)	Description	Function	Gene Type	Predicted number of Transcripts
Tmsb15l	X:136954988-136976869	72955	*Thymosin β 15b like*	---	protein-coding	1
Tmsb15b2	X:136955265-136975497	89996	*Thymosin β 15B2*	Organization of the cytoskeleton	Protein-coding	2
Tmsb15b1	X:136974022-136976874	89768	*Thymosin β 15b1*	---	protein-coding	1
Slc25a53	X: 136981116-137038302	44348	*Mitochondrial Carrier Triple Repeat 6*	Trans-membrane transport	protein-coding	14
Zcchc18	X:136993155-136998472	31428	*Zinc finger CCHC domain containing 18*	---	protein-coding	3
Fam199x	X:137049594-137082503	42595	*Family with sequence similarity 199 X-linked*	---	protein-coding	1
SNORA17	X:137093612-137093726	89191	*snoRNA - small nucleolar RNA, H/ACA box 17*	---	ncRNA	---
Esx1	X:137115397-137122083	44348	*Extraembryonic Spermatogenesis homeobox 1*	Placental development and spermatogenesis.	protein-coding	3
SNORA63	X:137150815-137150933	65095	*snoRNA, Small Nucleolar RNA, H/ACA Box 63*	---	ncRNA	1
Il1rapl2	X:137570608-138846946	59203	*Interleukin1 receptor accessory protein-like 2*	Associated with X-linked non-syndromic mental retardation.	protein-coding	3

Misc RNA: miscellaneous RNA

ncRNA: non-cording RNA

of variants would be predicted for the polymorphic CAST genome vs the C57 reference or SWR genome. It was also appreciated that identification of another strain besides CAST that confers GCT suppressor activity from the distal Chr X region would greatly facilitate the search for *Gct6* identity based on the hypothesis that a common allele would co-segregate with GCT suppressor activity. If the hypotheses is true, it would support an informatics filtering strategy to eliminate DNA sequence variations in the CAST strain that are not relevant to the GCT phenotype.

Beamer *et al*. (1988-b) screened for DHEA-induced GCT incidence in F1 hybrid female offspring of SWR dams mated to males from 9 different inbred strains: AKR/J, MA/MyJ, RIIIS/J, RF/J, SJL/Bm, ST/bJ, BUB/BnJ, PL/J and C57BL/6J (C57). Females from these inbred strains do not share the trait of GCT susceptibility with the SWR strain, but the F1 offspring from eight of the crosses (cohort size 32 to 155 females) exhibited GCT incidences from 1 % to 22 %, following inheritance of only one maternal copy of $Gct1^{SW}$, one copy of Chr X from the paternal strain and complete heterozygosity on the remaining autosomes (Table 1.3). It was fortuitous that the crosses were designed for paternal Chr X inheritance, since the experiment established that $Gct6^{AKR}$, $Gct6^{MA}$, $Gct6^{RIIIS}$, $Gct6^{RF}$, $Gct6^{ST}$ $Gct6^{BUB}$ and $Gct6^{PL}$ were GCT permissive alleles, and confirmed the permissive nature of the $Gct6^{J}$ allele from the SJL strain (Beamer *et al*., 1985). The paternal strain that stood out as a potentially harbouring a GCT suppressor locus was C57, with 0 GCTs recorded of 155 (SWR x C57) F1 females treated with DHEA from puberty until 10 wks of age (Beamer *et al*. 1988, Table 1.3). This data suggested that CAST and C57 could share the GCT suppressor allele at the *Gct6* locus, although it cannot be firmly concluded that X-linked *Gct6* is responsible for GCT suppression in this cross with this experimental design. Since the F1 females examined were heterozygous for SWR and C57 loci across the entire genome, it was possible a different suppressor locus was at work. The definitive

strategy to deconstruct polygenic traits in mouse models is to isolate the locus of interest in a congenic strain and test genotype/phenotype associations. We therefore proposed to construct a SWR.C57-X congenic line that carries the *Gct6*^{C57} locus based on positional information learned from the SWR.CAST-X mapping panel. With the *Gct6*^{C57} locus isolated, SWR.C57-X females would be investigated for their GCT susceptibility upon treatment with DHEA. If the SWR.C57-X strain females are GCT-resistant, there would be a strong rationale to search for common genetic determinants and perhaps common alleles between CAST and C57 genomes across the *Gct6* locus.

1.3 NGS Technology and Prior Collaborative Work

Genetic analysis has been transformed in the past decade with technological advances to sequence whole genomes consecutively with high-throughput, NGS platforms. NGS is a nucleic acid sequencing technology that simultaneously determines the sequence of many thousands or millions of short templates in a single sequencing run. Although the chemistry of current NGS platforms differ, the mass acquisition of sequencing information makes use of the reference genome scaffolds for many species to realign short sequencing reads for the detection of rare variant sequences in whole exomes, genomes or transcriptomes, in a qualitative and a statistically quantitative manner (Brown, 2013). To apply NGS methodology in the search for *Gct6* identity, a research collaboration was established with The Jackson Laboratory (TJL; Dr. Laura Reinholdt) to avail of the Illumina platform sequencing core facility at TJL and their informatics pipeline for mutation discovery in the mouse.

Since *Gct6* was already mapped to high resolution with phenotype-driven mapping protocols, a targeted sequencing approach was deemed the most efficient step to isolate the entire *Gct6*

locus of the SWR inbred strain and SWR.CAST-X congenic line DNA, prior to NGS sequencing (Simon *et al.*, 2012). *Gct6* locus enrichment was accomplished using a SureSelect DNA capture microarray (Agilent Technologies Inc.), customized to capture and enrich the SWR inbred strain and SWR.CAST-X congenic strain DNA with a generous interval (2.1 Mbp) across the mapped 1.019 Mbp *Gct6* locus. The array was designed with a 3 bp offset and 60 bp probe length, masking simple repeat regions, but representing all possible coding, regulatory and intronic regions within the locus. Enriched, captured DNA for the *Gct6* interval was subsequently fragmented into 100 bp lengths followed by sequencing on an Illumina HiSeq2500 using 2 x 100 paired end sequencing protocol. The fragment reads were aligned to the C57 reference genome (Archive Ensembl release 67 - May 2012 © WTSI / EBI). Primary, secondary and tertiary NGS analysis was performed at TJL according to their standard protocols (Simon *et al.*, 2012). Individual variant reports comparing SWR to C57 reference genome and CAST to C57 reference genome were provided to our lab group for follow-up analysis of variants in *Gct6* candidate alleles.

2.0 Hypotheses and Research Objectives

Hypotheses:

The C57 strain harbours a GCT suppressor allele within the mapped *Gct6* locus interval on Chr X.

Genetic variants in common between the CAST and the C57 strain across the *Gct6* locus are candidate tumor suppressor alleles that abrogate androgen-sensitive GCT initiation in SWR mice.

Research Objectives:

1) To develop a six-generation SWR.C57-X congenic strain that transfers C57 genome to the mapped *Gct6* region on mouse Chr X, and to test the GCT susceptibility of female offspring that paternally inherit the $Gct6^{57}$ locus in the presence of DHEA.

2) To perform informatics analysis and sequence confirmation for genetic variants identified by a whole-*Gct6* locus capture and NGS technique for the $Gct6^{SW}$ GCT-permissive locus vs the $Gct6^{CA}$ –suppressor locus, relative to the C57 strain reference genome.

3.0 Materials and Methods

3.1 Animal Housing

Mice were housed at the central animal care facility in the Health Sciences Centre at Memorial University of Newfoundland. All animal care protocols were approved by the Institutional Animal Care Committee, in accordance with Canadian Council on Animal Care guidelines. Mice were given Laboratory (autoclavable) Rodent Diet 5010 food (27.5 % protein, 13.5 % fat, 59 % carbohydrate; OM Nutrition International, Richmond, IN) with sterile water ad libitum, and were housed under a 12:12 hour (h) light/dark cycle. Mice were weaned at 20-23 day of age and housed in groups of 2 to 5 animals per 27.94 cm L x 17.78 cm W x 12.7 cm H rodent cages containing Bed-O-Cobs® corn-cob bedding material (The Andersons, Maumee, OH).

3.2 Development of the SWR.C57-X Congenic Strains

An outcross mating between SWR and C57 inbred strains was performed to create N1 generation progeny that are fully heterozygous at every autosomal and Chr X locus (N1F1). The N1F1 progeny were selectively bred for 5 additional backcross generations (N) to SWR inbred strain mates, followed by an intercross generation (N6F2) to generate 2 homozygous SWR.C57-X congenic lines (N6F3). SWR.C57-X Line 1 harboured a 65 Mbp continuous segment of C57 strain genome that introduced both $Gct6^{C57}$ and $Gct4^{C57}$ loci, delimited by simple sequence length polymorphic (SSLP) markers informative to distinguish SWR vs C57 background (*DXMit119*, *DXMit96*, *DXMit117*, *DXamd22* and *DXamd14* (Table 3.1). SWR.C57-X Line 2

carried a smaller 38 Mbp segment of C57 genome that introduced only the *Gct6*^{C57} locus, delimited by *DXMit117, DXamd22* and *DXamd14* SSLP DNA markers (Table 3.1). At each backcross generation, only those offspring who inherited both *Gct4*^{C57} plus *Gct6*^{C57} (Line 1) or just *Gct6*^{C57} (Line 2) alleles were selected for the next round of backcrossing. At the N6 generation, approximately 95 % SWR genome was reconstituted while 5 % was from the C57 donor strain, including the desired region on Chr X that contains the *Gct4* and *Gct6* loci (Silver, 1995; Figure 3.1). To evaluate the phenotype of *Gct6*^{C57} independently from the driver mutation *Gct1*^{SW}, all the selected N6 offspring chosen as founder mating pairs for the SWR.C57-X congenic strain were confirmed to be homozygous for the *Gct1*^{SW} allele using the *D4Mit232* SSLP marker genotyping by Simple SSLP.

3.2.1 Tail Tip Collection and DNA Extraction

Approximately 1-2 mm of tissue was removed from the tip of the tail using scissors upon necropsy or during weaning. Tail tips were placed in 1.5 mL microcentrifuge tubes with 500 µL of 50 mM sodium hydroxide (BDH Inc., Toronto, ON), and heated at 95 ºC for 10 min. After the addition of 50 µL of 1 M Tris-HCL pH 8.0 (Sigma-Aldrich Inc., St. Louis, MO; Fisher Scientific, Fair Lawn, NJ), the samples were centrifuged at 10,000 x *g* for 10 min, and the supernatant transferred to new 1.5 mL microcentrifuge tubes. DNA samples were stored at -20 ºC prior to genotyping protocols.

3.2.2 Polymerase Chain Reaction (PCR)

Genomic DNA extracted from mice tails was amplified using a *MasterTaq* PCR Kit (5 PRIME Inc., Gaithersburg, MD). The following reagents were combined in a 0.2 mL PCR tube (Bio-Rad Laboratories, Hercules CA): 4.35 µL of ultrapure distilled water (Invitrogen, Carlsbad CA), 2 µL of 5X *TaqMaster* PCR Enhancer pre-heated to 65 ºC, 1 µL of 10X Reaction Buffer, 0.2 µL of 10 mM deoxynucleotide triphosphates-dNTPs (Invitrogen, Carlsbad, CA), 0.05 µL of *Taq* DNA polymerase (5 Units/µL), 0.2 µL of 10 µM forward and reverse primers (see Table 3.1; Integrated DNA Technologies, Coralville, IA) and 2.0 µL of DNA template. Ten microliter reactions were amplified in a Veriti™ 96-well thermal cycler (Applied Biosystems) using the following PCR protocol: 97 ºC for 30 s; 39 cycles of 94 ºC for 15 s, 55 ºC for 30 second, 72 ºC for 30 s; 72 ºC for 10 min; 4 ºC hold.

Table 3.1 PCR primer sequences for SSLP DNA markers at *Gct1, Gct4* and *Gct6* loci.

Region	Marker Symbol	Forward primer sequence (3' > 5')	Reverse primer sequence (5' >3')	Chr: Positions *
Gct1	*D4Mit232*	GCGTCACCACACTGCTCTT	ACTCAGAGTCCCCTGGCC	4: 145057739-145058007
Gct4	*DXMit119*	CTT TAA CCA TAA TAA TGG CCT TGC	GGG TTC TGT GAT CGC AAG TT	X: 72410246-724103
Gct4	*DXMit96*	CATGTCAATTGGGATCTTTGG	AGGAGCAAATCCAACCTGG	X: 99231474-99231566
Gct6	*DXMit117*	GTAAAACCGCACTGACTAAGCC	TTCGCCTCTATCTTACTCTGGC	X: 131288026-131288151
Gct6	*DXamd22*	CCT TCC CTC TCA CTG TGT CC	AAT GCC ATG GTT TTG CTC AT	X: 137088324-137088343
Gct6	*DXamd14*	CCA TGT GTG ACG ACT TGA CC	CCC ACT CCA AAT TAG TAC CAA CA	X: 137618303-137618322

DXMit primers were obtained through MGI website (Blake et al., 2014).

DXamd primers (Custom SSLP markers) were created by Dr. Ann Dorward at Memorial University of Newfoundland.

*data were obtained through Ensembl Genome Browser release 75 (Flicek *et al.*, 2014).

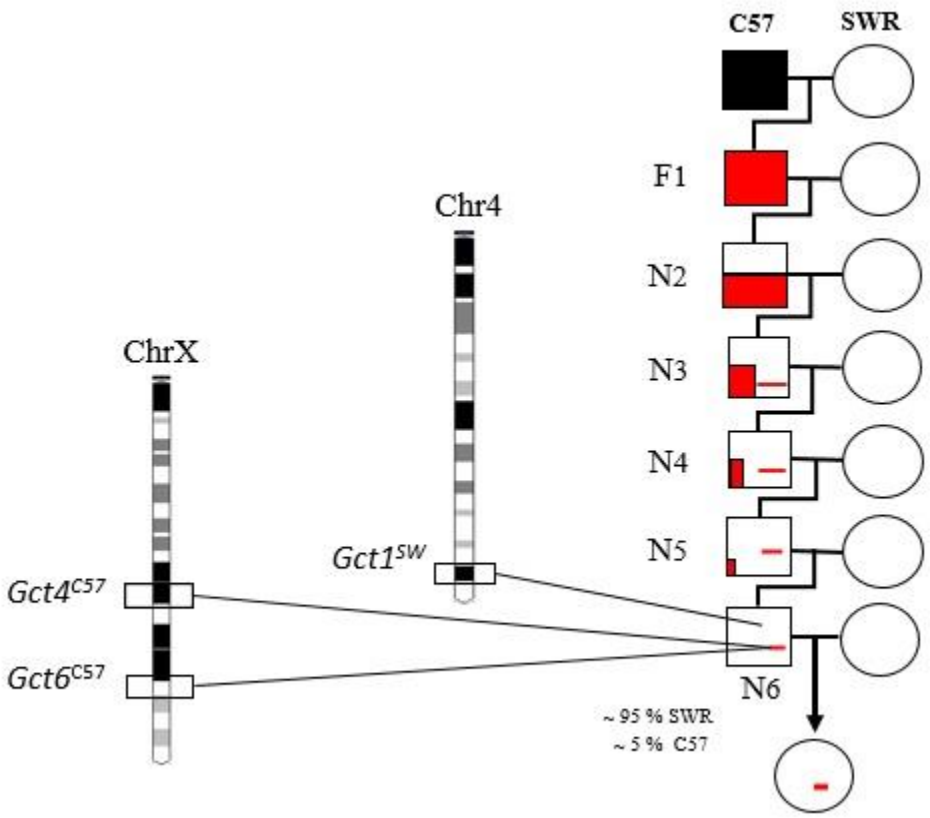


Figure 3.1 Isolating *Gct1*^{SWR}, *Gct4*^{C57} and *Gct6*^{C57} in a congenic strain

A schematic representation of the relative C57 and SWR contributions at sequential generations of backcrossing. *Gct4*^{SW} and *Gct6*^{SW} are swapped with *Gct4*^{C57} and *Gct6*^{C57}. By the sixth generation of backcrossing, the differential segment around the selected loci (*Gct4* and *Gct6*) represent the major contribution from C57 genome.

Black: C57; White: SWR; Red: Heterozygous.

(Adapted from Silver, 1995)

3.3.1 Gel Electrophoresis

Amplified PCR products were separated by horizontal electrophoresis through a 4 % high-resolution agarose gel. Bromophenol blue loading dye of 0.25 % (W/V) Bromophenol blue and 40.0 % (W/V) sucrose, was added to each PCR reaction of 10 µL and mixed to a final concentration of 1X. PCR products were resolved on 4 % agarose gels, composed of 8 gr of Metaphor agarose powder (Lonza, Mississauga ON) mixed with 20 mL of 10X TBE buffer (890 mM Tris, 20 mM EDTA (pH 8) and 890 mM Boric acid) and 172 mL H_2O. The gel mixture was boiled using a microwave for 4 min, and poured into a gel tray (HE 99X Max Horiz Sub unit Hoefer, Inc, San Francisco, CA). The PCR reactions and 500 ng of 100 base pair 1 Kb DNA ladder marker (Bio Basic Inc., Markham, ON) were loaded into individual wells. The horizontal gels were run in 1 X TBE buffer for 3.5 h at 110 V, after which they were stained for 30 min in a bath containing 50 µL of ethidium bromide solution-10 mg/ml (Sigma-Aldrich Inc., St. Louis, MO) diluted in 1 L of water. The SSLP PCR products were visualized using a U: Genius GelVue UV transilluminator (302 nm; Syngene, Frederick, MD) and scored for allele size by comparison to SWR, C57 and SWR.C57-X strain control DNA amplified under the same PCR conditions.

3.4 *Gct6*^{C57} Allele Phenotyping

Ten congenic SWR.C57-X males (5 from Line 1 and 5 from Line 2) were mated once more to SWR females in a paternal progeny test for the GC tumor suppressor activity of *Gct6*^{C57}. All female offspring born (N7F1) were implanted with a subcutaneous DHEA-filled capsule at weaning (age 20 – 24 d) to stimulate androgen-induced GC tumorigenesis. Ultra-pure DHEA (Steraloids Inc., Newport, RI, USA) was packed into 1.0 cm capsules made from Silastic tubing (1.98-mm inner diameter 93.18-mm outer diameter; Dow Corning, Midland, MI, USA) capped

with glass beads (Beamer *et al.*, 1988). Capsules were implanted subcutaneously on the back of mice under isoflurane anesthesia (Baxter Corporation, Mississauga, ON, Canada) with postoperative carprofen (5 mg/kg body weight) analgesic (Pfizer Canada, Kirkland, QC, Canada) and stainless steel wound closure clips.

All heterozygous females were euthanized by CO_2 gas at 8 wks and examined macroscopically for GC tumors with a visual inspection of both ovaries (Figure 3.2). At this age, GC tumors usually present as cystic or solid hemorrhagic masses of 5–10 mm in diameter and may be bilateral. If a unilateral or bilateral GC tumor was detected, the offspring was considered as one affected female for the purpose of tumor incidence calculations. A sample size (n) of minimum 50 individual females born from each congenic paternal cohort was determined based on statistical power calculations using an estimated DHEA-induced GC tumor incidence of 20 % in inbred SWR females recorded both historically and following transfer to the HSC from another institute, relative to a tumor-resistant phenotype (0 % incidence) in SWR.CAST-X congenic lines (Beamer *et al.* 1988 and Smith, 2012).

3.4.1 Reciprocal Test of SWR.C57-X

A reciprocal test was done by mating SWR.C57-X *Gct6*^{C57} heterozygote females with SWR males and treating their offspring with a DHEA capsule to prove that paternal inheritance of *Gct6* is enough to induce the tumor suppressive phenotype; the incidence of GC tumors were recorded.

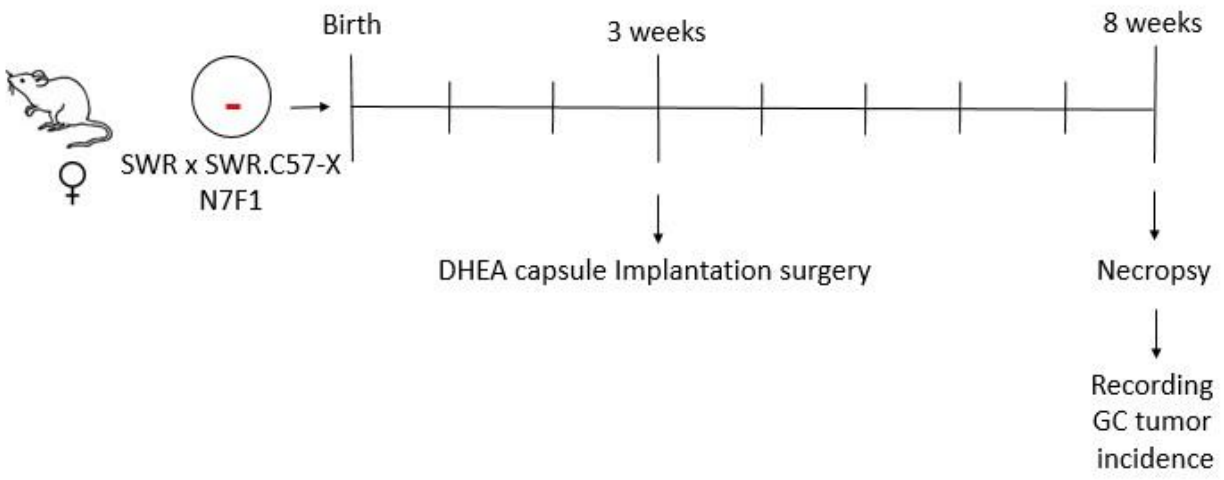


Figure 3.2 Sample collection paradigm for GC tumor incidence of congenic SWR.C57-X daughters implanted with DHEA capsules.

Congenic daughters SWR.C57-X were implanted with DHEA capsules at 3 weeks of age. They were necropsied at 8 wks and checked for the presence of GC tumors.

3.4.2 Homozygote Test of Gct6^{C57} SWR.C57-X

A homozygote line of *Gct6*^{C57} was created through mating SWR.C57-X *Gct6*^{C57} females with SWR.C57-X *Gct6*^{C57} males and treating their offspring with a DHEA capsule. The GC tumor incidence was examined and recorded in the same manner as the SWR.C57-X congenic line. This experiment was an additional layer of proof to indicate the tumor suppressive phenotype of the paternal inheritance of *Gct6*.

3.5 Statistical Analysis

Sample size was initially calculated based on a power calculation, with a predetermined tumor incidence rate of 18 % in the SWR strain used in this study that were treated with DHEA. Study power and significance level were selected to be 80 % and 0.05 respectively. For a zero

rate of incidence in the resistive group, the power calculations show a sample size of (at least) 49 is required for the selected power and significance level. Fisher's exact test (Preacher *et al.*, 2001) was then used to determine the p-values in order to confirm significance requirements have been met. The GC tumor incidence of both Line 1 and Line 2 N7F1 female cohorts was individually compared to the SWR inbred line incidence using the Chi-squared analysis for proportions, with a chosen significance level of $P < 0.05$. Statistical analyses was performed with GraphPad Prism ver. 5.00 Software (San Diego, CA, USA).

3.6 Histology

Representative ovaries with unique genotypes and phenotypes were taken at necropsy (approximately 8 wks of age) for histological analysis. Two ovaries with GC tumors, one from a SWR.C57-X *Gct6*^{C57} congenic mouse and one from an SWR mouse, were prepared for histological analysis. Additionally, normal ovaries were taken from 2 individual SWR.C57-X *Gct6*^{C57} congenic mice. The ovaries were carefully dissected under the microscope and immediately placed in 4 % paraformaldehyde in phosphate buffer (PFA) pH 7.4 for 5-6 h. After fixation in PFA, the ovaries were stored in 70 % ethanol overnight. The samples were then sent to Histology department of the Faculty of Medicine for paraffin embedding, sectioning, slide mounting and hematoxylin and eosin staining. The histology slides were observed under Leica DM-IRE2 inverted microscope (Leica Microsystems, Richmond Hill, ON, Canada) attached to a Retiga Exi CCD camera (QImaging, Burnaby, BC, Canada). Openlab Image Analysis software (Version 5.5; Improvision, Inc., Lexington, MA, USA).

3.6.1 NGS Variants List for Gct6

The list of variants identified for the *Gct6* locus were represented in an excel table format with the following headings: Chromosome, Position, Reference_Base, Consensus_Base, SNP quality, Depth, Allele frequency (non-reference allele), Mapping quality, Exon, Exon ID, Strant, Intron, Intergenic, Function class, Transcript(s), Gene symbol, Strand, Ensembl version (for sample data, see Appendix E). The first filtering step for the variant list was to exclude variants based on their reported allele frequency (AF). From this dataset, an AF of 1.0 indicated a homozygous variant while AF of 0.5 indicated heterozygosity. The DNA samples sent to The Jackson Laboratory were isolated from homozygous congenic or inbred strain animals, so we hypothesized that all variants with AF less than 1.0 indicated a technical error during the sequence assembly steps, and so these variants were not considered further.

The remaining variants with AF of 1.0 were reviewed as to whether they were in common between SWR and CAST genome vs the C57 reference genome (Figure 3.3). The common variants were excluded from further consideration given the rationale that $Gct6^{SW}$ and $Gct6^{CA}$ are hypothesized to be genotypically different, based on the distinct GC tumor permissive vs. suppressive phenotypes conferred by two strain alleles. The remaining variants were prioritized by their location relative to annotated features of the locus, labelled as intronic, intergenic, or exonic, with exonic variants given higher priority if they influenced the protein coding sequence vs the untranslated regions of genes. Finally, exonic coding variants were filtered to prioritize non-synonymous over synonymous variants. The non-synonymous variants were labeled as high priority for sequence validation and consideration for *Gct6* identity (Figure 3.3).

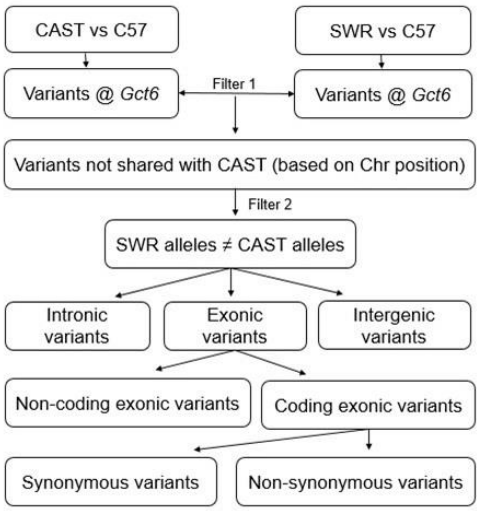


Figure 3.3 NGS filtering steps diagram

The NGS variants list included all the differences between C57 and each tested strain (CAST vs C57 and SWR vs C57). The variants that reside within *Gct6* locus were isolated for both, then the common variants between CAST and SWR according to their bp nucleotide positions were removed from the data set (Filter 1). Only variants with different alleles from CAST were kept (Filter 2) for further analysis. The remaining variants were then further broken down into Intronic, Exonic and Intergenic categories as well as synonymous and non-synonymous. The priority was given to non-synonymous coding exonic variants.

3.7 Sanger sequencing preparation

3.7.1 PCR Amplification

DNA was extracted from tail tips of SWR, SWR.C57-X and SWR.CAST-X mice as previously described (Section 3.3.1). DNA from additional GCT permissive strains (PL/J, BUB/BnJ, SJL/Bm and ST/bJ) was ordered from The Jackson Laboratory. The following reagents were combined in a 0.2 mL PCR tubes: 15.2 µL of distilled water, 7.0 µL of 5X TaqMaster PCR Enhancer preheated to 65 ºC, 3.5 µL of 10X Reaction Buffer, 0.7 µL of 10 mM dNTPs (Invitrogen, Carlsbad, CA), 0.175 µL of Taq DNA polymerase (5 PRIME, Inc. MD, USA) kept on ice, 0.7 µL of 10 µM forward and reverse primers and 7.0 µL of DNA template or ultrapure distilled H_2O (negative control; Invitrogen, Carlsbad CA). The 35 µL PCR reactions were amplified in a Veriti**™** 96-well thermal cycler with primers flanking the nonsynonymous variants in BC065397, Slc25a53, Esx1 and Tmsb15l (Table 3.2) under the following conditions: 5 min at 95 ºC; 6 cycle of 15 s at 95 ºC, 30 s at 68 ºC decreasing 1 ºC per cycle, 68 ºC for 1 min; 35 cycle of 15 s at 95 ºC, 62 ºC for 30 s, 68 ºC for 1 min; 72 ºC for 10 min; 4 ºC hold.

3.7.1.1 Gel electrophoresis

PCR products (10 µL) were run on a 1 % agarose gel, followed by horizontal gel electrophoresis and visualization of the products. The 50 mL gel prepared in 1 x TBE solution contained 0.5 g agarose (Sigma, St. Louis MO), boiled for 2 min to dissolve the agarose, after which 2 µL of 10 mg/mL ethidium bromide was added (Sigma-Aldrich Inc., St. Louis, MO). The gels were cast, loaded, electrophoresed for 30 min at 110 V and PCR products visualized as per the protocol for SSLP genotyping (Section 2.2).

3.7.2 DNA Purification from PCR Products

PCR products were purified using a QIAquick® PCR purification Kit-250 (Qiagen Inc., Mississauga, ON) as per the manufacturer's protocol. Five volumes of binding buffer was added to one volume of PCR product. If the PCR products' color indicator was orange or violet meaning the solution is basic, therefore 10 µL of 3 M sodium acetate (pH 5.0) was added to bring the solution pH close to 7.0, before transfer to a 2 mL QIAquick column and centrifuged for 1 min at 13000 rpm. After centrifugation, the flow through was discarded. Seven hundred and fifty µL of wash buffer was added to the QIAquick columns and the centrifugation step was repeated and flow through was discarded. The centrifuge and wash steps were repeated once more. The filter column was transferred to a new 1.5 mL micro-centrifugation tube, 50 µL of EB (10 mM Tris·Cl, pH 8.5) was added to the center of each column, incubated for 5 min and centrifuged at 13000 rpm for 1 min. The flow through samples containing the eluted DNA samples is preserved so that the purity of the samples can be assessed.

Table 3.2 Sanger sequencing primers used to amplify amplicons from protein coding transcripts within *Gct6*.

Genes Symbol	Forward primer sequence (5' > 3')	Reverse primer sequence (5' > 3')	Products length (bp)
BC065397	GGATGCCTGGATGCTTTGGA AGGCGAGGTCCCTCATTTTG * TGCTAGTATCTAAAGGAAAGGGCT*	GGGCATAAGGGAGATTCGCA	722
Esx1	GATGCAGTCAAATGCCACCC	ACCCCAGGAAAAAGCCACTC	626
Tmsb15l	CAG GAA GCA CTT ACA TAT CCC CA	CGT CTC CAC AGC CAG TTT CT ACGT CTC CAC AGC CAG TTTC ◊	513
Slc25a53	TGAGGGATGCGGGAACTAGA	GCAAACCAGCCTCGGTATCT	315

Primers were designed using NCBI primer blast tool (Geer *et al*., 2010).

* indicates primers used only for Sanger sequencing.

◊ Primer used only to amplify and sequence CAST.

3.7.3 Quality Assessment of Purified DNA

The quality of the purified DNA samples were assessed Thermo Scientifi NanoDrop 1000 Spectrophotometer. The readings were viewed by ND-1000 V3.5.2 software. The ratio of absorbance at 260 nm and 280 nm (260/280) was used to assess the purity of DNA samples. The 260/280 ratio of 1.8-2 was accepted as pure and a lower ratio value may indicate the presence of protein, Phenol or other contaminants that absorb strongly at or near 280 nm. A secondary measure of nucleic acid purity of ratio 260/230 values of 2-2.2 were used for the purity of nucleic acid. Lower 260/230 ratio values may indicate the presence of organic contaminants which absorb at 230 nm in the samples. The samples with the 260/280 ratio of higher than 1.21 and 260/230 ratio of higher than 1.41 were sent for Sanger sequencing.

3.7.4 Sanger Sequencing

All of the Sanger sequencing reaction were run at the Genomics and Proteomics Facility, Core Research and Equipment and Instrument Training Network, Memorial University of Newfoundland, using a BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems Inc., Foster City, CA). The following reagents were combined in a 0.2 mL PCR tube: 1 µL of 5X Sequencing Buffer, 0.5 µL of BigDye® Terminator v3.1 Cycle Sequencing mix, 1 µL of either 1.6 pmol/ µL forward or reverse primer, 10 ng of purified DNA template and H_2O to bring the volume to 10 µL. A control reaction containing 1 µL of 5X Sequencing Buffer, 0.5 µL of BigDye® Terminator v3.1 Cycle Sequencing mix, 1 µL of -21 M13 Control Primer, 1 µL of pGEM® -3Zf(+) template, and 6.5 µL of distilled water was included with each sequencing run. PCR was performed in a GeneAmp® PCR System 9700 Thermal Cycler (Applied Biosystems Inc., Foster City, CA) using the following protocol: 96 ºC for 5 min, 25 cycles of 96 ºC for 10 s, 50 ºC for 5 s, 60 ºC for 4 min; and a 4 ºC hold. The sequencing products were then purified with

an ethanol precipitation and subsequent ethanol wash with the addition of 2.5 µL of 125 mM EDTA and 65 µL of 95 % ethanol (Commercial Alcohols, Brampton, ON). The tubes were vortexed, centrifuged briefly, covered in aluminium foil and stored at 4 ºC overnight. The following day, the tubes were centrifuged at 3,000 x *g* for 30 min at RT, and the supernatants were discarded by inverting the tubes on a paper towel and a subsequent centrifugation of the inverted tubes (or plates) at 200 x *g* for 25-35 s. DNA pellets were suspended in 150 µL of 70 % ethanol by vortexing and pelleted by centrifugation at 3,000 x *g* for 15 min. Purified cycle sequencing products were resuspended in 15 µL of Hi-Di™ Formamide (Applied Biosystems Inc., Foster City, CA), briefly vortexed and centrifuged. Purified cycle sequencing were denatured for 3 min at 96 ºC in a GeneAmp® PCR System 9700 Thermal Cycler (Applied Biosystems Inc., Foster City, CA), loaded into a cassette and sequenced in a 3130 or 3730 Genetic Analyzer (Applied Biosystems Inc., Foster City, CA). Sanger sequencing for *Slc25a53* and *Tmsb15l* genes for ST/bJ, SJL/Bm, BUB/BnJ and PL/J strains done at the Genomics and Proteomics Facility, Core Research and Equipment and Instrument Training Network, Memorial University of Newfoundland.

3.7.5 Sequence Analysis

Sequences generated by Sanger sequencing methods were aligned for variant analyses using Sequencer® version 4.10.1 (Gene Codes Corporation, Ann Arbor, MI). All experimentally-determined sequences were aligned to the C57 strain reference genome, downloaded from Ensembl Genome Browser release 75 (Flicek *et al.*, 2014).

3.7.6 Protein Alignment across Species

Protein sequence comparisons between species were created for the indicated high priority genes (*Esx1*, *Slc25a53* and *Tmsb15l*) using ClustalW2 alignment program (Larkin *et al.*, 2007). These genes were chosen for further investigation due to the non-synonymous variant reported by NGS and confirmed by Sanger sequencing resulting in AA change and truncated protein. The species-specific AA sequences were obtained through the orthologue database of Ensembl Genome Browser release 75 (Flicek *et al.*, 2014). Two species with the highest target % ID were chosen from each species set (Primates, Rodents, Laurasiatheria, Sauropsida and Fish) for protein sequence alignment. Target % ID refers to the percentage of the sequence in the target specie (mouse) that matches to the query sequence (e.g. human).

4.0 Results

4.1 Phenotypic Mapping of Congenic SWR.C57-X Mice

Two groups of 5 congenic SWR.C57-X males from Line 1 and Line 2 were chosen to enter the paternal progeny test designed to independently test the tumor suppressor phenotype of the *Gct6*^{C57} locus in the presence of the *Gct1*^{SW} driver mutation for GC tumor susceptibility. The males were mated to SWR dams with a minimum, combined cohort of n=50 N7F1 females generated from each line to provide sufficient statistical power when dealing with a low penetrance trait. Both paternal groups carried *Gct1*^{SW} and *Gct6*^{C57} alleles and one line also carried *Gct4*^{C57} alleles, a recognized modifier locus for GC tumor susceptibility. The haplotype map for Chr X of each male entered into the progeny test is shown using SSLP DNA markers informative for *Gct4* and *Gct6* loci between SWR and C57 strains (Figure 4.1). N7F1 daughter offspring from each mating were implanted with DHEA capsules at puberty and examined for GC tumors at 8 wks of age. The GC tumor incidence was calculated as the number of females with GC tumors versus the total number of daughters examined, as is reported below, the locus haplotype for each male in Figure 4.1. The Line 1 female offspring with homozygous *Gct1*^{SW} on Chr 4 and paternally inherited *Gct4*^{C57} and *Gct6*^{C57} alleles on Chr X had zero incidence of GC tumors, even when treated with DHEA (0 of 50 examined); whereas the Line 2 N7F1 female offspring that carried homozygous *Gct1*^{SW} on Chr 4 and paternally inherited *Gct4*^{SW} and *Gct6*^{C57} alleles on Chr X had one GC tumor incidence identified (1 of 58 examined; see Figure 4.1). However, the GC tumor incidence of individual N7F1 females that inherited a paternal *Gct6*^{C57} allele in total (1/108 Line 1 and 2 combined) was significantly reduced compared to SWR inbred

females (9/51; Fisher's exact test p-value = 0.00016 < 0.05, Preacher *et al.*, 2001), but similar to paternal inheritance of *Gct6*^{CA} suppressor allele (0/58).

Thus both SWR.C57-X congenic lines were considered GC tumor suppressive while carrying the *Gct6*^{C57} allele. This proved our hypothesis that similar to the CAST strain, the C57 strain also harbours a GC tumor suppressor allele at *Gct6* locus on Chr X. Therefore, common genetic variants within CAST and C57 strain across the *Gct6* locus are candidate tumor suppressor alleles that initiate androgen sensitive GC tumors in GC tumor permissive strain (SWR) mice. Due to the highly polymorphic characteristics of the CAST strain, the search for the candidate tumor suppressor variant was approached by analyzing the genetic differences between SWR and C57 across the *Gct6* locus, to significantly reduce the number of variants to be investigated. An examination of GCT incidence in DHEA-treated females from the homozygous *Gct6*^{C57} congenic line further supports the conclusion that the *Gct6*^{C57} allele has GC tumor suppressor activity, since no GC tumors were observed among the 39 female mice examined at 8 wks age (Table 4.1).

Table 4.1 GC tumor incidence of homozygote *Gct6*^{C57} SWR.C57-X line

SWR.C57-X congenic	GC tumor incidence treated with DHEA	Total examined
Gct6^{C57} homozygote	0	33
Gct6^{C57} and *Gct4*^{C57} homozygote	0	6
Total	0	39

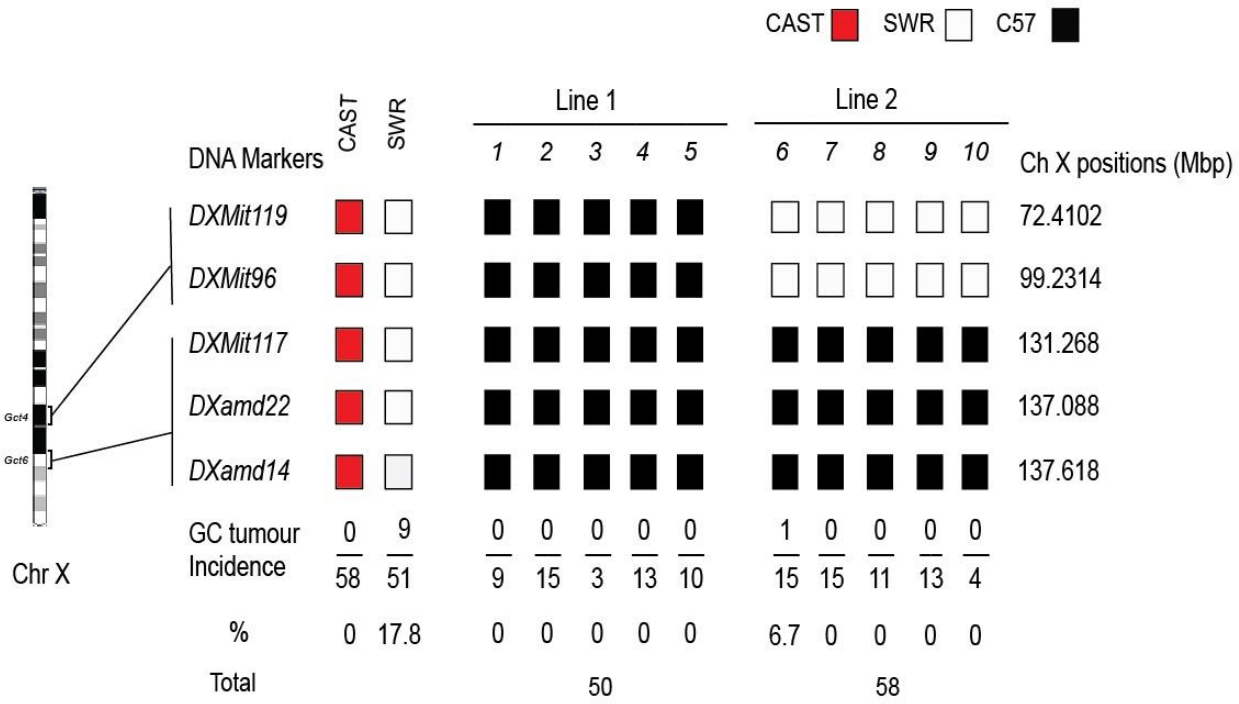


Figure 4.1 Haplotype map of SWR.C57-X congenic males and GC tumor incidence measured in their daughter offspring following pubertal treatment with DHEA.

The N7F1 female offspring of line 1 that carried *Gct1*^{SW}, *Gct4*^{C57} and *Gct6*^{C57} alleles were GC tumor suppressive as they had zero incidence of GC tumor after treatment with DHEA. Line 2 harbouring *Gct1*^{SW} and *Gct6*^{C57} alleles had one incidence of GC tumor after treatment with DHEA.

Red box: CAST; Black box: C57; White box: SWR.

4.2 Histology

The ovaries collected upon necropsy at 8 wks of age were fixated for histologic analysis. The GC tumor witnessed in line 2 harbouring only paternally-derived *Gct6*^{C57} was identified as GC tumor and it resembled the GC tumor histology of SWR mice (Figure 4.2).

4.3 Candidate *Gct6* Variants Identified by Whole Locus Sequencing

The locus capture plus NGS sequencing strategy for *Gct6* identified thousands of DNA sequence variations distributed across the locus when SWR or CAST strain genome was aligned to the C57 reference genome (Appendix E). After filtering based on AF criteria, the SWR vs C57 file contained 1151 variants while the CAST vs C57 file contained 4053 variants (Figure 4.3).

Phenotypic mapping of congenic SWR.C57-X mice proved that *Gct6*^{C57} also behaves as a GC tumor suppressor; therefore, the easier strategy was to evaluate the report with fewer variants list (SWR vs C57) and confirm in CAST genome when warranted. Subsequently, the files were cross referenced for common variants between SWR and CAST genomes, which would be removed from further consideration based on the confirmed phenotype differences between *Gct6*^{SW} and *Gct6*^{CA}. Three variants were identified in common between CAST and SWR that were removed from the data set.

The remaining 1148 variants were further filtered according to C57 and CAST alleles available for only the annotated variants. According to the selection criteria [(C57 = CAST) ≠ SWR], 402 additional variants were removed from the data set after compared to annotated SNPs in dbSNP databse. The remaining 746 variants were mapped out according to their gene position (Figure 4.4). Five hundred sixty of the 746 variants had rs IDs and were identified as "reference SNPs", 17 were categorized as "other annotated variants" as they were not given any

designated rs IDs and 169 were labeled as novel variants since they were not found in any of the online databases (dbSNP, Ensembl and Wellcome Trust Sanger Institute; see Figure 4.3). The 746 variants were 463 intronic, 75 exonic and 208 intergenic variants. The 75 exonic variants were given priority and into two categories of coding (10 variants) and non-coding (65 variants). From the 10 exonic/coding variants, 5 were nonsynonymous variants that caused an AA change. These 5 nonsynonymous variants were given highest priority for further analysis and confirmation through Sanger sequencing (Table 4.2).

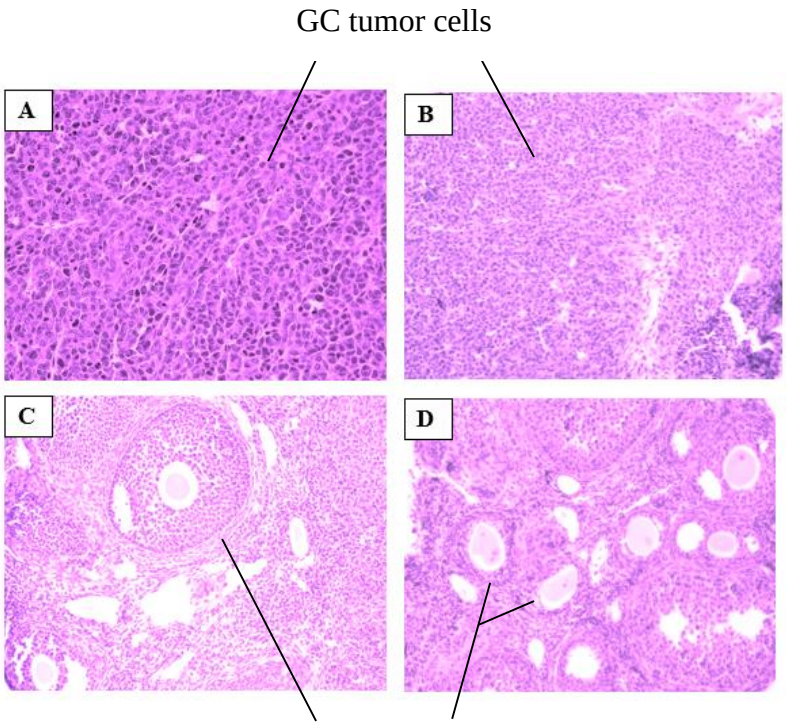


Figure 4.2 Histology of GCTs vs Normal ovaries.

A. GC tumor from line 2 congenic SWR.C57-X N7F1 female offspring carrying *Gct6*^{C57} allele taken at 40X magnification. **B.** GC tumor from SWR mouse taken at 20X magnification. **C. and D.** Represent two normal ovaries taken from two different mice from line 2 congenic SWR.C57-X N7F1 female offspring carrying *Gct6*^{C57} allele taken at 20X magnification.

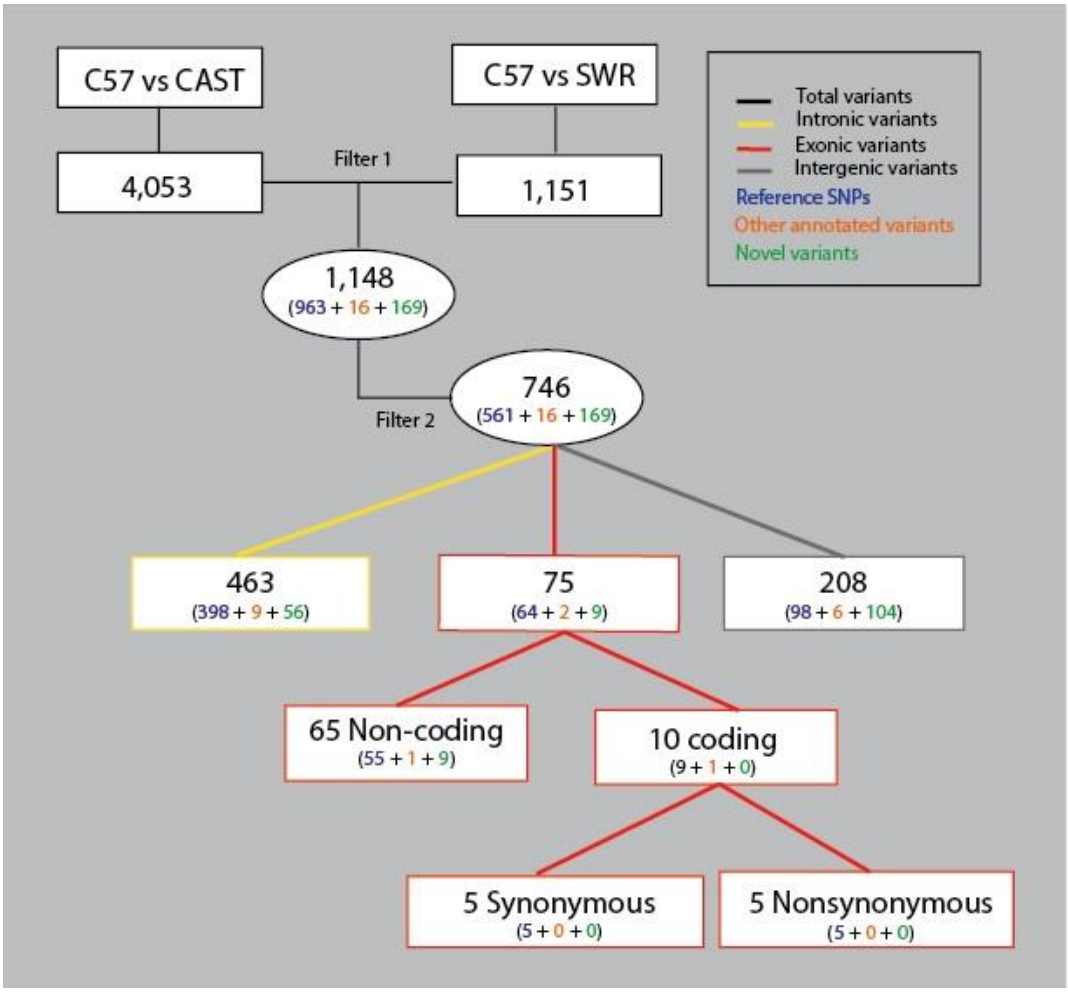


Figure 4.3 Filtering strategy for homozygous Single Nucleotide Polymorphisms (SNPs) between CAST vs. C57 and SWR vs. C57 reference genome across the mapped *Gct6* locus.

Targeted whole-locus capture of the *Gct6* region (Agilent Sure-Select Array) from SWR inbred and SWR.CAST-X congenic DNA was processed for NGS (Illumina Hi-Seq) at The Jackson Laboratory. Variant calls were made relative to the C57 reference genome for both $Gct6^{CAST}$ and $Gct6^{SWR}$ input. The CAST region (4,053 variants) was significantly more polymorphic than SWR (1,151 variants), such that C57 vs SWR variants were prioritized for analysis and subsequently compared to CAST genome. Three SNPs common between CAST and SWR were filtered out, leaving 1148 variants (Filter 1). The referenced portion of the 1148 were checked for CAST allele via dbSNP website and the variants that did not have the same allele as C57 were removed from data set (Filter 2) leaving only 746 variants left to be categorized as intronic, exonic or intergenic SNPs and further subdivided as reference SNPs, "other annotated" or "novel" variants. Five nonsynonymous variants were isolated for further analysis.

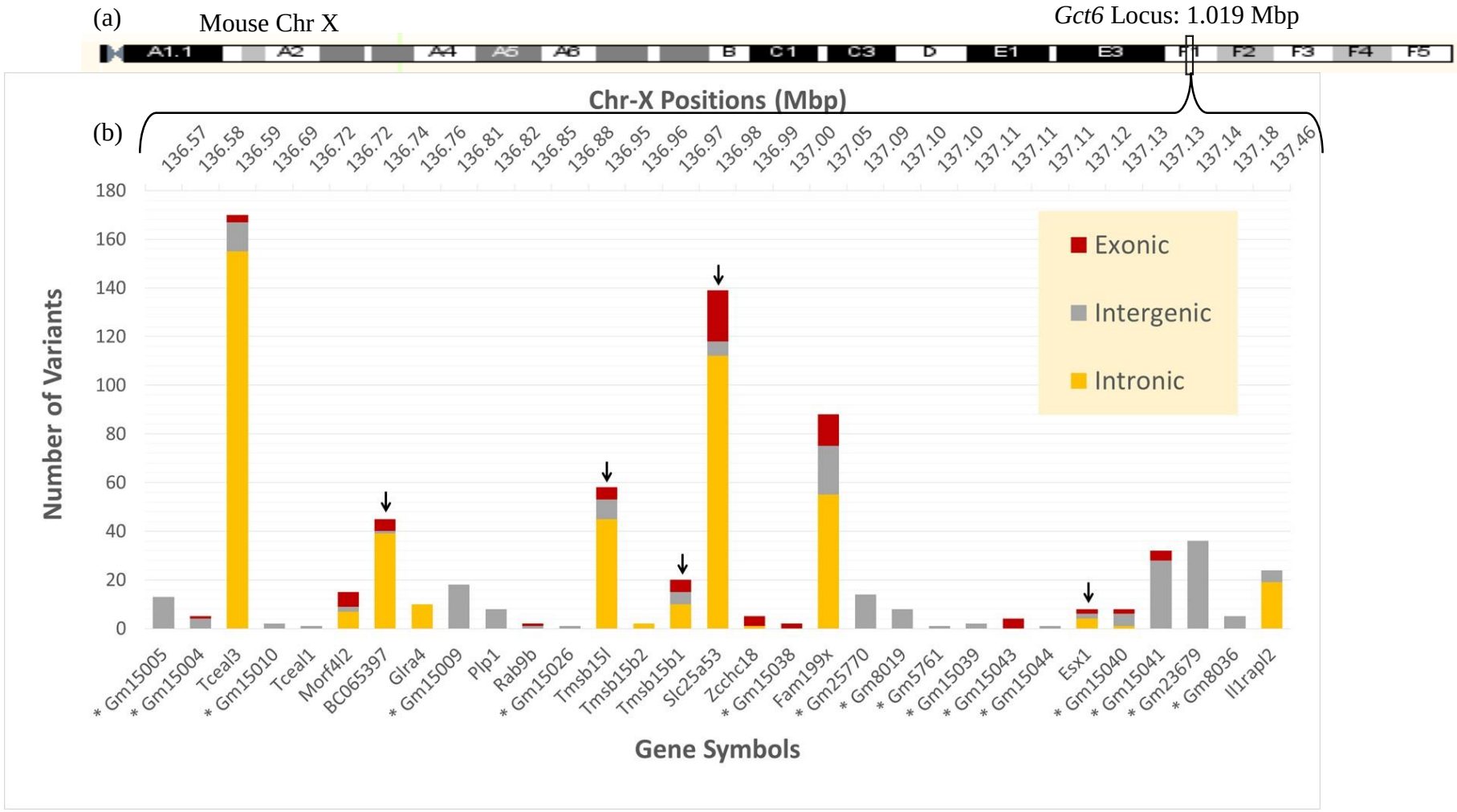


Figure 4.4 Physical distribution and annotation of 746 SNPs identified between C57 and SWR genomes across the *Gct6* locus.

Schematic diagram of mouse Chr X (GRCm38; Ensembl Genome Browser release 75, Flicek *et al.*, 2014). The *Gct6* locus of 1.019 Mbp on distal region of Chr X of mouse is shown by a black rectangle. (b) The 746 NGS variants isolated from "SWR vs C57" data set are mapped out and distributed unevenly among the 20 candidate genes of *Gct6* GC tumor suppressor. The variants are further subdivided into different color-coded categories of exonic (red), intronic (orange) and intergenic (grey). The 5 nonsynonymous variants are shown by black arrows.

Table 4.2 Variant sequence validation and predicted A. A. deviations in SWR vs C57.

rs IDs of 5 nonsynonymous variants were taken from the Ensembl website (GRCm38; Ensembl Genome Browser release 75, Flicek *et al.,* 2014). All the 5 nonsynoymous NGS reported variants were confirmed by Sanger sequencing.

Gene Symbol	SNP reference ID	Chr X position (Mb)	NGS reported variant	Sanger Sequence confirmation	Consequence
Slc25a53[*]	*rs29261702*	137.0	C/T	C/T	ACC/ATC - Thr/Ile
Tmsb15b1	*rs13484019*	136.9	C/T	C/T	CCG/CTG - Pro/Leu
Tmsb15l	*rs13484019*	136.9	C/T	C/T	CCG/CTG - Pro/Leu
Esx1	*rs240547705*	137.1	T/G	T/G	TCC/Gcc - Ser/Ala
BC065397	*rs244048438*	136.7	GTTT/GTTTGTTT	GTTT/GTTTGTTT	GTTT/GTTTGTTT - Phe/Leu[⅄]

* Sanger sequencing *Slc25a53* in SWR mice revealed the NGS reported variant and the deletion of 28 bp from two adjacent 28 bp repeats.

⅄ GTTT frame shift insertion bypasses stop codon and adds nine additional AA (Leu, Phe, Leu, Arg, Pro, Val, Val, Thr and Arg) before reaching a downstream secondary stop codon.

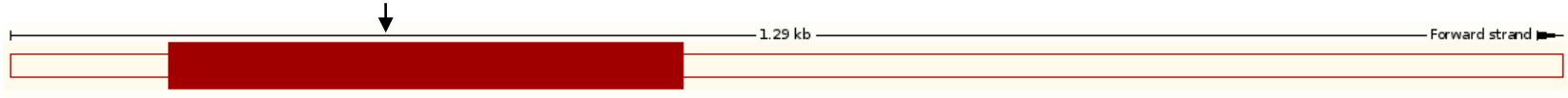
4.1 Sanger Sequencing

4.1.1 BC065397

The gene *BC065397* is not annotated in humans; in mice, however, it is a protein-coding gene with only one transcript (BC065397-201 ENSMUST00000089350) that has one coding exon that encodes 141 AA residues (Figure 4.5 a). NGS reported an INDEL variant of GTTT/GTTTGTTT (*rs244048438*) at genomic location of 136,742,379. This INDEL causes a frame shift which substitutes Phe (aromatic AA) by Leu (nonpolar aliphatic AA) causes the original stop codon to be bypassed and add nine AA residues (Leu, Phe, Leu, Arg, Pro, Val, Val, Thr and Arg) before reaching a secondary stop codon.

Sanger sequencing, BC065397, revealed that both C57 and CAST strains carry GTTT allele while SWR strain carries GTTTGTTT INDEL allele (Figure 4.5 b). However, the allele information obtained through Ensembl Genome Browser revealed that the INDEL did not co-segregate with other GC tumor permissive strains (ST/bJ, SJL/Bm, BUB/BnJ and PL; see Table 4.3). Therefore, BC065397 is given a low priority in our search for the Gct6 candidate gene.

(a)



(b)

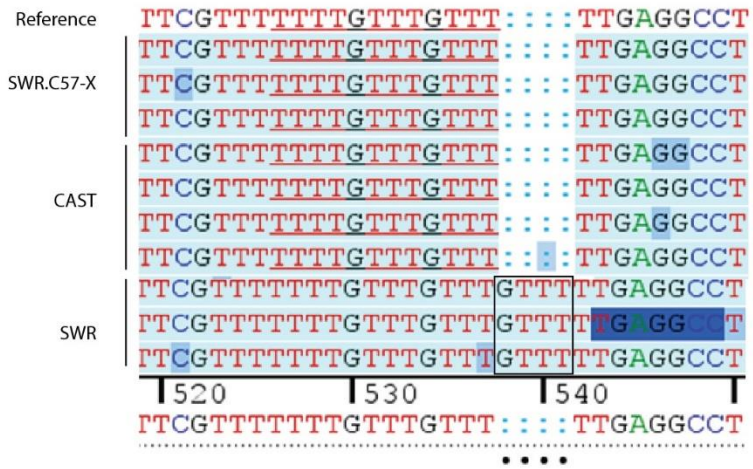


Figure 4.5 Sanger sequence alignment of *BC065397* gene in SWR.C57-X, CAST and SWR mice.

(a) Schematic diagram of BC0650397 transcript (BC065397 ENSMUSG00000072960); the coding exon is shaded in red and non-coding exons are represented by non-shaded boxes (Ensembl Genome Browser release 75, Flicek *et al.*, 2014). Exon 1 (black arrow) with flanking sequences was amplified and sequenced. (b) Sanger sequencing of BC065397 gene confirmed NGS reported variant (black box). Only SWR strain carries GTTT INDEL variant.

Light Blue: excellent sequence quality, Medium Blue: Average sequence quality, Dark Blue: poor sequence quality, Not-shaded: Reference sequence

4.1.2 Esx1

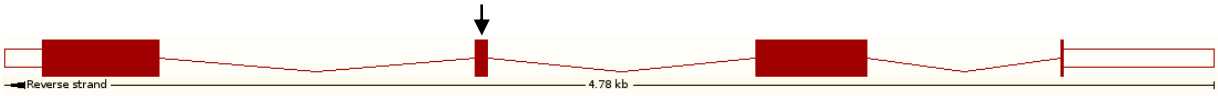
Human *Extraembryonic, Spermatogenesis, Homeobox 1 gene (ESX1)* gene has one transcript (ESX1-001 ENST00000372588) that is a protein coding transcript of 1,495 bp containing sequences of four coding exons and translating 406 amino acid residues (Figure 4.6 a); while mouse *Esx1* gene has 3 protein coding transcripts. The NGS reported variant resides within the first transcript of *Esx1* at 1,692 bp (Esx1-002; ENSMUST00000113066) that has four coding exons translating 314 AA residues (Figure 4.6 b). NGS showed a variant of T > G (rs240547705) at 137,118,653 bp position. This SNP replaces Ser (polar uncharged AA) with Ala (nonpolar AA) and was confirmed through Sanger sequencing (Figure 4.6 c). *Esx1* has multiple highly conserved regions among species however this particular SNP is within less conserved region of the protein which suggests that this SNP is likely to cause any disease (Appendix A). *Esx1* Sanger sequence revealed that both C57 and CAST strains carry the "T" allele while the SWR strain carries the "G" allele. Further analysis of the *Esx1* alleles among other GC tumor permissive strains (ST/bJ, SJL/Bm, BUB/BnJ and PL) indicated that this allele does not co-segregate with all the GC tumor permissive strains (Table 4.3). Therefore, this gene is given low priority for the candidate gene *Gct6* locus GC tumor suppressor.

4.1.1 Tmsb15l and Tmsb15b1

Thymosin β 15b like (Tmsb15l) and *Thymosin β 15 1 (Tmsb15b1)* genes are both protein-coding genes in mice. In humans, the closest gene to *Tmsb15l and Tmsb15b1* is *TMSB15A* which has one protein coding transcript (TMSB15A-001 ENST00000289373) of 685 bp with 3 coding areas translating 45 AA residues (Figure 4.7 a).

(a)

(b)



(c)

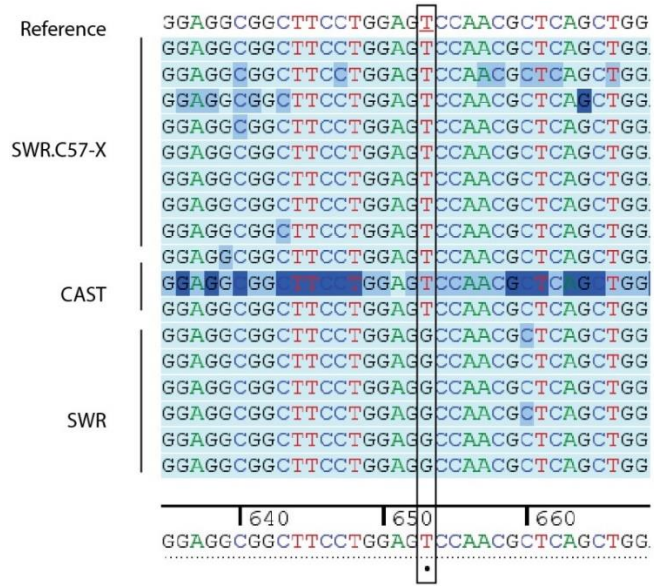


Figure 4.6 Sanger sequence alignments of *Esx1* gene in SWR.C57-X, CAST and SWR mice.

(a) Schematic diagram of human *ESX1* protein coding transcript (ESX1-001 ENST00000372588) with 4 coding exons shaded in yellow and noncoding regions are represented by non-shaded boxes (Ensembl Genome Browser release 75, Flicek *et al.*, 2014). (b) Schematic diagram of mouse *Esx1* protein coding transcript (*Esx1*-002; ENSMUST00000113066) with sequence of 4 coding exons shaded in red and noncoding regions are represented by non-shaded boxes (Ensembl Genome Browser release 75, Flicek *et al.*, 2014). Exon 2 (black arrow) with flanking intronic sequences was amplified and sequenced. (c) Sanger sequence alignment of SWR.C57-X, CAST and SWR. NGS reported SNP T > G (C57 vs SWR) was confirmed (shown in black box). Both SWR.C57-X and CAST carry the "T" allele while SWR carries "G" allele.

Light Blue: excellent sequence quality, Medium Blue: average sequence quality, Dark Blue: poor sequence quality, Not-shaded: Reference sequence

Table 4.3 *Gct6* alleles associated with GC tumor- permissive or - suppressive activity.

Strain Cross (maternal x paternal)	DHEA-induced GCT incidence % (n)	*Gct6* allele phenotype	*Slc25a53* (-strand)	*Slc25a53*◈ 28 bp repeat Deletion-Insertion	*Tmsb15l* (- strand)	*Tmsb15b1* (- strand)	*Esx1* (- strand)	*BC065397* (+ strand)
SWR (inbred)	17.8 (51)[1]	Permissive	T	Deletion	T	T	G	GTTTGTTT
SWR x SJL/Bm (F1)	14.0 (107)[2]	Permissive	T	Deletion	T	T	G	GTTTGTTT
SWR x BUB/BnJ (F1)	13.7 (95)[2]	Permissive	T	Deletion	T	T	G	GTTTGTTT
SWR x PL/J (F1)	10.0 (80)[2]	Permissive	T	Deletion	T	T	G	GTTT
SWR x ST/bJ (F1)	21.9 (32)[2]	Permissive	C	Insertion	C	C	T	GTTTGTTT
SWR x C57 (F1)	0.00 (155)[2]	Suppressive	C	Reference	C	C	T	GTTT
SWR x SWR.C57-X (F1 congenic)	0.90 (108)	Suppressive	C	Reference	C	C	T	GTTT
SWR x SWR.CAST-X (F1 congenic)	0.00 (58)[3]	Suppressive	C	Reference	C	C	T	GTTT

Shaded area indicates suppressive strains.

◈ After AA 11 and 20 of 346 the *Slc25a53* frame shift deletion and insertion add 63 and 78 additional residues before reaching their secondary stop codons, respectively.

[1] Smith K.N., *et al.,* 2013. Mammalian Genome. 24: 63-71.

[2] Beamer W.G. *et al.,* 1988. Cancer Research 48:2788-2792.

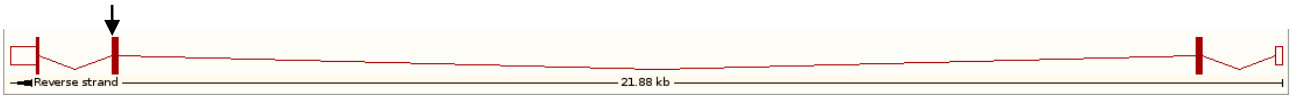
[3] Dorward A.M., *et al.,* 2013. Epigenetics 8:184–191.

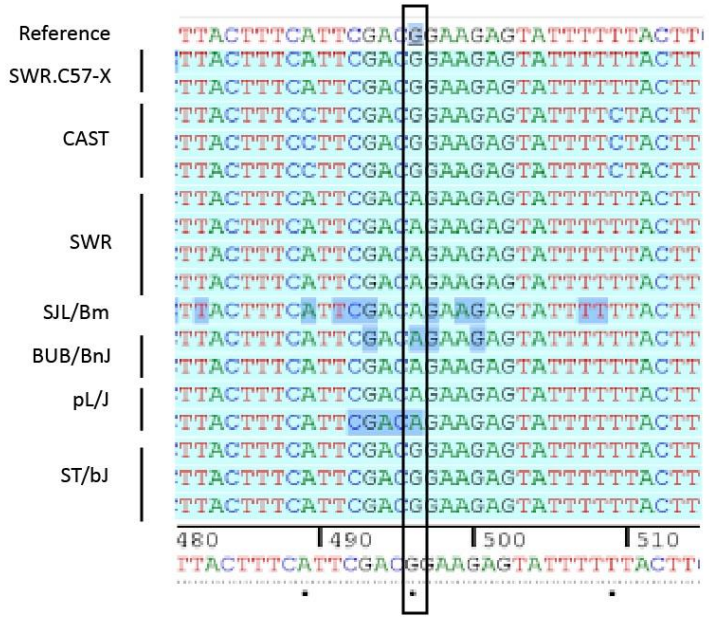
In mice, *Tmsb15l* and *Tmsb15b1* are positioned on Chr X in such a way that approximately 50 % of their sequences overlap. Analysis of the NGS data showed a T > C SNP (*rs13484019*) located within this overlapping region. The NGS report therefore includes two records of this SNP. The same amino acid change occurs in both Tmsb15l and Tmsb15b1. Hence only one primer was used to target this region.

Tmsb15l has only one transcript (Tmsb15l-001; ENSMUST00000127404) of 816 bp that has 4 exons coding for 80 AA residues (Figure 4.7 b). NGS reported T > C change at 136,975,409 bp position that causes a substitution of Pro (a polar uncharged AA) to Leu (Nonpolar AA) confirmed by Sanger sequencing (Figure 4.7 c). Upon *Tmsb15l* protein comparison between species it was revealed that *Tmsb15l* protein carries an extra 35 AA residues at the N terminus of the protein that only exists in mice (Appendix B). This extra AA residues might serve as a signal peptide, although not clearly annotated as a signal peptide in BLAST/BLAT Ensembl genome browser. *Tmsb15l* is a highly conserved protein and therefore it is presumed to have an evolutionary important function (Appendix B). However, the lack of correlation between the alleles and the phenotyped strains reduced the priority of the variant as sharing identity with *Gct6*.

(a)

(b)





(c)

Figure 4.7 Sanger sequence alignment of *Tmsb15l* gene in SWR.C57-X, CAST, SWR and the permissive and suppressive strains associated with GC tumor activity in mice.

(a) Schematic diagram of *TMSB15A* transcript with 3 coding exons shaded in yellow and noncoding regions are represented by non-shaded boxes (yellow boxes; Ensembl Genome Browser release 75, Flicek *et al.*, 2014). (b) Schematic diagram of *Tmsb15l* transcript with 4 coding exons shaded in red and noncoding regions are represented by non-shaded boxes (Ensembl Genome Browser release 75, Flicek *et al.*, 2014). Exon 2 with flanking sequences was amplified and sequenced (black arrow). (c) Sanger sequencing of *Tmsb15l* gene confirmed NGS reported SNP (shown in black rectangle) of G > A. SWR.C57-X, CAST and ST/bJ carry the "G "allele while the SWR, SJL/Bm, BUB/BnJ and PL/J strains carry the "A" allele.

Light Blue: excellent sequence quality, Medium Blue: Average sequence quality, Dark Blue: poor sequence quality,

Not-shaded: Reference sequence

(a)

(b)

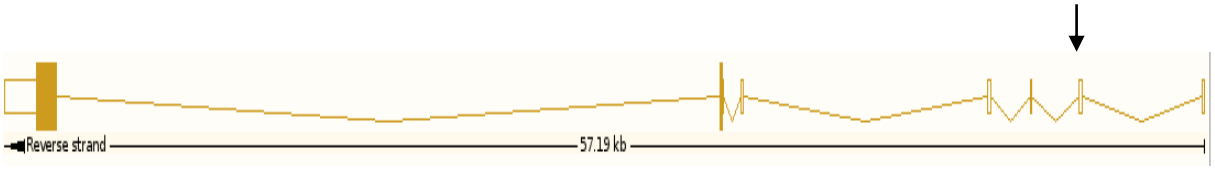


Figure 4.8 Human *SLC25A53* gene comparison with mouse *Slc25a53* gene

(a) Schematic diagram of human *SLC25A53* transcript indicating its two coding exons (yellow boxes) exons and noncoding regions represented by non-shaded boxes. (b) Schematic diagram of *Slc25a53* gene indicating the 7 coding exons (yellow boxes) and noncoding regions represented by non-shaded boxes. NGS reported SNP reside within Exon 6 (black arrow).

Ensembl website (GRCm38; Ensembl release 75 - Febuary 2014 © WTSI / EBI)

4.1.1 SLC25A53

In humans, *Solute Carrier Family 25, Member 53 (SLC25A53)* also known as *MCART6,* is a protein-coding gene that has two transcripts. Only one transcript, however, is a protein coding transcript (SLC25A53-001 ENST00000357421); it is 6,224 bp long containing sequences of two coding exons which when translated produces a peptide with 307 amino acids (Figure 4.8 a). In mice, *Slc25a53* has 14 transcripts with only five annotated as protein coding transcripts. The NGS data indicated -SNP in *Slc25a53* gene resides in the transcript *Slc25a53*-001 (ENSMUST00000113069) containing sequences of 7 coding exons and translating 346 AA residues (Figure 4.8 b). The PCR amplicons were checked for appropriate size before Sanger sequencing using agarose electrophoresis. In the case of *Slc25a53,* there were amplicon size differences between SWR, CAST and C57. The SWR genomic samples produced a smaller PCR fragment while the SWR.C57-X and SWR.CAST-X samples had the expected PCR amplicon size of 315 bp (Figure 4.9).

Sanger sequencing of *SLc25a53* in SWR, SWR.C57-X and CAST confirmed the NGS reported SNP of C > T change, causing an AA substitution of Thr (polar uncharged AA) with Ile (nonpolar aliphatic AA). Sanger sequencing also revealed a 28 bp fragment deletion of 2 adjacent 28 bp repeats in SWR mice. Further analysis of *Slc25a53* gene in GC tumor permissive strains (SJL/Bm, BUB/BnJ, PL/J and ST/bJ) showed yet another difference in PCR fragment size among the strains once separated by 1 % gel electrophoresis (Figure 4.10). It was observed that STb/J strain carries a longer amplicon than SWR.C57-X and SWR.CAST-X strains while the rest of the permissive strains harboured a shorter amplicon. Sanger sequencing of all the permissive strains revealed that the same 28 bp fragment deletion exists in SJL/Bm, BUB/BnJ and PL/J but not ST/BJ. Instead, ST/bJ Sanger sequence showed an 8 bp fragment deletion 84

bp upstream the *Slc25a53* exon 6 (the UTR region; see Figure 4.11 a) and a 28 bp repeat insertion immediately after the end of the two repeat sequence in the coding region of *Slc25a53* exon 6 (Figure 4.11 b). The 8 bp deletion and 28 bp insertion in ST/bJ strain compensated for the slightly (20 bp) heavier ST/bJ (335 bp) band. Similarly, the 28 bp fragment deletion in SWR, SJL/Bm, BUB/BnJ, PL/J accounted for their lighter (-28 bp) band of 287 bp when compared to SWR.C57-X and SWR.CAST-X controls that has the intended 315 bp amplicon. *Slc25a53* gene is highly conserved among species; however, the NGS reported SNP resides within a region that does not exists in any other species (Appendix C). Protein sequence comparison between species also revealed that mice have an extra 40 AA residues at the N terminus of *Slc25a53* protein (Appendix C). These extra AA residues are thought to be a signal peptide although it is not specifically annotated as a signal sequence in public databases. Even though the NGS reported SNP did not co-segregate completely with GC tumor permissive and suppressive phenotype, it is evident that all the GC tumor permissive strains have some sort of deviation of either an insertion or a deletion thus lacking an intact *Slc25a53* gene (Table 4.3). This evidence has given *Slc25a53* gene the highest priority and the best possible gene candidate for *Gct6* GC tumor suppressor activity.

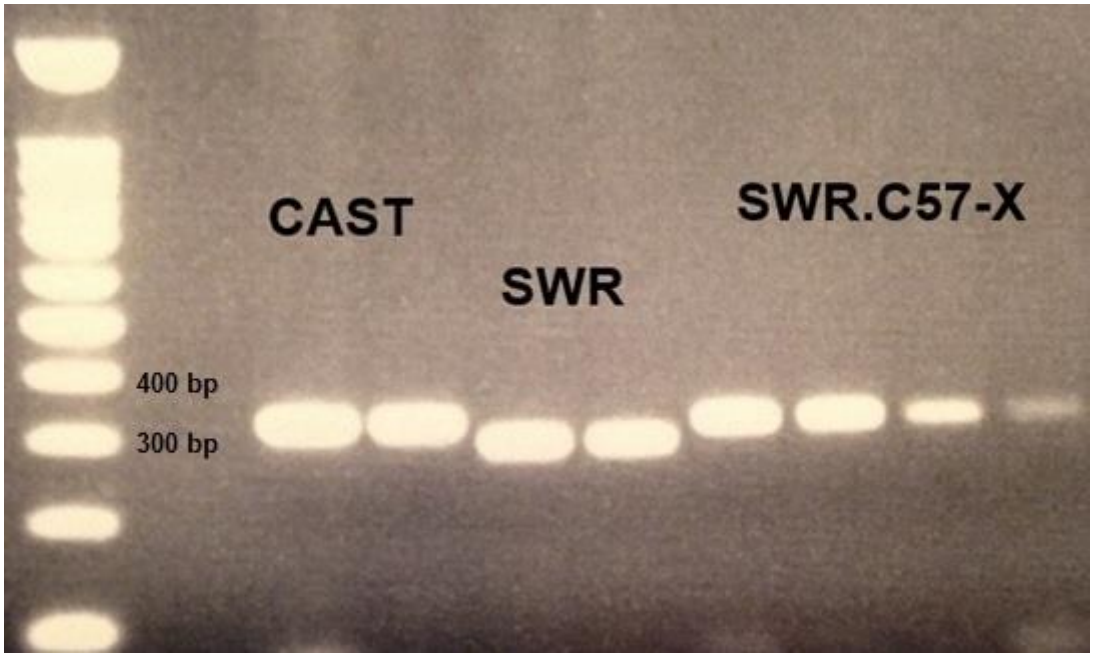


Figure 4.9 *Slc25a53* check gel prior to Sanger sequencing

The *Slc25a53* amplicons ran for one hour at 110 V on a 1% agarose gel prior to sequencing with two independent DNA samples from SWR.CAST-X and SWR mice, and four independent DNA samples from congenic line SWR.C57-X. SWR. The 100 base pair 1 Kb DNA ladder marker was used along with Bromophenol blue loading dye for the horizontal gel electrophoresis analysis.

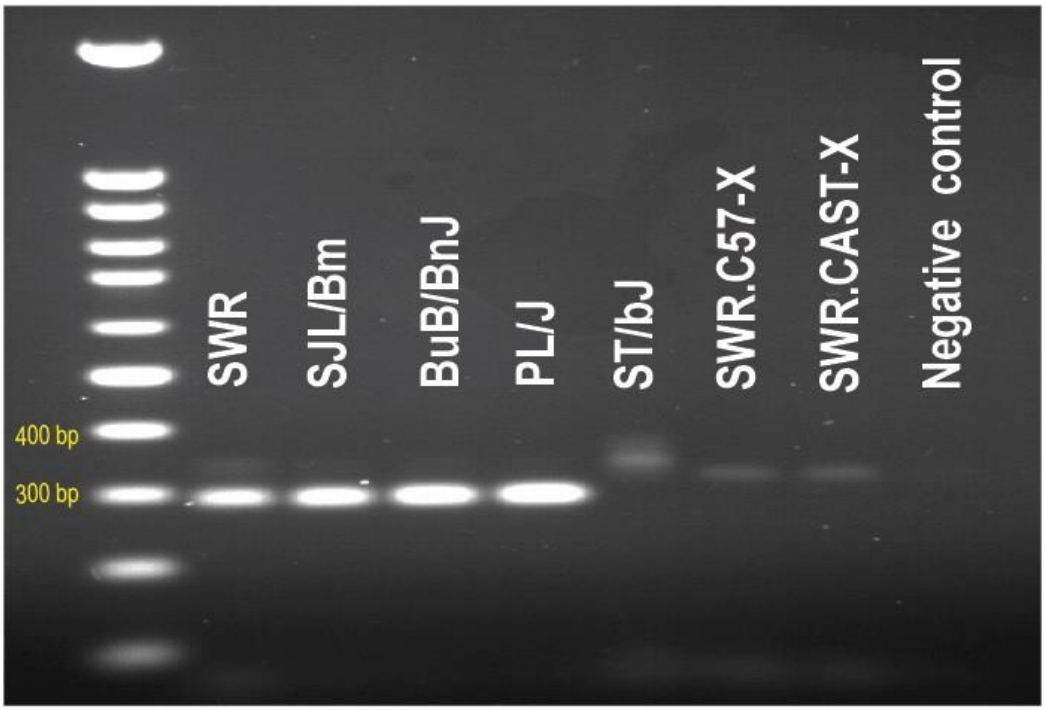


Figure 4.10 *Slc25a53* alleles across GC tumor – suppressive and –permissive strains.

Allele association analysis between 5 *Gct6* permissive (SWR, SJL/J, PL/J, BUB/BnJ, ST/bJ) and 2 *Gct6* suppressive (CAST, C57) strains revealed size differences across GC tumor – suppressive and –permissive strains. SWR, SJL/J, BUB/BnJ and PL/J strains carry the lighter *Slc25a53* gene amplicon of 287 bp while the ST/bJ strain carries a heavier band of 335 bp. The suppressive strains (SWR.C57-X and SWR.CAST-X) carry the intended amplicon of 315 bp fragment of *Slc25a53* gene. The amplicons were ran for one hour and half at 110 V on a 1% agarose gel. The 100 base pair 1 Kb DNA ladder marker was used along with Bromophenol blue loading dye for the horizontal gel electrophoresis analysis.

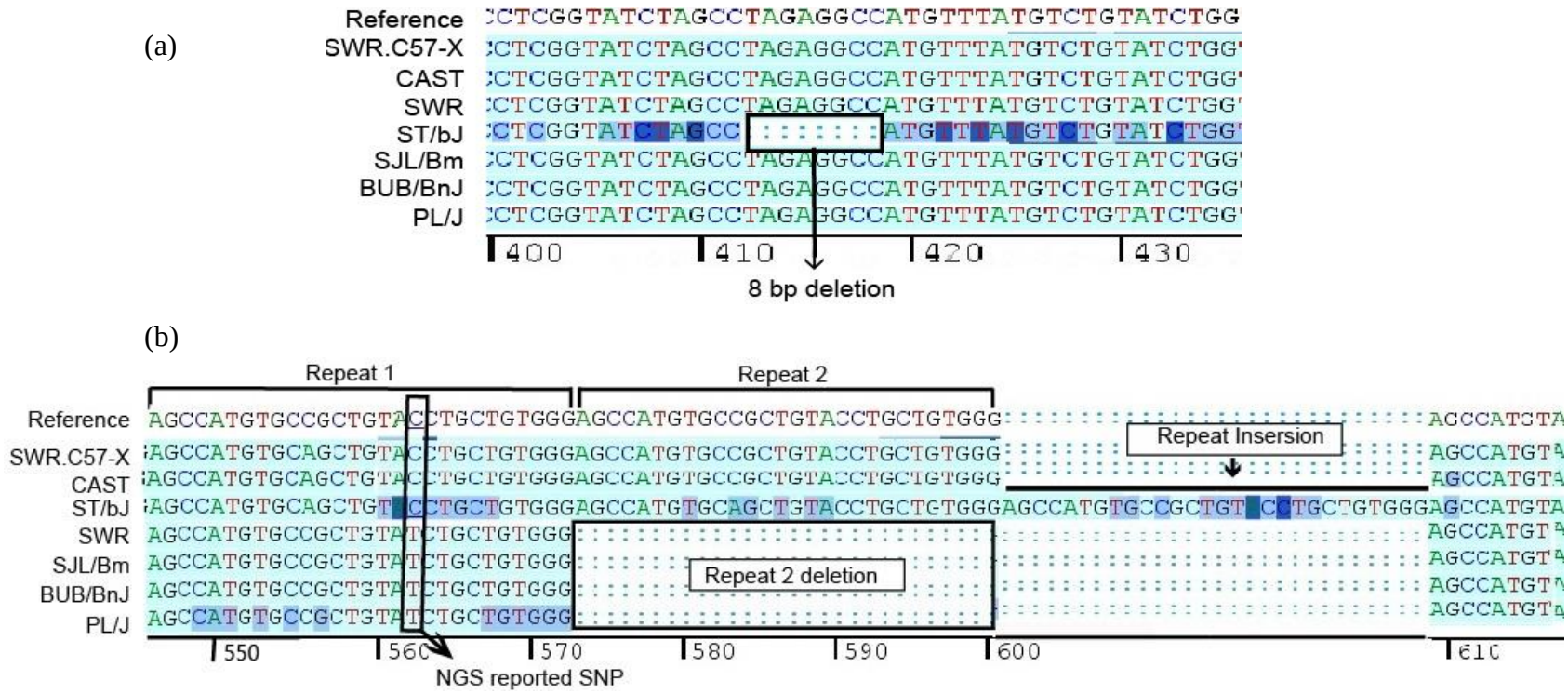


Figure 4.11 Multiple Sanger sequence alignments of *Slc25a53* gene in GC tumor permissive and suppressive strains.

(a) This figure is the upstream flank of *Slc25a53* Exon 6. Sanger sequencing of ST/bJ strain revealed an 8 bp deletion located 84 bp upstream of *Slc25a53* Exon 6 UTR. (b) This figure represents the sequence within the coding region of *Slc25a53* Exon 6. The NGS reported SNP C > T (C57 vs SWR) was confirmed through Sanger sequencing. Sanger sequencing of GC tumor permissive strains (SWR, ST/bJ, SJL/Bm, BUB/BnJ and PL) showed that all but ST/bJ carry the "T" allele. Ten bp downstream of the NGS reported SNP, there is a 28 bp deletion of repeat 2 in SWR, SJL/Bm, BUB/BnJ and PL/J strains. After repeat 2, there is a 28 bp insertion of the repeat sequence in ST/bJ. (This diagram was generated manually to best illustrate the insertion, deletion and repeats that take place in Slc25a53 gene among GC tumor –permissive and –suppressive strains).

Light Blue: excellent sequence quality, Medium Blue: Average sequence quality, Dark Blue: poor sequence quality, Not-shaded: Reference sequence

5.0 Discussion

A spontaneous, heritable GC tumor phenotype in SWR inbred female mice has revealed several *Gct* susceptibility loci influential for early tumor initiation in the presence of an androgenic environment: *Gct1* (Beamer *et al.,* 1988; Beamer *et al.* 1998; Dorward *et al.,* 2005; and Smith *et al*., 2013), *Gct4* (Beamer *et al.,* 1988; Beamer *et al.* 1998) and *Gct6 (*Dorward *et al.,* 2003 and Dorward *et al.,* 2013). In the SWR model, paternal inheritance or biparental inheritance of the *Gct6* locus from the CAST genome contributes a GC tumor resistant function. The tumor suppressor activity of the $Gct6^{CA}$ allele overrides the activity of the *Gct4* GC tumor modifier locus on Chr X and the oncogenic activity of the $Gct1^{SW}$ alleles on Chr 4, indicating that the $Gct6^{CA}$ allele suppresses the androgen-initiated GC tumorigenic program while preserving the normal function of the mouse ovary. The suppressor activity contributed by $Gct6^{CA}$ was unique when compared to all previous strains examined in breeding crosses with the SWR strain, suggesting a unique CAST allele within the mapped *Gct6* locus was a master controller of a biologic pathway that influenced GC tumor susceptibility. The only other strain examined by Beamer *et al*. that suggested equivalent tumor suppression was the C57 strain, although the activity was not mapped to any specific chromosome (Beamer *et al.,* 1985). The results of this project have confirmed the $Gct6^{C57}$ locus confers GCT suppressor activity similar to $Gct6^{CA}$, which also provided a more powerful strategy for *Gct6* variant analysis following targeted NGS sequencing of the locus.

5.1 $Gct6^{C57}$ allele contribution to GC tumor susceptibility

Through the development of the six-generation SWR.C57-X homozygous congenic lines, the independent contribution of $Gct6^{C57}$ allele to GCT resistance was tested. Based on the

experimental history that paternal inheritance has a dominant influence over Chr X-linked GCT susceptibility loci, my experiment was initially designed as a progeny test, mating SWR dams with SWR.CAST-X sires who contributed a Chr X genomic segment harbouring only C57 alleles ($Gct4^{C57}/Gct6^{C57}$) or a combination of alleles ($Gct4^{SW/}Gct6^{C57)}$ to daughter offspring in the presence of SWR strain genome at all other loci (Dorward *et al.,* 2003; Dorward *et al*., 2013). The female offspring were subsequently examined for GCT susceptibility with DHEA exposure beginning at puberty. Similar to the CAST genomic contribution at *Gct6,* we determined that paternal inheritance of the *Gct6* locus from the C57 genome leads to a significant reduction in GC tumor initiation. Female mice examined from either (SWR x SWR.C57-X) cross that inherited a paternal $Gct6^{C57}$ allele revealed only 1 confirmed unilateral GCT at necropsy of a total of 108 females examined from both congenic lines combined. This proportion was significantly reduced compared to the expected GCT incidence of SWR inbred females treated with the same DHEA supplementation paradigm and sacrificed at 8 wks of age (9/51; Fisher's exact test, p = 0.00016; Smith *et al.,* 2013). In contrast, low GCT incidence following paternal inheritance of $Gct6^{C57}$ was similar to the suppressed GCT incidence measured following paternal inheritance of the $Gct6^{CA}$ tumor-resistance allele (Dorward *et al.,* 2013). One alternate interpretation for the low GCT incidence measured in (SWR x SWR.C57-X) F1 females receiving DHEA treatment, was ineffective GCT stimulus by the DHEA protocol, since the positive control strain SWR was not simultaneously included in my experimental design. However, a concurrent DHEA implantation experiment ongoing in the lab using SWR inbred female mice, the same DHEA source and DHEA administration protocol, the same timing of surgical implant at puberty but a slightly earlier age at necropsy (5-6 wks as opposed to 8 wks) recorded 15.4 % GCT incidence: (8/52 confirmed tumors) with additional suspicious GCT

specimens (Ms. Kerri Smith). The historically consistent GCT incidence rate in SWR females administered DHEA (Beamer *et al.,* 1985 and Smith *et al.,* 2013) gives strong support to my conclusion that low GCT incidence in (SWR x SWR.C57-X) F1 females is due to the tumor suppressor activity of $Gct6^{C57}$ allele, not an ineffective tumor induction protocol. Thus, the SWR.C57-X paternal progeny test supported the hypothesis that a common *Gct6* tumor suppressor allele in C57 and CAST genomes is distinct from the permissive $Gct6^{SW}$ allele, leading to our hypothesis that common phenotypes are supported by a common genotype, or a common genetic target, between the CAST and C57 strains, which is unique from the SWR strain or other proven GCT-permissive strains (Dorward *et al.,* 2013).

The paternal dominance of the X-linked $Gct6^{CA}$ allele is observed though the recorded 0 % GCT incidence of F1 females that inherit two different *Gct6* alleles, the maternal permissive ($Gct6^{SW}$) and the paternal suppressive ($Gct6^{CA}$) alleles (Dorward *et al.*, 2013). The outcome is not the same when heterozygous females inherit the $Gct6^{CA}$ allele maternally, reinforcing the paternal parent-of-origin effect for the GCT susceptibility loci (Dorward *et al*., 2013) Such a reciprocal cross has not been completed with the newly developed SWR.C57-X strain, although the findings from the paternal progeny test described suggests a similar paternal effect to the SWR.CAST-X studies. The dominant paternal effect of C57-X and CAST-X congenic lines precludes an allele complementation test, which would have been an excellent functional test using our congenic resources to determine whether $Gct6^{CA}$ and $Gct6^{C57}$ are the same alleles. An allele complementation test works for recessive traits or semi-dominant traits, where two suspected ‘‘mutant’ alleles are brought together by selective breeding to see if the trait is fully reconstituted. In this case, this would only have worked if there was a measurable dose response relationship between the inheritance of two permissive alleles (20 % predicted GCT incidence),

the heterozygous condition of one permissive and one suppressor allele (predicted incidence of 10 % if semi-dominant action, or 0 or 20 % if either permissive or suppressor alleles had recessive activity, respectively) or two suppressor alleles (0 % GCT incidence). Regardless, the common paternal suppressor activity of *Gct6*^{C57} and *Gct6*^{CA} identified using SWR-derived congenic lines provided strong evidence for a common *Gct6* allele or genetic target, which was applied in our filtering strategy for the NGS-variants within the mapped *Gct6* region.

The paternal parent-of-origin effect has revealed itself a significant influence for two *Gct* loci, *Gct4* and *Gct6*, on two disparate regions of Chr X. This cumulative evidence suggests biased X-inactivation is in effect for this ovarian tumor-susceptibility phenotype, as opposed to allele-specific parent-of-origin expression patterns. Global X-inactivation bias is not anticipated across all tissues of SWR female mice, since inheritance of an X-linked Green fluorescent protein (GFP) transgene exhibited appropriate fluorescence mosaicism in the skin of hemizygous carrier female pups, whether the transgene was inherited maternally or paternally (Dorward, unpublished observations). It is our current hypothesis that the first ovarian follicle wave is the site of the androgen-sensitive activity of the *Gct* tumor susceptibility genes. Implementation of fluorescent trace analysis to isolate the GC populations from this first wave would permit a more sensitive assay for X-inactivation bias in this developmentally-restricted follicle cohort.

5.2 *Gct6* tumor suppressor candidate genes

Paternal inheritance of either the *Gct6*^{CA} or *Gct6*^{C57} locus on Chr X was sufficient to completely or significantly suppress GC tumor initiation in SWR.CAST-X and SWR.C57-X females, respectively. This finding supported the search for *Gct6* alleles shared between CAST and C57 strains that are unique from the GC tumor permissive strains such as SWR. A whole-

locus capture protocol for the *Gct6* region of SWR inbred and SWR.CAST-X congenic line DNA preceded high-throughput NGS sequencing at TJL. Post-informatics reports were provided to our lab that listed all nucleotide variations between SWR and C57 reference, and between CAST and C57 reference genome, with interval boundaries determined by prior phenotype mapping studies for *Gct6*. It was my role to review the reports, categorize, prioritize and independently verify the sequence of variants that were considered promising candidates for *Gct6* identity.

The variant report for CAST vs C57 genome contained four times more variants (4,682) compared to the SWR vs C57 file (1148 entries), likely due to the highly divergent and polymorphic nature of the CAST genome (Figure 1.8). Since the functional testing supported our hypothesis that C57 and CAST strains share a common *Gct6* allele, or have unique alleles in a common gene, my variant analysis proceeded with the shorter variant report (SWR vs C57), after which prioritized variants of interest were investigated in the CAST genome. The majority of the 1148 variants (C57 vs SWR) isolated after multiple filtering steps were intronic 62%) or intergenic (28%) and a minority (≈10 %) of the variants were exonic. The absolute GCT resistance conferred by $Gct6^{C57}$ and $Gct6^{CA}$ in the homozygous SWR.C57-X and SWR.CAST-X congenic lines, and near-absolute resistance conferred by paternal inheritance of either allele, implies a very significant allelic difference exists between the permissive and suppressive strains. For this reason, I originally prioritized the exonic variants for further investigation, since both synonymous and nonsynonymous SNP or Insertion/Deletion (INDEL) variations within exons can have severe effects on protein-coding genes, including: amino acid exchange in the translated protein, premature protein truncation, translation read-through or disrupted transcript splicing. Five exonic, nonsynonymous variants in 5 genes (*BC065397, Esx1, Tmsb15l,*

Tmsb15b1, and Slc25153) were verified by independent Sanger sequencing analysis and subsequently given priority for the allele segregation analysis between suppressive and permissive strain categories.

The *BC065397* gene is not annotated in the human genome and does not have any orthologs listed on the Ensembl genome browser (GRCm38; Ensembl Genome Browser release 75, Flicek *et al.,* 2014). This, together with the lack of allele segregation among GC permissive and suppressive strains, made this gene a low priority in regards to a translatable gene candidate for human JGCTs as a *Gct6* GC tumor suppressor gene.

Esx1 encodes 65 kDa protein which, upon proteolytic cleavage, produces a 45 kDa N-terminal fragment and a 20 kDa C-terminal fragment. The 45 kDa fragment has a homeo-domain where the 20 kDa fragment has a proline-rich domain. The N-terminal fragment is localized at the nucleus while the C-terminal fragment is localized to the cytoplasm. *Esx1* is likely to play a role in placental development, spermatogenesis and may regulate cell cycle progression and transcription during spermatogenesis (Bonaparte *et al.,* 2010). Human and mouse *Esx1* proteins have a very high sequence divergence of 34 % concentrated mostly at the C-terminal domain which plays a role in cell cycle control, suggesting that the region's function is different in primates and rodents (Bonaparte *et al.,* 2010). The allele distribution of this SNP among permissive and suppressive strains is diverse; thus, this gene was given a low priority as being the candidate gene responsible for the *Gct6* tumor suppresser activity.

Thymosin beta proteins (Tmsb) play important roles in the organization of the cytoskeleton and actin polymerization (Ensembl Genome Browser release 75, Flicek et al., 2014). There is a family of thymosin β genes, a member of which, Thymosin β -4 (Tmsb4x), encodes a major actin-sequestering protein that is up-regulated in a wide variety of carcinomas,

implicated in motility of certain cancer cells (Wang et al., 2003). The up-regulation of Tmsb4x promotes disruption of cell-cell adhesion and activates β-catenin signalling; therefore, it is thought to be responsible for the invasive phenotype of human cancers, such as colorectal carcinomas (Chen et al., 2004). Human ortholog of Tmsb15l gene, TMSB15A, has been used as a predictor of chemotherapy response in triple-negative breast cancer (Darb-Esfahani et al., 2012). Since the SNP reported by NGS is located in a highly conserved region of the gene that causes an AA change and most likely affect Tmsb15l function, this variant gene was considered a good candidate for Gct6 GC tumor suppressor activity. However, the lack of allele co-segregation with GC permissive and suppressive strains gave Tmsb15l a lower priority in the list of gene candidates for shared identity with Gct6.

The SLC25 family of mitochondrial carriers (MCs) are the largest of solute carriers that shuttle a variety of metabolites across the inner mitochondrial membrane (Palmieri, 2013). In humans, the SLC25 genes are distributed among all chromosomes; they differ significantly in their size and organization. They may contain 1 to 18 exons and spanning 1,184 to 65,456 bp (Palmieri, 2013). The expression levels of SLC25 family of genes vary in different tissues; they are highly conserved between species and are widespread in eukaryotes, viruses and bacteria (Palmieri, 2013). This superfamily are not only limited to mitochondria, as some members of the family are located in other cell organelles, such as peroxisomes, chloroplasts and mitosomes. MCs are nuclear proteins that are imported from the cytosol into the membrane where they are localized; they bridge between the metabolic reactions in the cytosol and mitochondrial matrix through the catalysis of numerous solutes translocation across the membrane. MCs are involved in many important metabolic pathways: oxidative phosphorylation, synthesis and breakdown of mtDNA, mtRNA and mitochondrial proteins, modulation of the nucleotide and deoxynucleotide

pools in the mitochondrial matrix, necrosis and apoptotic cell death (Palmieri, 2013). The general function of this super family is known, but not all the members of the family have their specific functions determined. This is the case for Slc25a53, presently described as an orphan carrier with no known substrate. The structure of all SLC25 proteins are very similar as they all have common sequences that create tripartite structural features made of a 3-fold repeated signature motif and 6 transmembrane α-helices, 2 in each of the 3 repeats (Palmieri, 2013). The cloning and functional identification of many SLC25 genes has made it possible to identify the genes and their defects responsible for some diseases (e.g. Stanley syndrome and Amish microcephaly) and to understand the molecular basis of the symptoms caused by certain diseases (hyperornithinaemia, hyperammonaemia and homocitrullinuria (HHH) syndrome and type II citrullinemia; Palmieri, 2013). The Slc25a53 allele association between GC permissive and suppressive strains revealed that all the GC permissive strains had an exon-coding deviation, either an insertion or a deletion of a repeat, from the Slc25a53 wild type allele present in C57 and CAST strains. This suggests that Slc25a53 function is compromised in GC permissive strains, leading to the development of GC tumors. Therefore, the Slc25a53 gene ranks highly in the list of candidate genes as having shared identity with the Gct6 tumor suppressor. As an orphan solute carrier, it is uncertain how a Slc25a53 mutation might contribute to GC tumor susceptibility. Given our knowledge that the juvenile-onset GC tumor initiation is sensitive to androgenic stimulation, it could be postulated that Slc25a53 protein activity counters a cellular process initiated by androgens in the first follicular wave. It is part of the future directions for the project to examine Slc25a53 protein localization and quantification in the GC populations of the strains that are tumor-permissive or tumor-suppressive, to determine if the SWR strain represents a spontaneous null strain for this candidate tumor suppressor gene.

5.3 Remaining NGS Identified Variants

Intronic variants can cause abnormal variations in splicing and are proportionally related to multiple human genetic disorders; they contribute to the development of cancer if not corrected by posttranscriptional quality control mechanism (Chen et al., 2006, Skotheim & Nees, 2007, He C., et al., 2009 and Fackenthal et al., 2008). Intergenic (non-coding) variants were originally thought to serve no purpose and were simply junk sequence. However, comparative genomics studies have shown that noncoding DNA sequences can have important functions (Ludwig, 2002). For instance, genome-wide association studies have successfully identified an intergenic loci located on chromosome 8q24 that harbours an important enhancer element associated with multiple cancer types (Grisanzio et al., 2010 and Chiara et al., 2010). Some regions of noncoding DNA are highly conserved even after 300–450 million years of evolution, implying significant functional constraints (Duret et al., 1997 and Müller et al., 2002). Synonymous variants, or silent mutations, are mutations in which the AA sequence is not altered lead to a change of one of the letters in the triplet code that represents a codon, but despite the single base change, the AA that is coded for remains unchanged or similar in biochemical properties. Silent mutations were thought to be of little to no significance. However, recent research suggests that such alterations to the triplet code do effect protein translation efficiency and protein folding and function (Komar, 2007 and Czech et al. 2010).

6.0 Summary and Future Directions

The unique congenic line, SWR.C57-X, developed and tested in this research project replicated the *Gct6* tumor suppressor activity observed for the CAST strain, suggesting *Gct6* suppressor gene targets or alleles are in common between the C57 and CAST genomes. This phenotype information was very useful for our NGS variant filtering strategy of the whole *Gct6* locus between SWR (tumor permissive) and CAST or C57 (tumor suppressive) strains. My participation in the downstream (quaternary) analysis of the NGS variant reports prioritized the non-synonymous exonic variants in 5 unique genes within the mapped *Gct6* locus boundaries. Disruptions (deletions or insertions) in the protein coding region of the gene *Slc25a53* were identified in all GCT-permissive strains, but not in C57 or CAST strains, making *Slc25a53* an excellent candidate as a tumor-suppressor gene. Solute carrier family 25 is a large family of substrate-specific carriers found primarily in the mitochondria, but also other organelles. Slc25a53 is currently an orphan member of the family, with unknown solute specificity or cellular localization. A future direction for this project is to determine the cellular and tissue expression profile for *Slc25a53* gene and protein in GCT permissive (SWR) and resistant (SWR.C57-X or SWR.CAST-X) congenic lines, to follow up on genetic evidence for *Slc25a53* as the candidate gene for *Gct6* identity.

As the identity of *Gct6* is established and confirmed, additional assays will be needed to determine how the *Gct6* activity influences androgen-driven GCT initiation in this model system. Furthermore, the persistence of the dominant paternal parent-of-origin effect with respect to X-linked *Gct* loci, now proven in a third congenic line (SWR.C57-X), suggests bias in X-

inactivation underlies the paternal influence of *Gct4* and *Gct6*. This is an intriguing concept, particularly if this bias is limited to the GC populations of the first follicular wave.

In summary, the trait of spontaneous GC tumorigenesis observed in the SWR strain model represents a unique opportunity to study genetic and endocrine mechanisms related to GCT susceptibility. Given the scarcity of information about specific genes associated with JGCTs, pursuing genetic determinants in this model provides insight for human pediatric JGCT, and may provide mechanistic insights for other androgen-sensitive cancers, such as prostate cancer.

The future directions recommended for continued research progress are:

- Determine if the SWR strain is a spontaneous, viable, *Slc25a53* null strain using gene and protein expression analyses.
- Identify the tissue and cellular expression profile for *Slc25a53* in strains (C57, CAST) with wild-type alleles.
- With accumulating evidence for *Slc25a53* as a candidate GCT susceptibility gene in the SWR model, translational and collaborative exploration in human JGCT specimens (protein expression or sequencing analysis) is warranted.
- The paternal parent-of-origin effect for two physically distant X-linked *Gct* loci suggests X-inactivation bias. This bias could be limited to the first follicular wave of ovarian follicles. This research aim could be pursued using fluorescent reporter strains to isolate GC populations from the first follicular wave, followed by gene expression analysis to calculate the ratios of X-linked maternal vs. paternal alleles in heterozygous female offspring.

7.0 References

Baerwald, A. R., Adams, G. P., & Pierson R. A. (2012). Ovarian antral folliculogenesis during the human menstrual cycle: a review. *Human Reproduction Update*, 18(1), 73-91. doi: 10.1093/humupd/dmr039.

Banerjee, I., Zhang, J., Moore-Morris, T., Lange, S., Shen, T., Dalton, N. D., … Chen, J. (2012). Thymosin beta 4 is dispensable for murine cardiac development and function. *Circulation Research*, 110(3), 456-64. doi: 10.1161/CIRCRESAHA.111.258616.

Beamer, W. G., Hoppe, P. C., & Whitten, W. K. (1985). Spontaneous Malignant Granulosa Cell Tumors in Ovaries of Young SWR Mice. *Cancer Research*, 45(11 pt 2), 5575-5581.

Beamer, W. G., Shultz, K. L., & Tennent, B. J. (1988). Induction of ovarian granulosa cell tumors in SWXJ-9 mice with dehydroepiandrosterone. *Cancer Research*, 48(10), 2788-2792.

Beamer, W. G., Shultz, K. L., & Tennent, B. J. (1988-a). Gene for ovarian granulosa cell tumor susceptibility, Gct, in SWXJ recombinant inbred strains of mice revealed by dehydroepiandrosterone. *Cancer Research*, 48(18), 5092-5095.

Beamer, W. G., Shultz, K. L., Tennent, B. J., Nadeau, J. H., Churchill, G. A., & Eicher, E. M. (1998-b). Multigenic and imprinting control of ovarian granulosa cell tumorigenesis in mice. *Cancer Research*, 58(16), 3694-3699.

Beamer, W. G., Hoppe, P. C., & Whitten, W. K. (1985). Spontaneous malignant granulosa cell tumors in ovaries of young SWR mice. *Cancer Research*, 45(11), 5575-5581.

Beamer, W. G. (1986). Gonadotropin, steroid, and thyroid hormone milieu of young SWR mice bearing spontaneous granulosa cell tumors. *Journal of National Cancer Institute*, 77(5), 1117-1123.

Benayoun, B. A., Georges, A. B., L'Hote, D., Andersson, N., Dipietromaria, A., Todeschini, A. L., ... Veitia, R. A. (2011). Transcription factor FOXL2 protects granu- losa cells from stress and delays cell cycle: role of its regulation by the SIRT1 deacetylase. *Human Molecular Genetics*, 20(9), 1673-1686. doi: 10.1093/hmg/ddr042.

Blake, J. A., Bult, C. J., Eppig, J. T., Kadin, J. A., & Richardson, J. E. (2014). The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Research*, 42(D1), D810-D817. doi: 10.1093/nar/gkt1225.

Bonaparte, E., Moretti, M., Colpi, G. M., Nerva, F., Contalbi, G., Vaccalluzzo, L., ... Miozzo, M. (2010). ESX1 gene expression as a robust marker of residual spermatogenesis in azoospermic men. *Human Reproduction*, 25(6), 1398-1403. doi: 10.1093/humrep/deq074.

Brown, S.M. (2013). *Next-generation DNA sequencing informatics*. Portlan, OR: Cold Spring Harbor Lab. Press

Buller, R. E., Shahin, M. S., Geisler, J. P., Zogg, M., De Young, B. R., & Davis, C. S. (2002). Failure of BRCA1 Dysfunction to Alter Ovarian Cancer Survival. *Clinical Cancer Research*, 8(5), 1196-1202.

Canadian Cancer Statistics 2014. Canadian Cancer Society's Advisory Committee on Cancer Statistics. Toronto, ON: Canadian Cancer Society. http://www.cancer.ca/en/cancer-information/cancer-type/ovarian/statistics/?region=on/ Accessed on Nov 21, 2013.

Cárdenas, H., Herrick, J. R., & Pope, W.F. (2002). Increased ovulation rate in gilts treated with dihydrotestosterone. *Reproduction*, 123(4), 527-533. doi: 10.1530/rep.0.1230527.

Cassidy, S.B., Schwartz, S., Miller, J.L., & Driscoll, D.J. (2012). Prader-Willi syndrome. *Genetics in Medicine*, 14, 10-26. doi:10.1038/gim.0b013e31822bead0.

Cederroth, C. R., Pitetti, J. L., Papaioannou, M. D., & Nef, S. (2007). Genetic Programs that regulate testicular and ovarian development. *Molecular and Cellular Endocrinology*, 265-266, 3-9. doi: 10.1016/j.mce.2006.12.029.

Chen, X., Truong, T. N., Weaver, J., Bove, B. A., Cattie, K., Armstrong, B. A., ... Godwin, A. K. (2006). Intronic Alterations in BRCA1 and BRCA2: Effect on mRNA Splicing Fidelity and Expression. *Human Mutation*, 27(5), 427-435. doi: 10.1002/humu.20319.

Cocquet, J., Pailhoux, E., Jaubert, F., Servel, N., Xia, X., Pannetier, M., ... Veitia, R. A. (2012). Evolution and Expression of FOXL2. *Journal of Medical Genetics*, 39(12), 916-921. doi: 10.1136/jmg.39.12.916.

Czech, A., Fedyunin, I., Zhang, G., Ignatova, Z. (2010). Silent mutations in sight: co-variations in tRNA abundance as a key to unravel consequences of silent mutations. *Mol Biosystem* 6 (10): 1767–72.doi:10.1039/c004796c.

Darb-Esfahani, S., Kronenwett, R., von Minckwitz, G., Denkert, C., Gehrmann, M., Rody, A., ... Loibl, S. (2012). Thymosin beta 15A (TMSB15A) is a predictor of chemotherapy response in triple-negative breast cancer. *British Journal of Cancer*, 107(11), 1892-1900. doi: 10.1038/bjc.2012.475.

Dorward, A. M., Shultz, K. L. Ackert-Bicknell, C. L. Eicher, E. M. Beamer, W. G. (2003). High-Resolution Genetic Map of X-Linked Juvenile-Type Granulosa Cell Tumor Susceptibility Genes in Mouse. *Cancer Research* 63, 8197-820.

Dorward, A. M., Shultz, K. L., Horton, L. G., Li, R., Churchill, G. A., & Beamer, W.G. (2005). Distal Chr 4 harbors a genetic locus (Gct1) fundamental for spontaneous ovarian granulosa cell tumorigenesis in a mouse model. *Cancer Research* 65(4), 1259-1264. doi: 10.1158/0008-5472.CAN-04-2992.

Dorward, A. M., Yaskowiak, E. S., Smith, K. N., Stanford, K. R., Shultz, K. L., & Beamer, W. G. (2013). Chromosome X loci and spontaneous granulosa cell tumor development in SWR mice. Epigenetics and epistasis at work for an ovarian phenotype. *Epigenetics*, 8(2), 184-191. doi: 10.4161/epi.23399.

Drummond, A. E., & Fuller, P. J. (2012). Activin and inhibin, estrogens and NFκB, play roles in ovarian tumourigenesis is there crosstalk? *Molecular and Cellular Endocrinology,* 359(1-2), 85-91. doi: 10.1016/j.mce.2011.07.033.

Duffin, K., Bayne, R. A., Childs, A. J., Collins, C., & Anderson, R. A. (2009). The forkhead transcription factor FOXL2 is expressed in somatic cells of the human ovary prior to follicle formation. *Molecular Human Reproduction,* 15(12), 771-777. doi: 10.1093/molehr/gap065.

Duret, L., & Bucher, P. (1997). Searching for regulatory elements in human noncoding sequences. *Current Opinion in Structural Biology,* 7(3), 399-406. doi: 10.1016/S0959-440X(97)80058-9.

Erickson, B. K., Conner, M. G., & Landen, C. N. Jr. (2013) The role of the fallopian tube in the origin of ovarian cancer. *American Journal of Obstetrics Gynecology,* 209(5), 409-414. doi: 10.1016/j.ajog.2013.04.019.

Fackenthal, J. D., & Godley, L. A. (2008). Aberrant RNA splicing and its functional consequences in cancer cells. *Disease Models and Mechanisms,* 1(1), 37-42. doi: 10.1242/dmm.000331.

Fleming, N. I., Knower, K. C., Lazarus, K. A., Fuller, P. J., Simpson, E. R., & Clyne, C. D. (2011). Aromatase is a direct target of FOXL2:C134W in granulosa cell tumors via a single highly conserved binding site in the ovarian specific promoter. *PLoS One.* 5(12), e14389. doi: 10.1371/journal.pone.0014389.

Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., ... Searle S. M. J. (2014). Ensembl 2014. *Nucleic Acids Research,* 42(D1) D749-D755. doi:10.1093/nar/gkt1196.

Fox, H. (1985). Sex cord-stromal tumors of the ovary. *The Journal of Pathology,* 145(2), 127-148.

Garcia-Ortiz, J. E., Pelosi, E., Omari, S., Nedorezov, T., Piao, Y., Karmazin, J., ... Ottolenghi, C. (2009). Foxl2 functions in sex determination and histogenesis throughout mouse ovary development. *BMC Developmental Biology,* 9(36). doi: 10.1186/1471-213X-9-36.

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., ... Bryant, S.H. (2010) The NCBI BioSystems database. *Nucleic Acids Research,* 38(suppl 1), D492-496. doi:10.1093/nar/gkp858.

Gilbert, Scott F. (2003). *Developmental Biology*, (7 ed). USA: Elsevier Science under Auspices of Society for Developmental Biology. Chapter 11, pp 346.

Goudet, G., Belin, F., Bezard, J., & Gerard, N. (1999). Intrafollicular content of luteinizing hormone receptor, alpha-inhibin, and aromatase in relation to follicular growth, estrous cycle stage, and oocyte competence for in vitro maturation in the mare. *Biology of Reproduction,* 60(5), 1120-1127. doi: 10.1095/biolreprod60.5.1120.

Greenacre, C. B. (2004). Spontaneous tumors of small mammals. *The Veterinary Clinics of North America. Exotic Animal Practice.* 7(3), 627–651. doi: 10.1016/j.cvex.2004.04.009.

Grisanzio, C., & Freedman, M. L. (2010). Chromosome 8q24−Associated Cancers and MYC *Genes & Cancer.* 1(6), 555-559. doi: 10.1177/1947601910381380.

Hirshfield, A. N., & DeSanti, A. M. (1995). Patterns of ovarian cell proliferation in rats during the embryonic period and the first three weeks postpartum. *Biology of Reproduction*, 53(5), 1208-1221. doi: 10.1095/biolreprod53.5.1208.

Kim, J. H., Yoon, S., Park, M., Park, H. O., Ko, J. J., Lee, K., & Bae, J. (2011). Differential apoptotic activities of wild-type FOXL2 and the adult-type granulosa cell tumor-associated mutant FOXL2 (C134W). *Oncogene*, 30(14), 1653–1663. doi: 10.1038/onc.2010.541.

Köbel, M., Gilks, C. B., & Huntsman, D. G. (2009). Adult-type granulosa cell tumors and FOXL2 mutation. *Cancer Research*, 69(24), 9160-9162. doi: 10.1158/0008-5472.CAN-09-2669.

Komar A. A. (2007). Silent SNPs: impact on gene function and phenotype. *Pharmacogenomics*, 8(8), 1075–80. doi:10.2217/14622416.8.8.1075.

Kuhn, E., Ayhan, A., Shih, I., Seidman, J. D., & Kurman, R. J. (2013). Ovarian brenner tumour: A morphologic and immunohistochemical analysis suggesting an origin from fallopian tube epithelium. *European Journal of Cancer*, 49(18), 3839-3849. doi: 10.1016/j.ejca.2013.08.011.

Kurman, R. J., & Shih, I. M. (2010). The Origin and Pathogenesis of Epithelial Ovarian Cancer: A Proposed Unifying Theory. *American Journal of Surgical Pathology*, 34(3), 433-443. doi: 10.1097/PAS.0b013e3181cf3d79.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., ... Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics,* 23(21), 2947-2948. doi: 10.1093/bioinformatics/btm404.

Leung, P. C. K., & Adashi, E. Y. (2004). *The Ovary*, second edition, United States of America: Elservier Academic Press.

Ludwig, M. Z. (2002). Functional evolution of noncoding DNA. *Current Opinion in Genetics & Development,* 12(6), 634–639. doi: 10.1016/S0959-437X(02)00355-6.

Lynch, H. T., Casey, M. J., Snyder, C. L., Bewtra, C., Lynch, J. F., Butts, M., & Godwin, A. K. (2009). Hereditary ovarian carcinoma: hererogeneity,molecular genetics,pathology, and management. *Molecular Oncology*, 3(2), 97-137. doi: 10.1016/j.molonc.2009.02.004.

Merzouk, S., Deuve, J. L., Dubois, A., Navarro, P., Avner, P., Morey, C., (2014). Lineage-specific regulation of imprinted X inactivation in extraembryonic endoderm stem cells. *Epigenetics Chromatin*, 7(11). doi:10.1186/1756-8935-7-11

Müller, F., Blader, P., & Strähle, U. (2002). Search for enhancers: teleost models in comparative genomic and transgenic analysis of cis regulatory elements. *Bioessays*, 24(6), 564-572. doi: 10.1002/bies.10096.

Myers, M., & Pangas, S.A. (2010). Regulatory roles of transforming growth factor beta family members in folliculogenesis. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, 2(1), 117-125. doi: 10.1002/wsbm.21.

Nef, S., & Vassalli, J. D. (2009). Complementary pathways in mammalian female sex determination. *Journal of Biology*, 8(8), 74. doi: 10.1186/jbiol173.

O'Donovan, P.J., & Livingston, D.M. (2010). BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis*, 31(6), 961-967. Doi: 10.1093/carcin/bgq069.

Online Mendelian Inheritance in Man, OMIM®. Johns Hopkins University, Baltimore, MD. MIM Number: {110100}: {08/17/2011}: World Wide Web URL:http://www.omim.org/entry/110100

O'Shea, J. D. (1981) Structure-function relationships in the wall of the ovarian follicle. *Australian Journal of Biological Sciences*, 34(4), 379-394. doi: 10.1071/BI9810379.

Palmieri, F. (2013). The mitochondrial transporter family SLC25: Identification, properties and physiopathology. *Molecular Aspects of Medicine*, 34(2-3), 465-484. doi: 10.1016/j.mam.2012.05.005.

Petkov, P. M., Ding, Y., Cassell, M. A., Zhang, W., Wagner, G., Sargent, E. E., ... Wiles, M. V. (2004). An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Research*, 14(9), 1806-1811. doi: 10.1101/gr.2825804.

Piprek, R. P. (2009). Molecular mechanisms underlying female sex determination - antagonism between female and male pathway. *Poland: Folia Biologica (Kraków)*, 57(3-4), 105-113. doi: 10.3409/fb57_3-4.105-113.

Prat, J., Ribé, A., & Gallardo, A. (2005). Hereditary Ovarian Cancer. *Human Pathology*, 36(8), 861-870. doi: 10.1016/j.humpath.2005.06.006.

Preacher, K. J., & Briggs, N. E. (2001, May). Calculation for Fisher's Exact Test: An interactive calculation tool for Fisher's exact probability test for 2 x 2 tables [Computer software]. Available from http://quantpsy.org.

Quinn, P., & Horstman, F. C. (1998). Is the mouse a good model for the human with respect to the development of the preimplantation embryo in vitro? *Human Reproduction* 13(suppl 4), 173-183. doi:10.1093/humrep/13.suppl_4.173.

Rebstock, L. E., Leufflen, L., Leroux, L., Harter, V., Verhaeghe, L. J., & Marchal, F. (2014). Tumeur De La Granulosa De l'ovaire: Étude Rétrospective à Propos De 17 cas. *Gynécologie Obstétrique & Fertilité*. 42(5), 331-333. doi: 10.1016/j.gyobfe.2013.08.010.

Risbridger, G. P., Schmitt, J. F., & Robertson, J. F. (2001). Activins and Inhibins in Endocrine and Other Tumors. *Endocrine Reviews*, 22(6), 836-858. doi: 10.1210/edrv.22.6.0450.

Sen, A., Prizant, H., Light, A., Biswas, A., Hayes, E., Lee, H. J., ... Hammes, S. R. (2014). Androgens regulate ovarian follicular development by increasing follicle stimulating hormone receptor and microRNA-125b expression. *Proceedings of the National Academy of Sciences of the United States of America* 111(8), 3008-3013. doi: 10.1073/pnas.1318978111.

Silver, L. M. (2004) *Mouse Genetics: Concepts and Applications.* Retrieved from http://www.informatics.jax.org/silverbook/index.shtml

Simon, M.M., Mallon, A.M., Howell, G.R. & Reinholdt, L.G. (2012). High throughput sequencing approaches to mutation discovery in the mouse. *Mammalian Genome*, 23, 499-513. doi: 10.1007/s00335-012-9424-0

Shah, S. P., Köbel, M., Senz, J., Morin R. D., Clarke B. A., Wiegand, K. C., Leung G., … Huntsman D.G. (2009). Mutation of *FOXL2* in Granulosa-Cell Tumors of the Ovary. *The new England journal of medicine.* 360(26), 2719-2729. doi: 10.1056/NEJMoa0902542.

Skotheim, R. I., & Nees, M. (2007). Alternative splicing in cancer: Noise, functional, or systematic? *The International Journal of Biochemistry & Cell Biology*, 39(7-8), 1432-1449. doi: 10.1016/j.biocel.2007.02.016.

Smith K. N., Halfyard, S. J., Yaskowiak, E. S., Shultz, K. L., Beamer, W. G., & Dorward, A. M. (2012). Fine map of the Gct1 spontaneous ovarian granulosa cell tumor locus. *Mammalian Genome*, 24(1-2), 63-71. doi: 10.1007/s00335-012-9439-6.

Smith, K. N., Halfyard, S. J., Yaskowiak, E. S., Shultz K. L., Beamer, W. G., Dorward, A. M. (2013). Fine map of the Gct1 spontaneous ovarian granulosa. *Mammalian Genome*, 24(63–71) Doi 10.1007/s00335-012-9439-6.

Serov, S. F., Sculty, R. E., & Sobin, L. H. (1973) Histological classification of ovarian tumours. *Histological typing of ovarian tumors.* Geneva: *World Health Organization*. pp 42. Switzerland: Roto-Sadag S.A., Geneva.

Tennent, B. J., Shultz, K. L. and Beamer, W. G. (1993) Genetic Susceptibility for C19 Androgen Induction of Ovarian Granulosa Cell Tumorigenesis in SWXJ Strains of Mice. *Cancer Research, 53,1059-1063.*

Tavassoli F. A., Mooney E. Gersell D. J., McCluggage W. G., Konishi I.,Fujii S., …(2003) Sex cord-stromal tumours. Tavassoli, F. A., & Devilee, P. (ED). *Pathology and genetics of tumours of the breast and female genital organs*. Chapter 2. pp 146, France: International Agency for Research on cancer 69008 Lyon.

Thayer, K. A., & Foster, P. M. (2007). Workgroup Report: National Toxicology Program Workshop on Hormonally Induced Reproductive Tumors Relevance of Rodent Bioassays. *Environmental Health Perspectives,* 115(9), 1351-1356. doi: 10.1289/ehp.10135.

Ud Din, N., & Kayani, N. (2014). Recurrence of Adult Granulosa Cell Tumor of the Ovary: Experience at a Tertiary Care Center. *Annals of Diagnostic Pathology*, 18(3), 125-128. doi: 10.1016/j.anndiagpath.2014.02.002.

Vendola, K., Zhou, J., Wang, J., Famuyiwa, O. A., Bievre, M., & Bondy, C. A. (1999). Androgens promote oocyte insulin-like growth factor I expression and initiation of follicle development in the primate ovary." *Biology of Reproduction*, 61(2), 353-357. doi: 10.1095/biolreprod61.2.353.

Verlhac, M., & Villeneuve, A. (2010). *Oogenesis: The universal process*. Hoboken, NJ, USA: Wiley.

Wang, W. S., Chen, P. M., Hsiao, H. L., Ju, S. Y., & Su, Y. (2003). Overexpression of the thymosin beta-4 gene is associated with malignant progression of SW480 colon cancer cells. Oncogene, 22(21), 3297-3306. doi:10.1038/sj.onc.1206404.

Wang, W. S., Chen, P. M., Hsiao, H. L., Wang, H. S., Liang, W. Y., & Su, Y. (2004). Overexpression of the thymosin beta-4 gene is associated with increased invasion of SW480 colon carcinoma cells and the distant metastasis of human colorectal carcinoma. *Oncogene*, 23(39), 6666-6671. doi:10.1038/sj.onc.1207888.

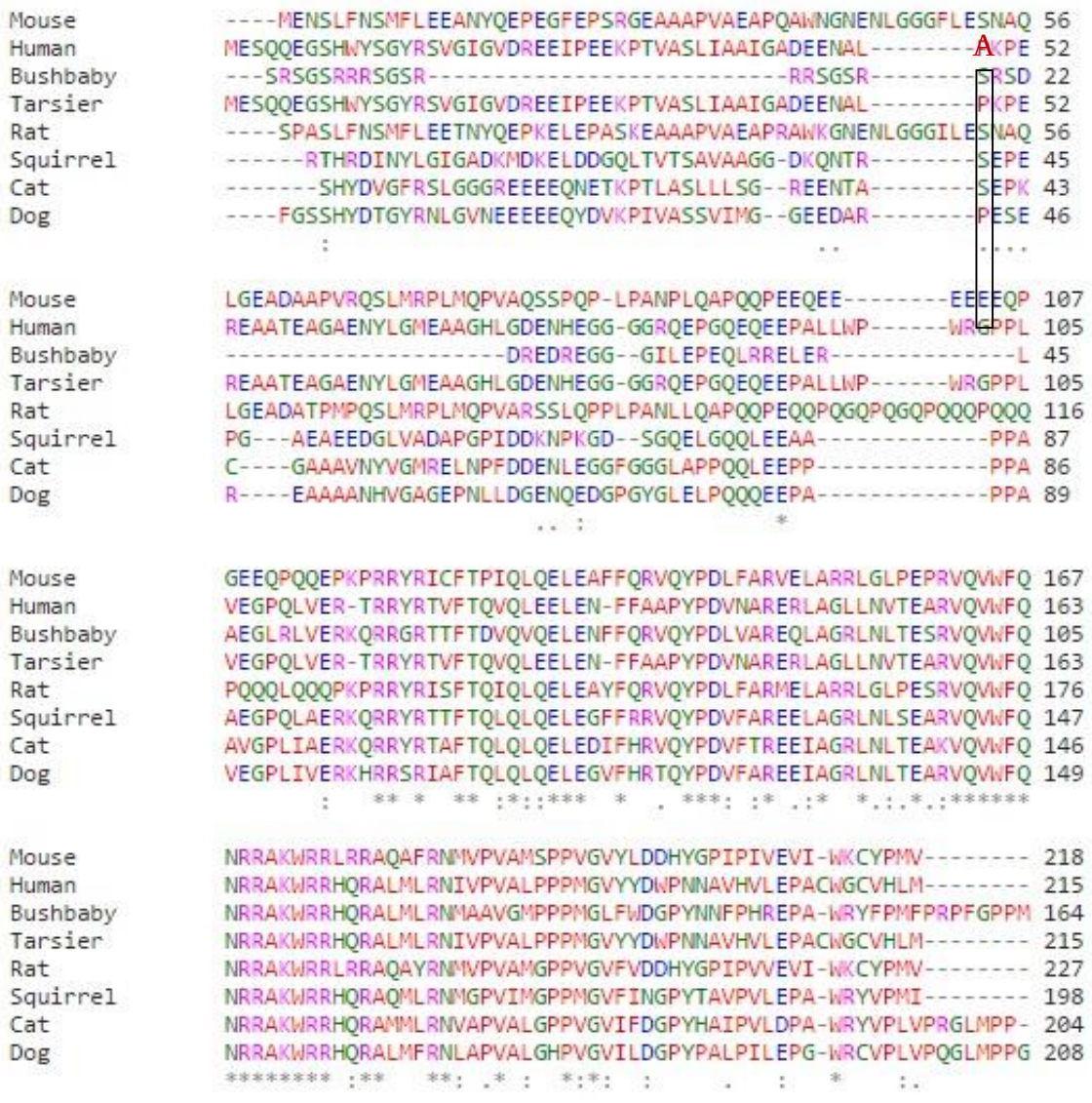
Wheater, P. R., Burkitt, H. G., & Daniels, V. G. (1979). *Functional Histology-a text and colour atlas.* London. Churchill Livingstone.

World Health Organization. Feb 24, 2014.

Young, R. H., Dickersin, G. R., & Scully, R. E. (1984). Juvenile granulosa cell tumor of the ovary. A clinicopathological analysis of 125 cases. *The American Journal of Surgical Pathology*, 8(8), 575-596.

Zheng, W., Zhang, H., Gorre, N., Risal, S., Shen, Y., & Liu, K. (2014). Two classes of ovarian primordial follicles exhibit distinct developmental dynamics and physiological functions . *Human Molecular Genetics*, 23(4), 920-928. doi: 10.1093/hmg/ddt486.

Appendix A *Esx1* AA alignment conservation among multiple species



Protein sequence conservation of *Esx1* gene was compared between multiple species. The location of the variant detected by NGS is shown in black box and the consequence of the variant is shown by the letter above the box. Two species with highest target % ID were chosen from each species set (Primates, Rodents, Laurasiatheria, Sauropsida and Fish) for protein sequence alignment. Target specie = Mouse, target % ID for: Human = 21, Bushbaby = 40, Tarsier = 31, Rat = 80, Squirrel = 39, Cat = 38, Dog = 37.

*: identities; -: deletion; gaps: nucleotide changes. RED: Small and hydrophobic, BLUE: Acidic, MAGENTA: Basic, GREEN: Hydroxyl, sulfhydryl and amine, Grey: Unusual AA

Appendix B *Tmsb15l* AA alignment conservation among multiple species

Protein sequence conservation of *Tmsb15l* gene was compared between multiple species. The location of the variant detected by NGS is shown in black box and the consequence of the variant is shown by the letter above the box. Two species with the highest target % ID were chosen from each species set (Primates, Rodents, Laurasiatheria, Sauropsida and Fish) for protein sequence alignment. Target specie = Mouse, target % ID for: Human = 78, Mouse lemur = 85, Orangutan = 78, Rat = 89, Rabbit = 80, Hedgehog = 88, Sheep = 80, Turkey = 77, Zebra finch = 76, Platy fish = 71, Amazon molly = 71.

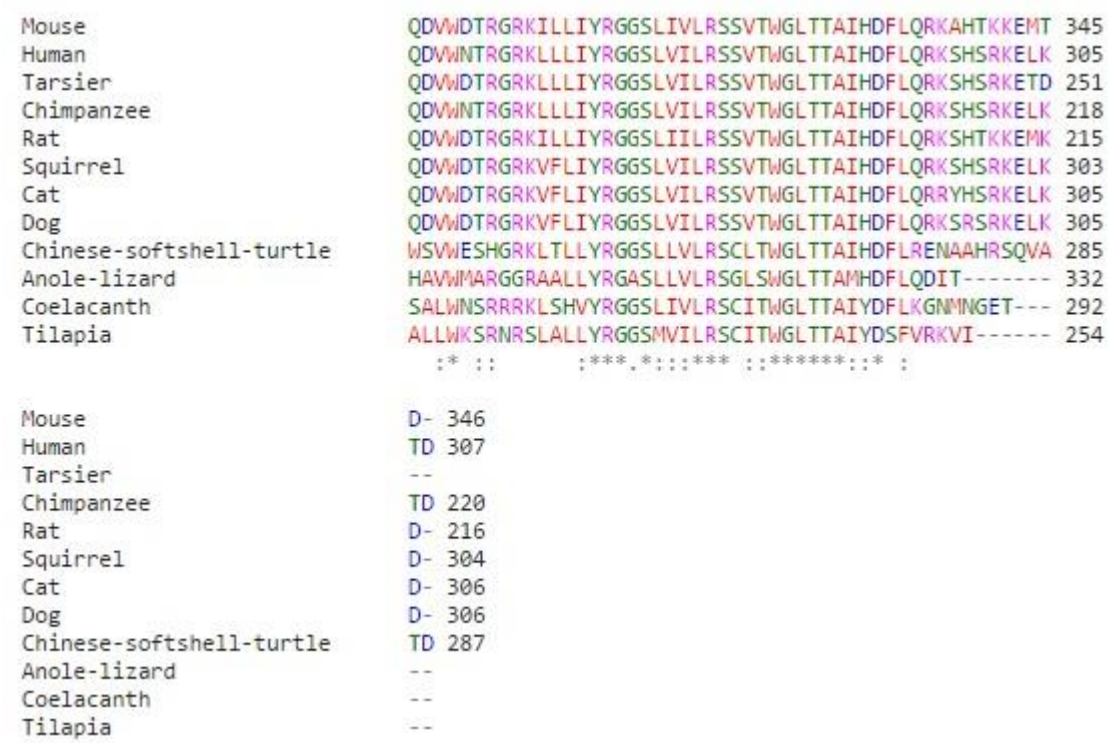
*: identities; -: deletion; gaps: nucleotide changes. RED: Small and hydrophobic, BLUE: Acidic, MAGENTA: Basic, GREEN: Hydroxyl, sulfhydryl and amine, Grey: Unusual AA

Appendix C *Slc25a53* AA alignment conservation among multiple species

```
Mouse                          PVGPHSLGQR-WTAGLMSGVVEAVALSPFERVQNVLQDARKQACFPSTFS 196
Human                          PVGPHTLGHR-WAAGLMSGVVEAVALSPFERVQNVLQDGRKQARFPSTFS 156
Tarsier                        PVGPHSLGHR-WAAGLMSGVVEAVALSPFERVQNVLQDGRKQARFPSTFS 102
Chimpanzee                     PVGPHTLGHR-WAAGLMSGLVEAVALSPFERVQNVLQDGRKQARFPSTFS 69
Rat                            PVGPHSLGQR-WTAGLMSGVVEAVALSPFERVQNVLQDARKQACFPSTFS 66
Squirrel                       PDGPHSLGHR-WTAGLMSGVVEAVALSPFERVQNVLQDGRKQARFPSTFS 154
Cat                            PIGPHSLGHR-WTAGLMSGVVEAVALSPFERVQNVLQDGRKQARFPSTFS 156
Dog                            PVGPHSLGHR-WAAGLMSGVVEAVALSPFERVQNVLQDGRKQARFPSTFS 156
Chinese-softshell-turtle       PSEPCTLGER-WAAGFLSGSVEALVLNPFERVQNVLQDGRKDARFPNIRS 136
Anole-lizard                   PSGSHSWGQQ-ATAGLLSGLSEGLVLGPFERVQNVLQDGQKARRFPTARS 189
Coelacanth                     TPDSYSLKNR-CVAGFLSGTMEAVVLCPFERVQNILQDSRKDSRFPNIRS 147
Tilapia                        SQNVISAKALPALAGFGAGVVEAMVFTPFERVQNVLQNSQNDSRLPSLKS 116
                               .   :      **: :*  *.:.: *******:**:.::   :*.  *

Mouse                          ILKEFNSYG-LWGRLSLGYYRGFWPVLVRNSLGSALYFSFKDPIQDGLAR 245
Human                          ILKEFNSYG-LWGRLSLGYYRGFWPVLARNSLGSALYFSFKDPIQDGLAE 205
Tarsier                        ILKEFNSYG-LWGRLSLGYYRGFWPVLARNSLGSALYFSFKDPIQDGLAE 151
Chimpanzee                     ILKEFNSYG-LWGRLSLGYYRGFWPVLARNSLGSALYFSFKDPIQDGLAE 118
Rat                            ILKEFNSYG-LWGRLSLGYYRGFWPVLVRNSLGSALYFSFKDPIQDGLAR 115
Squirrel                       ILKEFNSYG-LWGRLSLGYYRGFWPVLVRNSLGSALYFSFKDPIQDGLAE 203
Cat                            ILKEFNSYG-LWGRLSLGYYRGFWPVLLRNSLGSALYFSFKDPIQDGLAE 205
Dog                            ILKEFNSYG-LWGRLSLGYYRGFWPVLLRNSLGSALYFSFKDPIQDGLAE 205
Chinese-softshell-turtle       ILQEFNSYG-PKERLVVGYYRGLSPVLLRNSLGSALYFSFKDPLRDSLAA 185
Anole-lizard                   ILEEFRSYGGLRERLALGYYRGLGLILLRNGLGSALYFSAKDPLRDRLRE 239
Coelacanth                     ILQEFNSYR-EKEKLTLGYYRGFLPVLLRNGMGSALYFAFKDPIKNSLSE 196
Tilapia                        VLLRLKAEG-----MALGYYRAFLPIAARNAFGSSLYFGLKGPVCVAVAE 161
                               :* .:.:       : :****.:  :  **.:**:***. *.*:   :

Mouse                          QGLPHWVPALVSGSVNGTITCLILYPLMVLVANMQSHIGWQRMPSLWASA 295
Human                          QGLPHWVPALVSGSVNGTITCLVLYPLIVLVANMQSHIGWQNMPSLWASA 255
Tarsier                        QGLPHWVPALVSGSVNGTITCLVLYPLIVLVANMQSHIGWQSMPSLWASA 201
Chimpanzee                     QGLPHWVPALVSGSVNGTITCLVLYPLIVLVANMQSHIGWQNMPSLWASA 168
Rat                            QGLPHWVPALVSGSVNGTITCLILYPLIVLVANMQSHIGWQRMPSLWASA 165
Squirrel                       QGLPHWVPALVSGSVNGTITCLVLYPLIVLVANMQSHIGWQRMPSLWASA 253
Cat                            QGLPHWVPALVSGSVNGTITCLVLYPLIVLVANMQSHIGWQRMPSLWASA 255
Dog                            QGLPHWVPALVSGSVNGTITCLVLYPLIVLVANMQSHIGWQSMPSLWASA 255
Chinese-softshell-turtle       RGLPSCLATLVSGSVNGTVTCLLLYPLSVLVANIQSQVGGRDLLSLRESA 235
Anole-lizard                   KGLPARLPALLSGCITGTAIGLLLYPLGVLIANIQAQVGSGEALGLRGAL 289
Coelacanth                     KGLTGSIPAFISGSVNGIVVCLTLYPLSVLIVTMQSQVKG-EMLNIPDSF 245
Tilapia                        QGLSPMASSFISGTVISMAVSLTLYPLSVLVANMQAQVGG-DVKGVIACW 210
                               :**.   .:::** : .    * **** **:..:*:::      .:  .
```



Protein sequence conservation of *Slc25a53* gene was compared between multiple species. The location of the variant detected by NGS is shown in black box and the consequence of the variant is shown by the letter above the box. Two species with the highest target % ID were chosen from each species set (Primates, Rodents, Laurasiatheria, Sauropsida and Fish) for protein sequence alignment. Target specie = Mouse, target % ID for: Human = 89, Tarsier = 92, Chimpanzee = 90, Rat = 98, Squirrel = 90, Cat = 89, Dog = 89, Chinese soft shell turtle = 61, Anole lizard = 46, Coelacanth = 54, Tilapia = 46.

*: identities; -: deletion; gaps: nucleotide changes. RED: Small and hydrophobic, BLUE: Acidic, MAGENTA: Basic, GREEN: Hydroxyl, sulfhydryl and amine, Grey: Unusual AA

Appendix D Copyright permissions

(a)

Permissions Department
Contracts, Copyrights and Permissions
Segment Administration

PERMISSION LICENSE: PRINT REPUBLICATION

Request ID/Invoice Number: ZOH18545

Date: May 05, 2014

To: Zoha Rabie
Grad Student
Memorial University of Newfoundland
St. John's NL
CANADA
"Licensee"

McGraw-Hill Material

Author: A. Mescher
Title: Junqueira's Basic Histology: Text and Atlas
ISBN 0071780335
Description of material: Two Figures: 22-2 and 22-3 [only]

Fee: Waived

Licensee Work:

Author: Zoha Rabie
Title: 'The Chr X-linked Gct6 locus: A Granulosa cell tumor suppressor in mice'
Publisher: Memorial University of Newfoundland
Publication Date: 2014
Distribution Territory: Canada
Languages: English

Permission for the use described above is granted under the following conditions:

1. The permission fee of $0.00 must be received by McGraw-Hill Education and MUST BE ACCOMPANIED BY A SIGNED COPY OF THIS AGREEMENT. .

2. No adaptations, deletions, or changes will be made in the material without the prior written consent of McGraw-Hill Education.

ZOH18545

2 Penn Plaza, 20th Floor | New York, NY 10121-2298 | phone: (646) 766.2574 | mhe-permissions@mheducation.com

Permissions Department
Contracts, Copyrights and Permissions
Segment Administration

3. This permission is non-exclusive, non-transferable, and limited to the use specified herein. McGraw-Hill Education expressly reserves all rights in this material.

4. A credit line must be printed on the first page on which the material appears. This credit must include the author, title, copyright date, and publisher, and indicate that the material is reproduced with permission of McGraw-Hill Education.

5. This permission does not allow the use of any material, including but not limited to photographs, charts, and other illustrations, which appears in a McGraw-Hill Education's work copyrighted in or credited to the name of any person or entity other than McGraw-Hill Education. Should you desire permission to use such material, you must seek permission directly from the owner of that material, and if you use such material you agree to indemnify McGraw-Hill Education against any claim from the owners of that material.

Please sign both copies and return one to McGraw-Hill Education, Permissions Department, 2 Penn Plaza, 20th Floor, New York, NY 10121.

For McGraw-Hill:

Cynthia Aguilera

Cynthia Aguilera
Permissions Department
McGraw-Hill Education

For Licensee:

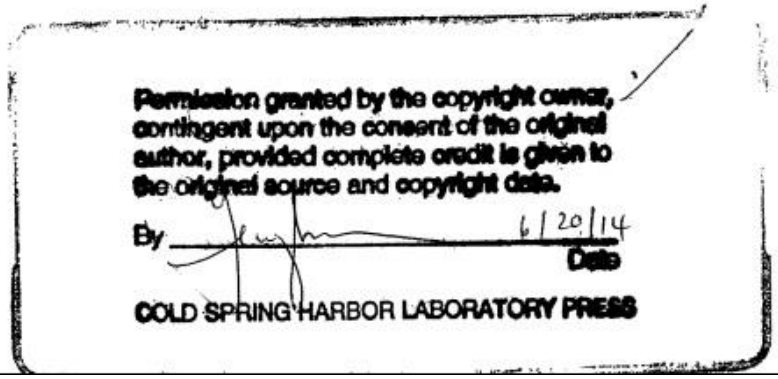
Name_____

Title _____

ZOH18545

(b)

Name: zoha rabie
CompanyInstitution: Memorial University of Newfoundland
Library Address: 230 Elizabeth Ave St. John's,
City: St. John's
State (US and Canada): Newfoundland and Labrador
Country: Canada
Zip: A1B0H2
Title: Mrs
Lab/Department: Human genetics
Phone: 7097495022
Email: z.rabie@mun.ca
Title of Publication: Zoha Rabie, The Chr X-linked Gct6 locus: A granulosa cell tumor suppressor in mice
Authors/Editors: Zoha Rabie
Date of Publication: Oct 2014
Publisher: MSc Thesis, 2014
Title of CSHLP Journal/Book: Genome Research
Title of Article/Chapter: An efficient SNP system for mouse genome scanning and elucidating strain relationships.
CSHL Authors/Editors: Petkov PM1, Ding Y, Cassell MA, Zhang W, Wagner G, Sargent EE, Asquith S, Crew V, Johnson KA, Robinson P, Scott VE, Wiles MV.
Page Numbers: 1806-1811
Figure Numbers: 3
Figure Page Numbers: 1810
Copyright Date: 2004
Additional comments: For use in Master's Thesis.



Appendix E NGS data sample sheet

Chromosome	Position	Reference_Base	Consensus_Base	SNP quality	Depth	Allele frequency (non ref allele)	mapping quality					Exon	Exon ID	strant	Intron	Intergenic	function class	transcript(s)	Gene symbol	strand	Ensembl version
chrX	3,062,340	A	T	18.8	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=33	FQ=-33	---	No	:Nearest Exon distance from SNP=14455		No	Yes	Intergenic	ENSMUST00000121982	Gm14332	-	mus_musculus_core_67_37
chrX	3,062,395	C	A	17.8	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=33	FQ=-33	---	No	:Nearest Exon distance from SNP=14400		No	Yes	Intergenic	ENSMUST00000121982	Gm14332	-	mus_musculus_core_67_37
chrX	3,213,575	G	T	5.29	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=17	FQ=-33	---	No	ENSMUST00000119250 :Nearest Exon distance from SNP=8351		No	Yes	Intergenic	ENSMUST00000119250	Gm14352	-	mus_musculus_core_67_37
chrX	3,304,401	G	A	8.44	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=22	FQ=-33	---	No	:Nearest Exon distance from SNP=17579		No	Yes	Intergenic	ENSMUST00000120670	Gm14360	+	mus_musculus_core_67_37
chrX	3,570,611	A	G	3.98	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,2,0	MQ=18	FQ=-33	---	No	ENSMUST00000116643 :Nearest Exon distance from SNP=22642		No	Yes	Intergenic	ENSMUST00000116643	Gm14354	-	mus_musculus_core_67_37
chrX	3,914,043	C	G	27.8	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=44	FQ=-33	---	No	:Nearest Exon distance from SNP=5073		No	Yes	Intergenic	ENSMUST00000105015	Gm14347	+	mus_musculus_core_67_37
chrX	3,914,072	G	A	6.02	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=18	FQ=-33	---	No	:Nearest Exon distance from SNP=5102		No	Yes	Intergenic	ENSMUST00000105015	Gm14347	+	mus_musculus_core_67_37
chrX	3,918,422	T	C	38.8	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=57	FQ=-33	---	No	:Nearest Exon distance from SNP=9452		No	Yes	Intergenic	ENSMUST00000105015	Gm14347	+	mus_musculus_core_67_37
chrX	3,918,429	G	A	21.8	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=57	FQ=-33	---	No	:Nearest Exon distance from SNP=9459		No	Yes	Intergenic	ENSMUST00000105015	Gm14347	+	mus_musculus_core_67_37
chrX	3,981,081	G	T	24	DP=3	AF1=1	CI95=0.5,1	DP4=0,0,1,2	MQ=27	FQ=-36	---	No	:Nearest Exon distance from SNP=12533		No	Yes	Intergenic	ENSMUST00000105014	Gm10922	-	mus_musculus_core_67_37
chrX	4,137,840	C	T	13.2	DP=4	AF1=0.5016	CI95=0.5,0.5	DP4=0,1,1,2	MQ=52	FQ=-6.21	PV4=1,0.013,0.48,0.4	No	ENSMUST00000116187 :Nearest Exon distance from SNP=21907		No	Yes	Intergenic	ENSMUST00000116187	Gm9427	+	mus_musculus_core_67_37
chrX	4,137,861	C	A	57	DP=4	AF1=0.5003	CI95=0.5,0.5	DP4=0,1,1,2	MQ=52	FQ=4.13	PV4=1,0.33,0.48,	No	ENSMUST00000116187 :Nearest Exon distance from SNP=21886		No	Yes	Intergenic	ENSMUST00000116187	Gm9427	+	mus_musculus_core_67_37
chrX	4,137,873	T	A	15.1	DP=4	AF1=0.5004	CI95=0.5,0.5	DP4=0,1,1,2	MQ=52	FQ=3.5	PV4=1,0.019,0.48,0.3	No	ENSMUST00000116187 :Nearest Exon distance from SNP=21874		No	Yes	Intergenic	ENSMUST00000116187	Gm9427	+	mus_musculus_core_67_37
chrX	4,243,282	C	T	29.8	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=60	FQ=-33	---	No	ENSMUST00000151048 :Nearest Exon distance from SNP=14899		No	Yes	Intergenic	ENSMUST00000151048	Gm4725	+	mus_musculus_core_67_37
chrX	4,244,058	T	C	7.59	DP=2	AF1=1	CI95=0.5,1	DP4=0,0,1,1	MQ=19	FQ=-33	---	No	ENSMUST00000151048 :Nearest Exon distance from SNP=14123		No	Yes	Intergenic	ENSMUST00000151048	Gm4725	+	mus_musculus_core_67_37