October 2018

MYC Distant Enhancers Underlie Ovarian Cancer Susceptibility at the 8q24.21 Locus

Anxhela Gjyshi Gustafson
University of South Florida, agjyshi@mail.usf.edu

MYC Distant Enhancers Underlie Ovarian Cancer Susceptibility at the 8q24.21 Locus

by

Anxhela Gjyshi Gustafson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Cell Biology, Microbiology and Molecular Biology
College of Arts and Sciences
University of South Florida

Major Professor: Alvaro N.A. Monteiro, Ph.D.
W. Douglas Cress, Ph.D.
Jiandong Chen, Ph.D.
Kathleen M. Egan, Ph.D.
Kevin Brown, Ph.D.

Date of Approval:
July 20, 2018

Keywords: Ovarian cancer, GWAS, SNPs, 8q24 locus

DEDICATION

I dedicate this work to:

To my loving and supporting parents, Gjergji and Gitjana Gjyshi, who have sacrificed so much to give me opportunities they were deprived of in their life and careers. You have instilled in me the importance of education and have encouraged me to follow my dreams. Thank you! You have always been a constant source of strength and inspiration and have taught me that with hard work anything is possible.

To my wonderful sister, Marlena Gjyshi-Kodra, who has encouraged and supported me through everything. Thank you for always being there for me. And to my niece and nephew, Megi and Eges Kodra, whose love and smiling faces helped me make it through the disappointing days of failed experiments in the lab

Last but not least, I dedicate this work to my dear husband, Alan P. Gustafson Jr., for his love, understanding, and support – both moral and monetary! Thank you for your patience and encouragement throughout this process. To my daughter, Riley, who always lights up my day when everything else fails - You have made my life complete; I am a better, stronger person because of you.

ACKNOWLEDGMENTS

First, I would like to thank my mentor, Dr. Alvaro Monteiro, for giving me the opportunity to go through this journey under his wing. It has been a wonderful learning experience that transcends science. He is truly one of a kind mentor who can give you a comprehensive summary or information on any kind of topic; including, but not limited to: art, history, religion, poetry, mythology, psychology, construction, travel, technology, current events, and much more. It has been a privilege. I am grateful for the guidance and support throughout the graduate program. Thank you for everything!

I would also like to thank my committee members, Dr. Douglas Cress, Dr. Jiandong Chen, and Kathleen Egan. I appreciate your guidance and support throughout graduate school. A special dedication in remembrance of Dr. Catherine Phelan, who, although no longer with us, her kindness and invaluable advice will always remain with me. I thank Dr. Matthew Freedman at the Dana Farber Cancer Institute for giving me the opportunity to learn the 3C technique in his lab. That training has made an impact in my graduate work and in our lab, adding to the repertoire of many techniques we have developed and incorporated in our analyses. I would like to acknowledge Dr. Jaime Matta and his team at the Ponce School of Medicine for their support during my intern research rotation. I would also like to thank Dr. Kevin Brown from The National Cancer Institute for serving as the outside chairperson for my dissertation defense.

Additional experimental support has been provided by Kate Lawrenson, Dennis Hazelett, Simon G. Coetzee, and Houtan Noushmehr who provided the RNA-seq, FAIRE-seq, and ChIP-seq data for ovarian cell lines. Fine mapping analysis was performed by Paul Pharoah and Jonathan Tyrer at the University of Cambridge. FPLC was performed by Nicholas Woods.

I would also like to thank members of the Ovarian Cancer Association Consortium (OCAC), especially, Thomas Sellers, Simon Gayther, Paul Pharoah, and Ellen Goode for their expertise and guidance and for trusting me as a graduate student to contribute to the cause of the consortium. It has been a privilege and wonderful learning experience to learn from this elite group of experts in the field and it is remarkable to see the small contribution I am making influence a greater cause.

Thank you to all the former and present members of the Monteiro lab, for their help and for making the lab feel like my second home. Also, special thanks to all the wonderful friends that I made throughout the years, especially Ankita Jhuraney, Kyle Bauckman, Vito Rebecca, Paulo Cilas Lyra, Carly Harro and Sweta Dash who have made the graduate school experience an enjoyable one. I would also like to thank Dr. Kenneth Wright, Cathy Gaffney, and Janet Opel for the guidance and support they provide for all the Cancer Biology Ph.D. Students.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Ovarian cancer is a leading cause of death among women diagnosed with cancer. Mortality rate is high because an overwhelming majority of new cases are diagnosed with late-stage disease when the survival statistics are very poor with an overall 5-year survival rate of less than 40%. Despite the large burden of disease, the etiology of ovarian cancer is not well understood. In addition to linkage studies that have identified highly penetrant cancer susceptibility genes such as *BRCA1* and *BRCA2*, the emergence of Genome Wide Association Studies (GWAS) in the last decade has facilitated the identification of common genetic variants with low to moderate penetrance, termed single nucleotide polymorphisms (SNPs). However, the vast majority of risk-associated SNPs are found in non-coding regions and their mechanistic basis underlying susceptibility to cancer remains widely unknown.

Functional analyses are critical in providing insight into the mechanism of action of these variants. In this study we have implemented bioinformatics methods, *in silico* tools, and cell culture models to interrogate functional effects of SNPs that predispose to cancer. Importantly, we have developed and adopted new techniques that have been instrumental in our functional studies.

GWAS have identified forty chromosomal loci associated with susceptibility to ovarian cancer. In this study we present a comprehensive fine mapping and functional analysis of the 8q24.21 ovarian cancer susceptibility locus to establish biological

mechanisms underlying disease susceptibility. This locus was initially identified in a previous study, which discovered this region to be association with ovarian cancer. This follow-up work serves to replicate and confirm the initial findings and, most importantly, to investigate the biological mechanisms underpinning the association with ovarian cancer at the 8q24.21 chromosomal region.

The 8q24.21 locus is characterized as a gene desert, which interestingly is also associated with various cancer types including prostate, colorectal, breast and urinary bladder cancer. Following the signal of association with fine mapping and functional assays, we identify eight putative causal SNPs, which reside within two distinct enhancer elements, displaying allele-specific activity in ovarian cells. Long-range physical interactions with the promoter region of the *MYC* oncogene, and to a lesser extent, the non-coding RNA, *PVT1*, establish these two genes as the targets of the enhancer elements containing the putative causal SNPs. We present evidence that the ATF1 transcription factor binds exclusively to the minor allele of variant rs2165806. Additionally, we report a comparative analysis testing the cell-type specific activity of different enhancer regions in breast, prostate, colorectal and ovarian cell lines. In summary, this work highlights the regulatory landscape at the 8q24.21 locus and provides a mechanistic basis to understand susceptibility to ovarian cancer in this region.

CHAPTER ONE:

BACKGROUND

Ovarian Cancer

Epidemiology

Ovarian cancer is the fifth leading cause of cancer death in women, carrying the highest mortality rate of all gynecologic malignancies (Siegel et al. (2014)). According to the American Cancer Society, in the United States every year approximately 22,000 new cases are diagnosed and over 15,000 deaths occur due to ovarian cancer. In terms of population subgroups, rates are highest among Caucasian women and lowest among African American women (Siegel et al., 2014). Worldwide, there are estimated to be roughly 204,000 new cases and 125,000 deaths annually (Sankaranarayanan and Ferlay, 2006). Incidence rates are highest in developed regions of the world, such as North America and Western and Northern Europe (Bray et al., 2005; Coleman et al., 1993; Howe et al., 2006). Incidence of ovarian cancer increases with age and over 80% of tumors occur after the age of 45 years (SEER Cancer Statistics Review, 1975-2007)(SF, 2010).

Signs and Symptoms

Although ovarian cancer is known as the 'silent killer', over 80% of patients have symptoms in the early stages of the disease when the tumor is still confined to the ovaries (Goff et al., 2000). The main contributing factor in the high mortality rate of ovarian cancer is the lack of specific symptoms associated with the disease when it is localized to the ovary and has the best survival statistics (Jelovac and Armstrong, 2011). Most symptoms experienced by women with ovarian cancer are common for many gastrointestinal, genitourinary and gynecological conditions such as abdominal discomfort or pain, frequent urination, bloating, and early satiety, which are mild and nonspecific symptoms that are often disregarded by women (Goff et al., 2000; Goff et al., 2007). Consequently, diagnosis is usually delayed until more severe symptoms (e.g. increased pain, weight loss, obstruction of the urinary tract/intestines) become apparent as the disease progresses to more advanced stages (Jelovac and Armstrong, 2011). Unfortunately, patients presenting with advanced stages of ovarian cancer have dismal survival rates.

Pathogenesis and Molecular Heterogeneity

The origin and pathogenesis of ovarian cancer are poorly understood. For years ovarian cancer was considered and treated as a single disease when in fact that is not the case. Advances in research and treatment have been impeded by the uncertainty of the tissue of origin of the serous subtype of ovarian cancer and the remarkable molecular heterogeneity among the different tumor subtypes. Aside from ovarian surface epithelium (OSE) which was traditionally thought to be the precursor tissue for

epithelial ovarian cancer, there is evidence that the secretory tubal epithelial cells in the fimbria of fallopian tubes as well as endometriosis contribute to ovarian cancer pathogenesis (Berns and Bowtell, 2012; Kim et al., 2012; Lee et al., 2007).

Malignant epithelial ovarian cancers, known as ovarian carcinomas, are categorized into four major histologic subtypes that resemble the normal cells from which they are derived. These histotypes include serous carcinomas originating from the epithelial lining of the ovaries or fallopian tubes (70%), endometrioid carcinomas originating from endometrium cells (10-15%), clear cell carcinomas originating from vaginal cells (10%) and mucinous carcinomas originating from endocervix cells (3%) (Bast et al., 2009; Cho and Shih Ie, 2009), which are also thought to derive from metastasis of gastrointestinal tumors (Lee and Young, 2003).

Serous ovarian carcinomas are predominantly high-grade (90%) accounting for over 70% of all ovarian cancer deaths and will be the focus of this work (Jemal et al., 2009; Seidman et al., 2004). The different ovarian histological types differ in terms of disease development, pathology, genetic alterations, and prognosis (Landen et al., 2008; Levanon et al., 2008; Shih Ie and Kurman, 2004). Based on these observations, ovarian cancers are classified into two groups designated as type I and type II (Shih Ie and Kurman, 2004).

Type I tumors are slow growing and encompass low-grade serous, low-grade endometrioid, mucinous and clear cell carcinomas (Kurman and Shih Ie, 2008). As a group, these tumors are genetically stable and are generally characterized by mutations in *KRAS*, *BRAF*, *ERBB2*, *CTNNB1*, *PTEN* and *PIK3CA* (Cho and Shih Ie, 2009; Obata et al., 1998; Singer et al., 2003; Teer et al., 2017; Wu et al., 2001). Type I tumors have

slow progression of disease and are mostly diagnosed in early stages of the disease (Jemal et al., 2009).

Type II tumors, constituting the majority (75%) of all cases, exhibit very aggressive histology and consist of high-grade serous, high-grade endometrioid and undifferentiated carcinomas (Kurman and Shih Ie, 2010; Shih Ie and Kurman, 2004). Expression profiles of type II tumors cluster separately from type I tumors. Type II tumors are characterized by mutations in *TP53* (~ 80% of tumors), *BRCA1/2* (~50% of tumors), and proteins involved in the DNA homologous recombination repair (HRR) pathway as well as *CCNE1* amplifications but rarely encounter mutations that characterize type I tumors (Bonome et al., 2005; Cancer Genome Atlas Research, 2011; Kuo et al., 2009; Kurman and Shih Ie, 2010; Press et al., 2008; Teer et al., 2017).

In addition, high-grade serous tumors are also characterized by copy number aberrations and genomic instability, which is explained by disruption of the DNA damage repair system (Gorringe et al., 2007). Importantly, type II tumors display very high proliferative index (positive Ki-67 staining) with extremely aggressive clinical progression and very poor prognosis (Koonings et al., 1989; Siegel et al., 2014).

Diagnosis and Staging

Surgery plays an important role in ovarian cancer because it is essential in diagnosing, staging the disease and treating ovarian carcinomas. Although different noninvasive methods are used to detect ovarian cancer, such as pelvic exams; blood test for cancer antigen, CA 125; and transvaginal ultrasound, they have poor specificity

and/or sensitivity, so the ultimate determinant factor for staging of ovarian cancer is surgery (Jelovac and Armstrong, 2011).

According to the Federation of Gynecology and Obstetrics (FIGO) staging scheme, stage I ovarian cancer tumors are confined to the ovaries (20% of diagnoses); stage II tumors spread outside of the ovaries but are confined to the pelvis (5% of diagnoses); stage III tumors are confined to the abdominal cavity (58% of diagnoses); and stage IV tumors disseminate to lungs or other distant sites outside of the abdomen (17% of diagnoses) (Benedet et al., 2000). (Table 1)

Table 1: Ovarian cancer stage distribution and survival rate

Stage	Stage distribution	10-year survival rate
I	20%	73%
II	5%	45%
III	58%	21%
IV	17%	<5%

Adapted from (Benedet et al., 2000).

It is important to emphasize that low stage (I-II) ovarian tumors are not just simply confined "earlier" forms of high stage (II-IV) tumors (Kobel et al., 2010). Low stage disease is usually associated with type I tumors (Seidman et al., 2004), while high stage disease is predominantly found in type II high grade serous ovarian tumors (> 70% of cases) (Gilks, 2004; Vaughan et al., 2011). Unfortunately, only 25% of high grade serous ovarian cases are diagnosed at early stages (I, II) of the disease when the prognosis is promising (Benedet et al., 2000; Seidman et al., 2004).

Stage II is the least frequently diagnosed stage of ovarian carcinomas; the main reason being lack of a defined anatomical barrier separating the pelvis from the upper abdomen (Benedet et al., 2000; Vaughan et al., 2011). High-grade ovarian carcinomas spread quickly by shedding and circulating tumor cells in the abdominal cavity that eventually implant onto peritoneal surfaces (Jelovac and Armstrong, 2011). Considering the extremely aggressive clinical progression of serous ovarian carcinomas, once the disease disseminates outside of the ovaries into the pelvic cavity it spreads rapidly throughout the abdomen.

Stage III is the most frequently diagnosed stage of ovarian carcinomas (Benedet et al., 2000). Unfortunately, most of the patients diagnosed with advanced stage (III, IV) tumors that have spread to the abdominal cavity and other distant sites will die of disease (Gilbert et al., 2012). The 5-year survival rate for these patients is less than 40% (Jemal et al., 2009).

Treatment and Prognosis

Prognosis for patients with ovarian cancer depends greatly on the stage of disease at the time of diagnosis and the volume and pathology of tumor remaining after the initial cytoreduction or debulking surgery (Bristow et al., 2002; Jelovac and Armstrong, 2011). Patients diagnosed with stage I ovarian carcinomas are generally treated with surgery alone without adjuvant therapy and have a 5-year disease-free survival rate of over 90% (Young et al., 1990). Standard treatment of patients diagnosed with stage II tumors involves surgery followed by 3 cycles of single agent adjuvant

chemotherapy (e.g. paclitaxel, carboplatin) resulting in a 5-year disease-free survival rate of 80% (Bell et al., 2006).

Roughly 75% of cases are diagnosed with advance stage (III-IV) ovarian carcinomas and are treated with cytoreductive surgery followed by 6 cycles of multi agent combination chemotherapy (ex. cisplatin or carboplatin in combination with paclitaxel) (Ozols et al., 2003). The outcome of the treatment for these patients varies depending on the amount of residual tumor left behind after debulking surgery - the smallest remaining tumor lesions having the best prognosis (Bristow et al., 2002; Vaughan et al., 2011). Patients with optimally debulked disease (i.e. residual tumor < 1cm) have a 10-year survival rate of 25% (Benedet et al., 2000). Patients with suboptimally debulked tumors (i.e. residual tumor > 1cm) have considerably worse prognosis (Benedet et al., 2000). Nevertheless, a small fraction of stage III patients will have disease-free survival.

Unfortunately, stage IV patients rarely have long term disease-free survival (Jelovac and Armstrong, 2011). Despite advances in surgery and chemotherapy, over 90% of women with advanced stage ovarian carcinomas will die after the cancer relapses (Bukowski et al., 2007). Early detection and prevention are therefore essential in reducing ovarian cancer mortality (Bast et al., 2009).

PARP inhibitors. Recent developments in identification of targeted approaches have shown that poly (ADP-ribose) polymerase (PARP) inhibitors lead to improved survival in patients carrying pathogenic variants in *BRCA1* and *BRCA2* genes (Fong et al., 2009; Gien and Mackay, 2010; Gunderson and Moore, 2015; Tangutoori et al., 2015). Cells rely on two distinct pathways of DNA damage repair in order to maintain

their DNA integrity. PARP is an enzyme involved in single-strand DNA repair (i.e. non-homologous end joining; NHEJ), while BRCA 1 and BRCA 2 proteins are involved in double-strand DNA break repair processes (i.e. homologous recombination; HR) (Schreiber et al., 2006).

Cells that lose function of *BRCA* genes have defective double-strand DNA repair and, therefore, rely heavily on the alternate DNA repair mechanism (i.e. single-strand DNA repair) which involves PARP (Bryant et al., 2005; Farmer et al., 2005). Because inactivation of *BRCA* genes is common in ovarian carcinomas, those patients can benefit from targeted therapy with PARP inhibitors by impeding the ability of the tumor cells to repair the chemotherapy-induced DNA damage, causing increased cytotoxicity of the tumor cells (Ashworth, 2008). Olaparib is a highly potent PARP inhibitor that has recently been approved by the U.S. Food and Drug Administration (FDA) and the European Commission for therapy in ovarian cancer patients with high-grade serous tumors that carry *BRCA1* and *BRCA2* pathogenic variants and are sensitive to platinum therapy (Gunderson and Moore, 2015). Treatment with PARP inhibitors has shown favorable outcomes both for patients with germline and those with somatic ovarian carcinomas (Carden et al., 2010; Gunderson and Moore, 2015).

.

Protective Factors

Several epidemiologic, lifestyle, environmental and genetic factors are known to influence a woman's lifetime risk for ovarian cancer. Factors associated with reduced risk, such as parity, breastfeeding or lactation, and use of oral contraceptives, have been hypothesized to confer a protective effect by reducing the number of lifetime

ovulations (Hankinson et al., 1995; Risch et al., 1994). Specifically, women who have three children or use oral contraceptives for five years have over 50% reduction of risk (Whittemore et al., 1992).

One theory is that reducing the number of lifetime ovulations reduces the damage and inflammation of the ovarian epithelium that occurs with each ovulation (Auersperg et al., 1997; Ziltener et al., 1993). There is evidence that nonsteroidal anti-inflammatory drugs (NSAIDS) and other anti-inflammatory medicines may reduce risk of developing ovarian cancer (Fairfield et al., 2002).

It has also been hypothesized that both pregnancies and oral contraceptives provide a protective hormonal environment (e.g. reduced gonadotropin, estrogen levels; high progestin levels) that decreases risk of developing ovarian cancer by promoting apoptosis of genetically damaged ovarian epithelial cells that would likely transform into malignant cells (Risch, 1998; Rodriguez et al., 1998; Salehi et al., 2008). The hormonal effect may also explain the excess level of protection obtained beyond the effect of pregnancy and oral contraceptive use in reducing lifetime ovulations (Whittemore et al., 1992).

A noticeable reduction (20%-50%) in ovarian cancer risk has been observed in women who have undergone tubal ligation or hysterectomy, thus, removing the tissues that are thought to be the precursors of ovarian carcinomas in the first place (Hankinson et al., 1993; Irwin et al., 1991; Loft et al., 1997). Another mechanism of protection provided by gynecologic procedures is hypothesized to be the disruption of ascending infections of the ovaries associated with talcum powder and pelvic inflammatory disease

as well as decrease of blood flow to the ovaries (Chiaffarino et al., 2005; Huncharek et al., 2003; Risch and Howe, 1995).

Risk Factors

Increased incidence of ovarian cancer has been associated with nulliparity, late menopause, early menarche, and hormonal therapy, which increase the number of lifetime ovulations and exposure to estrogen (Morch et al., 2009; Salehi et al., 2008). Furthermore, endometriosis (an inflammatory disorder) has been associated with increased risk of ovarian carcinomas (Ness, 2003; Van Gorp et al., 2004). Interestingly, there is evidence of increased prevalence of ovarian cancer in regions with low sunlight exposure, suggesting that reduced sunlight and vitamin D levels may increase risk of developing the disease (Egan, 2006; Lefkowitz and Garland, 1994; Rodriguez, 2003). Cigarette smoking is also associated with increased risk of certain subtypes of epithelial ovarian carcinomas (Jordan et al., 2006). Although it is unclear whether physical activity influences ovarian cancer risk, sedentary behavior is associated with increased risk in two studies (Patel et al., 2006; Zhang et al., 2004).

The Genetics of Ovarian Cancer. Although multiple factors play a role in modifying ovarian cancer risk, the strongest impact is observed in association with genetic factors (Stratton et al., 1998). Ovarian cancer tends to aggregate in families. Although family studies cannot differentiate whether the aggregation is due to genetic or environmental causes, studies on monozygotic twins and inbred populations provide evidence that the observed familial clustering of ovarian cancer is mainly due to genetic

factors rather than lifestyle or environmental factors (Goldgar et al., 1994; Lichtenstein et al., 2000).

The relative risk of women who have a first-degree relative affected with ovarian cancer ranges between three to twelve, increasing with the number of affected relatives within the family (Stratton et al., 1998). These families commonly carry pathogenic variants in *BRCA1* and *BRCA2* genes, which are involved in DNA damage repair processes, or *MLH1* and *MSH2* genes, which are part of the mismatch repair (MMR) system and act to repair nucleotide mismatch errors during DNA replication (Wheeler et al., 2000). Risk is highest in families that carry *BRCA1* or *BRCA2* mutations (Antoniou et al., 2002; Lawrenson et al., 2015b; Pharoah and Ponder, 2002). Mutations in *BRCA1/2* genes have been associated with high grade serous ovarian cancer, while mutations in MMR genes are more likely associated with mucinous and endometrioid ovarian carcinomas (Berns and Bowtell, 2012; Fujita et al., 1995; Lawrenson et al., 2015a).

A portion of ovarian cancers are known to occur as part of two common hereditary cancer syndromes: the hereditary breast and ovarian cancer syndrome, involving *BRCA1* and *BRCA2*, and the Lynch syndrome, alternately known as the hereditary non-polyposis colorectal cancer (HNPCC) syndrome, involving (*MMR*) genes such as *MLH1*, *MSH2*, *MSH6*, *PMS2*, and *EPCAM* (Lawrenson et al., 2012; Lawrenson et al., 2015b; Lawrenson et al., 2013; Pharoah and Ponder, 2002; Walsh et al., 2011). Procedures and regulations for genetic testing for known ovarian cancer susceptibility genes have already been established in the clinic (Kauff et al., 2008; Lindor et al., 2006).

Germline pathogenic variants have also been found in two additional highly penetrant cancer susceptibility genes, *TP53* and *PTEN*. (Lawrenson et al., 2012; Lawrenson et al., 2015b). Pathogenic variants in highly penetrant ovarian cancer susceptibility genes are rare in the population, but have a strong genetic effect and are therefore highly penetrant, meaning a good proportion of the women carrying the mutation will develop the disease (Figure 1) (Manolio et al., 2009; Pharoah et al., 2004).

Additional variants with intermediate genetic effect have been identified in various genes involved in ovarian cancer processes. Some of these genes include *ATM*, *CHEK1*, *CHEK2*, *BRIP1*, *BARD1*, *PALB2*, *RAD50*, *RAD51C*, *RAD51D*, which are involved in DNA damage repair and cell cycle processes (Castera et al., 2014; Kuusisto et al., 2011; Lawrenson et al., 2012; Staples and Goodman, 2013; Thorstenson et al., 2003; Walsh et al., 2011). Pathogenic variants in moderately penetrant genes are found in low frequency and have a moderate to intermediate genetic effect in ovarian cancer risk (Figure 1) (Manolio et al., 2009).

Still, high and moderately penetrant pathogenic alleles of known ovarian cancer susceptibility genes account for approximately 50% of hereditary cases and only 10-15% of all ovarian cancer cases, thus explaining only a fraction of the excess familial risk, suggesting that other genetic factors are involved in ovarian cancer processes (Gayther et al., 1999; Lawrenson et al., 2010; Lawrenson et al., 2013; Pharoah et al., 2004)

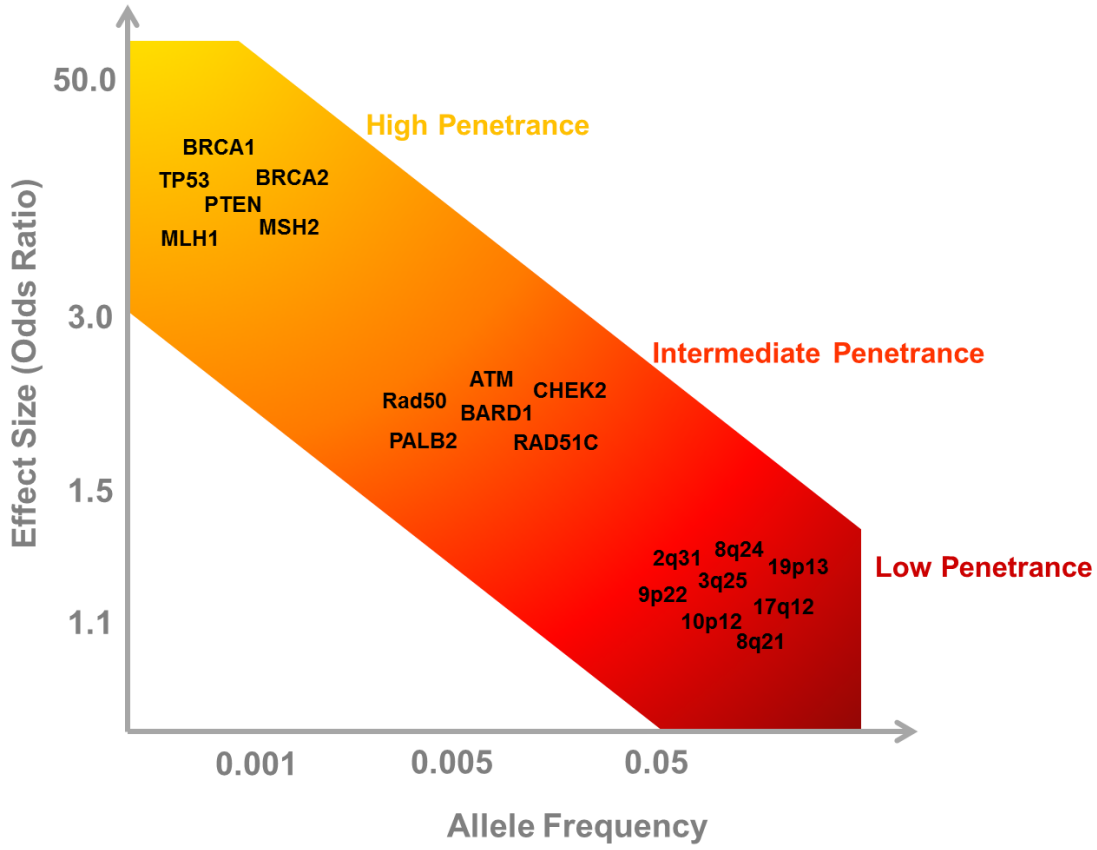


Figure 1. Ovarian cancer susceptibility genes and their contribution to disease. Examples of different ovarian cancer susceptibility genes are plotted by their approximate strength of genetic effect versus allele frequency. Rare genes with large effect size are highly penetrant, common variants with low effect size have low penetrance, and genes with modest effect size and intermediate allele frequency have intermediate penetrance of the disease. Adapted from (Manolio et al., 2009).

Since the search for rare, highly-penetrant ovarian cancer susceptibility genes such as *BRCA1* and *BRCA 2* has been exhausted, the focus has recently shifted towards exploring genetic variants that confer moderate to low risk (low-penetrance) and are common in the population (Pharoah et al., 2004; Pharoah and Ponder, 2002). Although these variants individually have a small genetic effect, collectively they may have a significant effect on the disease (Fasching et al., 2009; Pharoah et al., 2004). Recently, new population-based association studies have identified several loci that are

associated with ovarian cancer susceptibility (Bolton et al., 2010; Goode et al., 2010; Kuchenbaecker et al., 2015; Permuth-Wey et al., 2013; Pharoah et al., 2013; Song et al., 2009). Although these variants have a small genetic effect, they are common in the population and collectively may explain the remaining genetic contribution to risk of developing ovarian cancer (Figure 1) (Fasching et al., 2009; Manolio et al., 2009; Pharoah et al., 2004).

Decreasing Ovarian Cancer Mortality

Several factors contribute to the high mortality rate of ovarian cancer. The screening modalities currently available are ineffective (Partridge et al., 2009). The disease symptoms are mild and non-specific and the diagnosis is usually delayed until disease has progressed extensively. Moreover, the highly heterogeneous and extremely aggressive histology of ovarian carcinomas present a challenge to clinicians to successfully treat the disease, resulting in very poor survival rates for ovarian cancer patients.

Considering the devastating lethality of ovarian carcinomas, it is suggested that targeting prophylactic measures and improved screening strategies to high-risk women (i.e. having inherited predisposition and greatest risk of developing the disease) have significant potential in reducing ovarian cancer mortality (Fasching et al., 2009; Pharoah and Ponder, 2002). A better understanding of the disease is critical for developing strategies for risk assessment, prevention, early detection, and development of therapeutic agents to ultimately improve survival.

Genome Wide Association Studies (GWAS)

High-penetrance Genes and Linkage Studies

Defective *BRCA1* and *BRCA2* genes are rare in the population but are highly penetrant, following to some extent a Mendelian or single-gene pattern of inheritance (Pharoah et al., 2004). Highly penetrant pathogenic variants and traits are deposited in a publicly available database named the Online Mendelian Inheritance in Man (OMIM).

Family-based linkage studies have been successful in identifying rare, highly penetrant genes such as *BRCA1* and *BRCA2* and many genes associated with true Mendelian disorders, such as Huntington's disease and cystic fibrosis. However, in ovarian cancer, as well as other cancers, the known susceptibility genes such as *BRCA1*, *BRCA2* and *TP53* explain only a fraction of the excess familial risk, implying that additional genetic factors are also involved in ovarian cancer processes (Pharoah and Ponder, 2002).

Low-penetrance Alleles and Association Studies

Most cancers, including ovarian cancer, are very complex diseases with polygenic inheritance, meaning that the inherited risk is due to multiple risk alleles with low to moderate individual effect, which collectively may have a significant impact on the disease (Fasching et al., 2009; Pharoah et al., 2004) (Figure 1). Millions of common genetic variants such as insertions, deletions, sequence repeats or nucleotide substitutions have been found in the human genome (Cargill et al., 1999; Genomes Project et al., 2012).

The most common type of variation is a single nucleotide base substitution or polymorphism (SNP), which is generally observed in at least 1% of the population and is described by minor allele frequency (MAF), referring to the allele with the lower frequency in a reference population (Chung et al., 2010). Tens of millions of SNPs identified in the human genome have been catalogued in an international public SNP database, called the dbSNP (www.ncbi.nih.gov/SNP/) (International HapMap, 2003; International HapMap et al., 2007).

The Principles of GWAS Design

The best model to study SNP variants is a population-based association study, which compares the frequency of the SNP alleles in disease- or trait-affected individuals to a comparable control group (Cardon and Bell, 2001; Risch, 2000). A genome-wide association study (GWAS) is a hypothesis-free approach used to identify SNPs across the genome that are associated with a trait or disease, such as cancer (Stadler et al., 2010). By comparing distribution of alleles between cases and controls, GWAS can identify genomic regions or loci that are associated with the disease (Chung et al., 2010; Corvin et al., 2010).

GWAS utilize the knowledge that nearby genomic loci are often inherited together more often than is expected by chance due to lack of recombination, mainly because of their close juxtaposition to each other (Stadler et al., 2010). The nonrandom association of alleles in adjacent loci within the same chromosome that are often correlated together is known as linkage disequilibrium (LD) (Reich et al., 2001) (Figure 2)

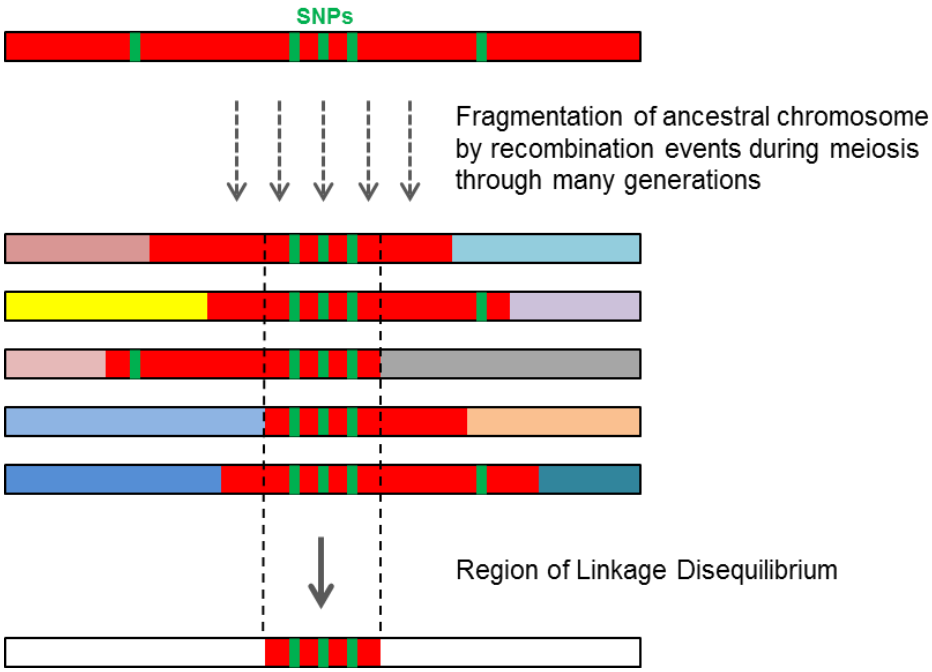


Figure 2. Illustration of linkage disequilibrium. The set of alleles or SNPs that are inherited together through many generations are in linkage disequilibrium. The LD fragment is delineated by recombination events that occur during meiosis.

The most commonly used measure of LD in GWAS is the r^2, which is a statistical correlation between two loci or SNPs that takes into account the LD as well as the allele frequency (Pharoah et al., 2004; Visscher et al., 2017). The r^2 ranges in value from 1 indicating complete correlation or 'perfect' LD, which can be disrupted by recombination events to the point where no correlation is observed, $r^2 = 0$ (Orr and Chanock, 2008; Reich et al., 2001) (Figure 3A).

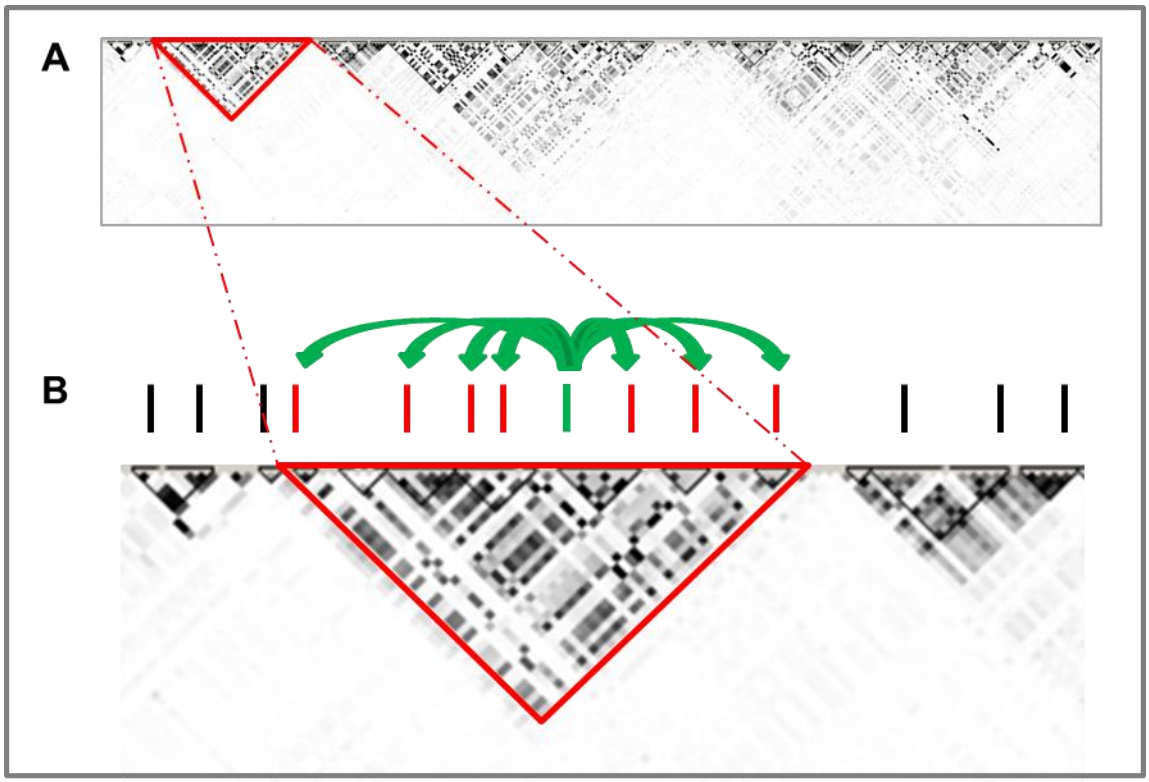


Figure 3. LD structure and tagSNP selection. ***A***. LD plot of a chromosome fragment represented by r^2. Color intensity indicates SNPs in increasing LD from $r^2 = 0$ (white), to $r^2 = 1$ (black). LD structure indicated with red triangle. ***B***. Tagging SNP (green tick) acts as a proxy for all SNPs within the LD structure (red ticks), but not for the SNPs in the neighboring regions (black ticks). A tagging SNP that is associated with a trait or disease represents all SNPs within the LD block as candidate causal SNPs.

The set of alleles or SNPs that are in strong LD and tend to be inherited together from generation to generation is called a haplotype (Sabeti et al., 2002). LD structures allow particular SNPs, designated as tagSNPs, to act as surrogates for the other SNPs on the haplotype (Pharoah et al., 2004) (Figure 3B). The ability of one SNP to act as a surrogate marker for another is determined both by the strength of LD and their relative allele frequencies (Pharoah et al., 2004).

The fundamental principle of GWAS testing is to use carefully selected tagSNPs that serve as proxies for untested SNPs, thus being able to capture genetic diversity

across the genome (Cardon and Abecasis, 2003; Carlson et al., 2004; Johnson et al., 2001).

GWAS Methodology and Data Interpretation

The GWAS approach is used to identify common, low-penetrance genetic variants (e.g. SNPs) associated with a particular disease. Most SNPs that have been identified by GWAS have a modest effect (OR < 1.3) and relatively high minor allele frequency (MAF ≥ 1%) so in order to ensure adequate statistical power a large sample size is required for these studies (Visscher et al., 2017; Wang et al., 2005).

A typical GWAS for cancer susceptibility selects tens of thousands of patients with a particular cancer (cases) and tens of thousands of individuals without the disease (controls) from a comparable control group to achieve statistical power for detection of modest ORs after adjustment for multiple testing. Information on age or ethnicity is necessary to adjust for 'admixture' confounding(Stadler et al., 2010). Taking into consideration the requirement for large sample sizes, it is typical for primary GWAS studies to be grouped together to form meta-analysis, thus enabling discovery of associations that would otherwise not be evident in separate individual studies (Brisbin et al., 2011).

In the last thirteen years, the identification of susceptibility loci for common cancers has been driven by large consortia that are part of the Collaborative Oncological Gene-environment Study (COGS). These consortia include the Ovarian Cancer Association Consortium (OCAC), the Breast Cancer Association Consortium (BCAC), Prostate Cancer Association Group to Investigate Cancer-Associated

Alterations in the Genome (PRACTICAL), and the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA).

OCAC, created in 2005, is a forum of interdisciplinary investigators conducting case-control studies of ovarian cancer with the aim of identifying genetic factors related to the risk of ovarian cancer. Over the years their aim has broadened to include identification of genetic variation associated with clinical phenotypes such as chemo-sensitivity and overall survival. Importantly, OCAC is also the only functional venue for large-scale validation of ovarian cancer susceptibility loci. Since its conception in 2005, the OCAC has been very successful in identifying genetic loci associated with ovarian cancer risk (Bolton et al., 2012; Pharoah et al., 2013).

GWAS utilize DNA samples (mainly blood or buccal swabs) from the study subjects that are genotyped using commercially available microarray chips, which can perform high-throughput analysis of up a million SNPs at a time. Each DNA sample is genotyped for a set of tagSNPs that are selected to capture the genomic diversity across the entire human genome by serving as proxies for untested SNPs (Yang et al., 2013).

The basic statistical analysis in GWAS is logistic regression with disease status as the dependent variable and SNP genotype as the independent variable (Corvin et al., 2010). Logistic regressions are performed for each SNP in the microarray chip (i.e. about one million regression models per chip), identifying the reference allele for each SNP along with the odds ratio and p-value (Corvin et al., 2010).

Since multiple tests are performed, stringent statistical thresholds are needed to avoid false-positive results. The most commonly used adjustment for multiple testing in

GWAS is the Bonferroni correction in which the standard type I error p = 0.05 is divided by the number of tests performed (typically one million per chip) thus generating an acceptable GWAS threshold of $p \leq 5 \times 10^{-8}$ (Corvin et al., 2010; Stadler et al., 2010). P-values that are smaller than expected by chance pass the threshold of significance (i.e. $P \leq 5 \times 10^{-8}$) for GWAS and highlight a genomic locus that may be casually associated with the disease.

The standard for a successful GWAS finding includes (1) a strong association in the initial sample, (2) replication of association in an independent study, and (3) a cumulative p-value, $p \leq 5 \times 10^{-8}$ (Studies et al., 2007). Since its first application in 2005, the GWAS method has been very successful in identifying genomic loci associated with different traits and diseases as is evident with over 3,000 publications in the GWAS catalog of the National Human Genome Research Institute (NHGRI) (Welter et al., 2014) (A Catalog of Published Genome-Wide Association Studies).

Ovarian Cancer Loci Identified by GWAS

Genome wide association studies to date have identified 40 genomic loci associated with ovarian cancer risk (Bojesen et al., 2013; Bolton et al., 2010; Chen et al., 2014; Goode et al., 2010; Kar et al., 2016; Kelemen, 2015; Kuchenbaecker et al., 2015; Permuth-Wey et al., 2013; Pharoah et al., 2013; Phelan et al., 2017; Shen et al., 2013; Song et al., 2009). These risk-associated loci are found mainly in noncoding regions of the DNA (e.g. between genes or within introns) and are associated with small genetic effect (odds ratio; OR), or low penetrance – meaning, a small frequency of individuals who carry the SNP variant will manifest the disease.

Compared to highly-penetrant genes such as *BRCA1* and *BRCA2* that have odds ratios greater than 7, the OR for the identified ovarian SNPs, and all SNPs in general, lie between 0.8 to 1.4 (Kuchenbaecker et al., 2015). This is in line with the expected small biological effect of SNP variants under the common-disease-common variant model. Although individually these SNPs have a modest effect, it is hypothesized that the common occurrence of SNPs in the population and the cumulative effect of particular SNP sets contribute significantly to the burden of disease (Pharoah et al., 2004). This observation may be also due to the GWAS association capturing the actual causal variant that carries most of the effect among the pool of variants identified.

The 8q24 ovarian cancer locus. A previous study reported the 8q24 locus, tagged by the SNP (rs10088218) variant allele associated with the protective effect, especially in the serous subtype (OR = 0.76, 95% CI 0.70-0.81, $p = 8 \times 10^{-15}$) (Goode et al., 2010). Two additional SNPs, rs1516982 and rs10098821, also reached genome wide significance ($p \leq 5 \times 10^{-8}$). This locus is located approximately 800 kb telomeric to the myelocytomatosis (*MYC*) proto-oncogene (see Figure 4 for illustration).

The most strongly associated SNP, rs10088218, was also tested for association with overall survival in patients classified in ovarian cancer subgroups different from serous (e.g. histological subtype, tumor stage and tumor grade), but no evidence of association was observed (Goode et al., 2010). Real time PCR analysis of *MYC* levels in 48 primary human ovarian surface epithelium cell cultures and 24 ovarian cancer cell lines revealed significant increase of *MYC* expression in the ovarian cancer cell lines

compared to normal cultures (p = 0.0011); however, no evidence of association was observed between SNP allele and gene expression (Goode et al., 2010).

Architecture of the 8q24 genomic region. The 8q24 locus is a well-known cancer susceptibility region that is associated with different types of cancer including prostate, breast, ovarian, bladder, pancreatic, colorectal and lung cancer (Easton et al., 2007; Eeles et al., 2008; Ghoussaini et al., 2008; Goode et al., 2010; Gudmundsson et al., 2007; Kiemeney et al., 2008; Low et al., 2010; McKay et al., 2017; Michailidou et al., 2013; Pharoah et al., 2013; Tenesa et al., 2008; Tomlinson et al., 2007; Yeager et al., 2007; Zanke et al., 2007; Zhang et al., 2016).

This region is characterized as a gene desert, which by definition is a large genomic region devoid of protein-coding genes without any apparent relevance in biological functions (Ovcharenko et al., 2005; Venter et al., 2001). Nevertheless, there is ample evidence that some gene deserts, including 8q24, contain regulatory sequences that play an important role in regulating expression of distantly located genes (Huppi et al., 2012; Jia et al., 2009; Nobrega et al., 2003; Pomerantz et al., 2009).

The only well-annotated gene in the 8q24 region is the myelocytomatosis (*MYC)* proto-oncogene, which is a transcription factor known to be involved in many biological processes that are important in cancer development and progression, including growth and proliferation, transformation, differentiation, and apoptosis (Adhikary and Eilers, 2005). Other genes in the area include transmembrane protein 75 (*TMEM75)*, which yields a protein of unknown function, and plasmacytoma variant translocation 1 (*PVT1*), producing a long non-coding RNA.

Many cancer loci are clustered in a region centromeric to *MYC* at distances between 30-800 kilobase pairs (kb) away (Figure 4). There are at least 6 distinct cancer susceptibility loci most of which are specific for a particular cancer (Amundadottir et al., 2006; Easton et al., 2007; Gudmundsson et al., 2007; Haiman et al., 2007b). Interestingly, one region contains highly correlated SNPs that are associated with prostate, colorectal and ovarian cancer suggesting a shared underlying factor that influences the risk of the three different cancers (Ghoussaini et al., 2008; Haiman et al., 2007a; Tomlinson et al., 2007).

There is growing evidence suggesting that SNPs present in the various cancer susceptibility regions at the 8q24 locus reside within regulatory sequences and regulate transcription of *MYC* through long range interactions with its promoter region, likely in a cell type specific manner (Ahmadiyeh et al., 2010; Jia et al., 2009; Pomerantz et al., 2009; Sotelo et al., 2010; Wright et al., 2010). However, *MYC* transcript levels have been difficult to evaluate, mainly due to the inability of the current available platforms to detect slight changes in expression levels, thus most studies fail to show a correlation between risk allele and *MYC* expression levels (Grisanzio and Freedman, 2010; Huppi et al., 2012).

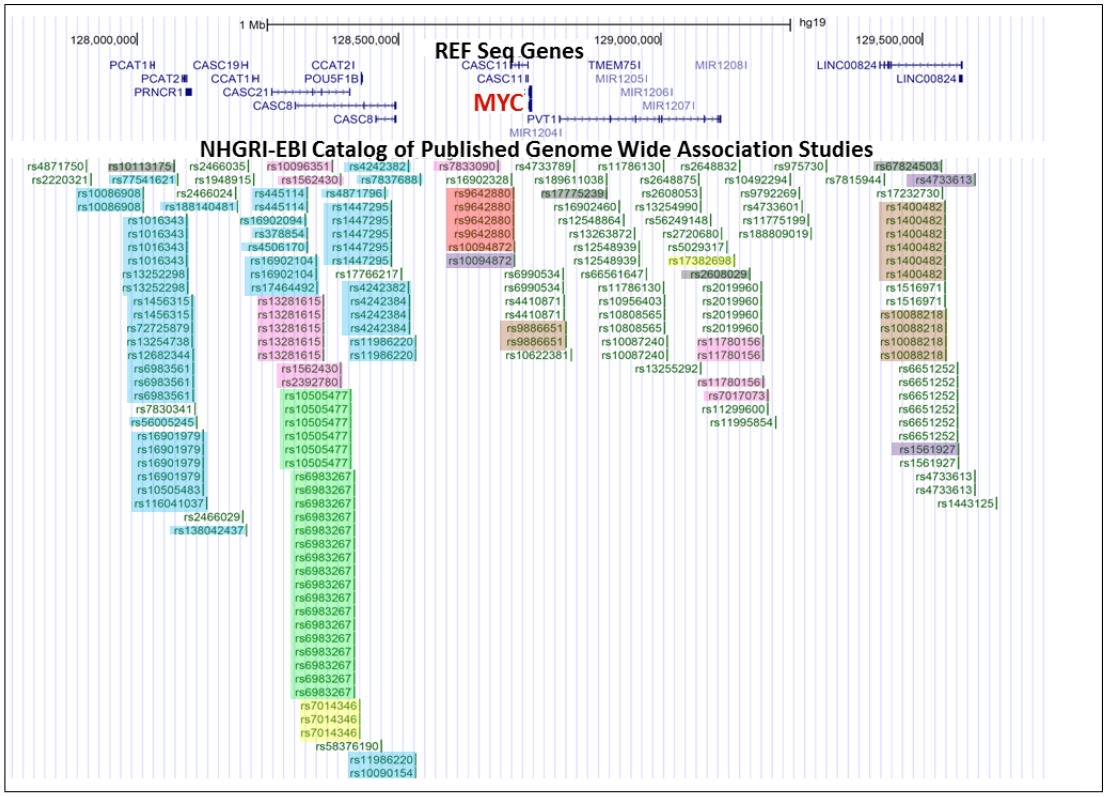


Figure 4. Chromosome 8q24 architecture. Genome Browser view of specific SNPs for various cancer loci identified at the 8q24 chromosomal region that have been deposited in the Catalog of Published GWAS. Blue highlight indicates association with prostate cancer; yellow = colon cancer; green = prostate and colon cancer; pink = breast cancer; red = bladder cancer; grey = lung cancer; purple = pancreatic cancer; orange = ovarian cancer.

Nevertheless, it has been suggested that even subtle changes in expression levels can influence tumorigenesis (Yan et al., 2002). An *in vivo* study by Wasserman et al. revealed that one of the prostate cancer variants in the 8q24 locus is associated with endogenous *MYC* expression during the early development stages of prostate tissue, suggesting that risk alleles may play a significant role in carcinogenesis well before tumor formation, in specific and brief time points during the development (Wasserman et al., 2010).

Genomic alterations in the 8q24 genomic region. In addition to accumulation of various cancer-associated SNPs, the 8q24 genomic region is also known to be subject to a variety of frequent somatic aberrations, including translocations, rearrangements, viral integration, mutations, and amplifications (Beroukhim et al., 2010; Dalla-Favera et al., 1982; Patel et al., 2004; Peter et al., 2006).

This region was initially recognized to be the target of various chromosomal translocations such as t(8:14) and less frequently t(8:22) and t(2:8), which are characteristic of Burkitt's Lymphoma, thus positioning the *MYC* oncogene next to an active promoter, resulting in increased expression and malignant transformation (Dalla-Favera et al., 1982; Zech et al., 1976).

Additionally, a large scale study has identified 8q24 as the most frequently amplified region across many human cancers (Beroukhim et al., 2010; Haverty et al., 2009). More specifically, The Cancer Genome Atlas (TCGA) project (http://cancergenome.nih.gov/), which has catalogued molecular aberrations in 489 high grade serous ovarian tumors, has reported that the *MYC* region is one of the most commonly amplified in ovarian carcinomas (Cancer Genome Atlas Research, 2011).

The 8q24 region has also been identified as the most common site of human papilloma virus (HPV) integration in genital neoplasia, accompanied with amplification of various foci within the 8q24 region (Durst et al., 1987; Herrick et al., 2005; Kraus et al., 2008; Peter et al., 2006). However, the role of HPV in ovarian carcinogenesis is inconclusive. There are no reports of coincidence of ovarian cancer and cervical cancer. There is evidence of high HPV prevalence in ovarian cancer tissue, but it varies widely by geographical area ranging from 45.6% (95% CI, 31.0-60.3) in Asia, to 18.5% (95%

CI, 8.5-28.6) in Eastern Europe, to 1.1% (95% CI, -1.6 to 3.8) in Western Europe and zero in North America (Rosa et al., 2013; Svahn et al., 2014).

Altogether, there is ample evidence pointing at a critical role of the 8q24 region in cancer processes that warrants further exploration of this genomic region to better understand the mechanisms underlying susceptibility to ovarian cancer in this locus.

Mechanisms of Transcriptional Regulation

Gene Transcription

The central dogma of molecular biology describes the sequential transfer of information from the DNA genetic code to RNA to protein (Crick, 1970; Crick, 1958), although there are exceptions to the rule (Baltimore, 1970; Gerstein et al., 2007). The process of generating RNA from a DNA template is termed transcription and there is evidence that up to 75% of the human genome can be transcribed (Djebali et al., 2012). Eukaryotic cells contain three classes of RNA polymerase enzymes (RNA pol I, II, and III) which are used to catalyze synthesis of different classes of RNA (Hahn, 2004).

Mechanism of Transcription Initiation and Elongation

Transcription of protein-coding genes into messenger RNA (mRNA) involves the assembly of the RNA polymerase II (pol II) transcription machinery at the promoter of a gene (Hahn, 2004). RNA Pol II is an enzyme consisting of twelve subunits that requires various additional proteins for promoter recognition and transcription initiation (Myer and Young, 1998; Ptashne and Gann, 1997).

Promoters contain DNA sequence elements that bind and orient RNA pol II subunits, specifying the location of transcription for a specific gene (Juven-Gershon et al., 2008). Core promoters typically contain a TATA element, an (A+T)-rich sequence located 50-70 bp upstream of the transcription start site (TSS). The TATA element is recognized and bound by the TATA binding protein (TBP), which is responsible for recruiting RNA pol II to the TSS (Struhl, 1989).

Additional core promoter elements, such as the initiator element (Inr), downstream promoter element (DPE), and TFIIB-recognition element (BRE) are recognized by general transcription factors (GTFs), which assist in assembling the RNA pol II transcription machinery (Smale and Kadonaga, 2003).

Typically, once transcription initiation starts, the RNA pol II machinery transcribes a short distance (20-50 bp) and then is paused by physical association with pause control factors (Adelman and Lis, 2012). The paused polymerase may proceed to elongation or terminate transcription altogether by releasing the nascent RNA.

Pause release and elongation proceed upon recruitment of transcription elongation factors, such as the positive elongation transcription factor b (P-TEFb), through formation of elongation complexes (Luo et al., 2012; Smith et al., 2011). Transcription factors such as MYC regulate transcriptional pause release of RNA pol II by interacting with P-TEFb, thus controlling transcription regulation of a target gene by regulating its transcription elongation (Rahl et al., 2010).

The Fundamentals of Transcriptional Regulation

Regulation of transcription is essential for all living organisms as it controls many biological processes, including cell growth, differentiation, response to internal and external stimuli, development, and disease (Struhl, 1995). The groundwork for the basic concepts of transcriptional regulation was initiated in bacterial systems over five decades ago (Jacob and Monod, 1961). Today, it is well-established that transcription factors binding to specific DNA sequences at regulatory elements are the fundamental basis of transcriptional regulation, along with their cofactors, the transcription machinery and various chromatin regulators (Lee and Young, 2013). Regulatory mechanisms that control the differential and dynamic expression of genetic information define the properties of cells and delineate individual cell types (Hobert, 2008).

Transcriptional regulation is achieved by combinations of interactions between regulatory elements (i.e. promoters, enhancers, silencers and insulators) and a wide variety of molecular factors such as transcription factors, cofactors and chromatin remodelers (Coulon et al., 2013; Hubner et al., 2013).

Transcriptional regulation requires two interconnected levels of regulation: one involves transcription factors and the transcription machinery, and the second involves chromatin organization and its regulators (Lee and Young, 2013).

Transcriptional regulation by transcription factors and enhancers. Transcription factors typically control gene transcription by binding enhancer elements and recruiting cofactors and the RNA pol II machinery to target genes (Lelli et al., 2012; Ong and Corces, 2011; Spitz and Furlong, 2012). The timing, location and the level of transcription for a specific gene is determined by DNA regulatory sequences, which

include the core promoter, proximal regions and enhancer regions (Fuda et al., 2009). The core promoter works in concert with the other regulatory regions to regulate transcription (Andersson et al., 2015).

Enhancer regions contain regulatory elements called enhancers that guide recruitment of specific transcription factors, which can be classified as activators or repressors depending on whether they act to stimulate or inhibit transcription (Fuda et al., 2009). Activators can facilitate transcription initiation by interacting directly with many GTFs (e.g. TBP, TFIIA, TFIIB, TFIID), or through recruitment of co-activators, such as the Mediator (Esnault et al., 2008; Kim et al., 1994). Cofactors play a central role in DNA loop formation and maintenance between enhancers and promoters, which is an important genomic structure for proper gene control (Kagey et al., 2010).

Additionally, activators recruit chromatin-modifying enzymes that modulate chromatin condensation (and, consequently, accessibility of the RNA pol II transcription machinery to promoters), thus directly affecting gene expression (Henikoff, 2008; Schulze and Wallrath, 2007). Co-activators often function as histone modifiers (e.g. histone acetyltransferases (HATs), histone methyltransferases (HMTs), ATP-dependent chromatin remodelers, or mediators in RNA pol II recruitment (Roeder, 2005).

Transcriptional repressors inhibit transcription by directly or indirectly (i.e. via recruitment of co-repressors) recruiting chromatin modifying complexes to remodel chromatin into a condensed or inactive form (Shapiro and Shapiro, 2011). Insulator proteins are also another class of transcription factors, which identify insulator nucleotide sequences, that set boundaries between adjacent genes by blocking actions of cis-regulatory sequences in one gene from interfering with a neighboring gene (Bell

et al., 1999; Kellum and Schedl, 1991). They are generally 300 bp to 2 kb in length and have multiple binding sites for large proteins such as CTCF (Bell et al., 2001; Burgess-Beusse et al., 2002).

Transcription factor binding sites are the fundamental units of regulatory sequences (Levo and Segal, 2014). These are short (6-12 bp) DNA sequences, often called sequence motifs, that have particular specificity for DNA-binding proteins (i.e. transcription factors), which upon recruitment regulate transcription of associated genes. Enhancers are important transcription factor binding platforms (generally 200-500 bp) that contain multiple recognition sites (i.e. regulatory elements) for various transcription factors, thus facilitating cooperative binding of multiple transcription factors to individual enhancers (Panne, 2008).

Enhancers are central players in transcriptional regulation in eukaryotic cells (Andersson et al., 2015; Levo and Segal, 2014). Enhancers can be located upstream or downstream from their target gene promoters and are found in proximity as well as over long genomic distances (up to 1 Mb) from their associated genes (Bondarenko et al., 2003; Zabidi et al., 2015). Binding of cell-type specific factors to enhancers play a central role in driving lineage-specific gene expression (Calo and Wysocka, 2013).

Transcription factors bound to enhancers can regulate transcription of nearby or distantly located genes through physical interactions that involve chromatin loops between enhancers and their target genes (Krivega and Dean, 2012). Distant enhancers are brought to close proximity with their target gene promoter mainly by means of looping structures and 3D chromatin architecture (Gorkin et al., 2014). Transcription of most genes requires involvement of several different transcription

factors coming together, each binding to its appropriate sequence (i.e. enhancer), collaborating to activate gene expression by enabling recruitment of the RNA pol II transcription machinery to the core promoter (Ptashne and Gann, 1997; Struhl, 1995).

Enhancers can be identified by mapping the locations of transcription factor binding genome wide (Chen et al., 2008). When transcription factors are not known, certain characteristic features can be used to identify putative enhancer regions such as sensitivity to DNase treatment and nucleosome modifications (Buecker and Wysocka, 2012; Thurman et al., 2012). Enhancer activity of putative enhancers can be experimentally tested in luciferase reporter assays (Buckley et al., 2016).

Sequence variations in enhancer regulatory regions may perturb transcription factor binding and transcriptional regulation of their target genes, thus contributing to disease. Studies show that a significant portion of genomic variants associated with common diseases are concentrated in regulatory DNA regions (Maurano et al., 2012).

Interestingly, recent studies have explored the role of transcription factors at enhancers and the potential impact of enhancer templated non-coding RNAs (eRNAs) in transcription (Kim et al., 2010; Natoli and Andrau, 2012; Ong and Corces, 2012; Ren, 2010; Spitz and Furlong, 2012). It has been suggested that enhancers are similar to gene promoters in reference to DNA regulatory units, nucleosome positioning, and transcription factor binding, as well as functional similarities (i.e. initiate transcription by RNA pol II), although they operate at a slower rate than true gene promoters (Core et al., 2014; De Santa et al., 2010; Kim et al., 2010; Koch et al., 2011).

Transcriptional regulation by chromatin organization and chromatin regulators. In eukaryotic cells, genomic DNA is coiled around nucleosomes, which are

complexes of four different histone molecules (two H3-H4 and two H2A-H2B) appearing like ‘beads on a string’ (Felsenfeld and Groudine, 2003; Olins and Olins, 1974). The nucleosome is the fundamental unit of chromatin that can be further condensed to generate higher order chromatin (Gerchman and Ramakrishnan, 1987).

The dense packaging of genomic DNA hinders accessibility of promoters and enhances by RNA pol II and other factors involved in transcription. However, nucleosomes are controlled by diverse chromatin regulators that can dynamically remove, displace, or modify nucleosomes, allowing the transcription machinery access to the regulatory DNA sequences (Cairns, 2007; Hartley and Madhani, 2009; Shivaswamy et al., 2008; Workman and Kingston, 1992).

Two common histone variants that are used to replace core histones include H3.3 and H2A.Z (Altaf et al., 2010; Goldberg et al., 2010; Jin et al., 2009). This results in less stable nucleosomes (i.e. more flexible structure of the genome), thus facilitating dynamic processes such as DNA transcription, replication and repair (Diesinger and Heermann, 2009; Hubner et al., 2013; Jin and Felsenfeld, 2007). It is hypothesized that sites of nucleosomal hypermobility enable accessibility of DNA by transcription factors, which in turn recruit chromatin remodeling complexes (e.g. SWI/SNF, INO80) resulting in nucleosomal exclusion from enhancer DNA (Calo and Wysocka, 2013; Li et al., 2012).

Methods for Studying Chromatin Structure: ChIP-seq and FAIRE-seq

Organization of chromatin structure is an important aspect of transcriptional regulation in eukaryotic cells (Wallrath et al., 2008). Chromatin that is not bound around

nucleosomes is accessible by regulatory factors as well as the transcription machinery. Traditionally, these segments have been detected experimentally through their increased sensitivity to nuclease digestion, particularly DNase I, which is a biofeature of regulatory sequences (Gross and Garrard, 1988; Urnov, 2003). Other methods have been developed to analyze and characterize chromatin structure such as Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-seq), which is a technique used to identify DNA fragments that are bound by specific proteins, allowing precise mapping of protein-DNA interactions *in vivo* throughout the genome (Barski et al., 2007; Ho et al., 2016; Kellum and Schedl, 1991; Robertson et al., 2007; Zingher, 2003).

On the other hand, Formaldehyde-Assisted Isolation of Regulatory Elements followed by high-throughput sequencing (FAIRE-seq) is a technique used to isolate nucleosome-depleted DNA genome-wide (Giresi et al., 2007; Giresi and Lieb, 2009; Simon et al., 2012; Waki et al., 2011). Genomic regions depleted of nucleosomes correspond largely to DNase I hypersensitive sites, regulatory elements such as active promoters, enhancers, silencers and insulators, as well as histone modifications associated with active transcription (Consortium et al., 2007; Giresi et al., 2007; Kim et al., 2005a; Kim et al., 2005b; Koch et al., 2007).

Chromosome Conformation Capture (3C) is a technique used to analyze higher order chromatin and will be discussed in much detail in Chapter 3 of this manuscript (Dekker et al., 2013; Splinter et al., 2004).

Higher Chromatin Organization – Topological Domains

The recent development of Chromosome Conformation Capture (3C) and 3C-based techniques used for genome-wide analysis have provided insight into the long-range interactions and organization of chromatin *in vivo* (de Wit and de Laat, 2012; Gibcus and Dekker, 2013; Sanyal et al., 2011). One model of chromatin organization proposes the folding of chromatin into discrete topological domains, which are organized into fractional globules and further into chromosome territories, however still permitting interaction of distant genomic sites within or between chromosomes (Bancaud et al., 2012; Lieberman-Aiden et al., 2009).

Topological domains have been found to overlap significantly with cis-regulatory promoter-enhancer units in the mouse genome (Shen et al., 2012). The inter-chromosomal interactions for transcription regulation are not well-understood yet; however, it has been hypothesized that many genes are found in the periphery of the chromosomal territories co-localizing in interchromatin granules or transcription factories (Brown et al., 2008; Hu et al., 2008; Schoenfelder et al., 2010). These factories are aggregates of RNA pol II complexes with multiple regulatory sequences that synergistically increase transcriptional activity by accumulation of transcription factors needed for transcription (Edelman and Fraser, 2012).

Post-translational Modifications

Chromatin organization and transcription is greatly regulated by post-translational modifications of the histone components in the nucleosome (Heintzman et al., 2009). These modifications do not affect the nucleosome core, instead they alter the N-terminal

tails of the core histones by covalent attachment of different chemical groups (Santos-Rosa and Caldas, 2005). Chemical modifications include methylation, acetylation, phosphorylation, ubiquitination and sumoylation (Lee and Young, 2013).

Methylation of lysine 4 residue on histone H3 (H3K4me1) was one of the first histone modifications identified to be associated with distal enhancer regions, while tri methylation of the same residue (H3K4me3) was associated with active promoters and was found specifically at the 5' end of annotated genes (Heintzman et al., 2007). Acetylation of lysine 27 on histone H3 (H3K27ac) is enriched in regions containing active enhancers and promoters and is associated with transcriptional activity (Creyghton et al., 2010; Wang et al., 2008). The various histone modifications have a functional impact on regulating transcription of specific genomic regions (Guttman et al., 2009).

ENCODE. The ENCyclopedia Of DNA Elements (ENCODE) Project is a collaborative effort of many research groups worldwide that aimed to create a comprehensive catalogue of functional elements in the human genome (Consortium, 2004). The ENCODE project has produced a genome-wide chromatin landscape of the human genome based on mapping of genomic regions associated with transcription, transcription factor binding, histone modification and chromatin structure, thus enabling the profiling of regulatory elements and other functional domains in various cell types (Consortium, 2012).

Transcriptional Dysregulation in Disease

Our ability to better understand regulatory sequences is critical in refining our understanding of essential biological processes. Genetic variation in enhancer sequences may influence disease mainly through dysregulation of gene expression (Lee and Young, 2013).

Recent evidence emerging from GWAS indicates that a significant portion of disease-associated SNPs are found in regulatory DNA sequences and have been associated with many human diseases, including cancer (Ahmadiyeh et al., 2010; Bojesen et al., 2013; Bolton et al., 2010; Ghoussaini et al., 2008), Alzheimer's disease (Gaj et al., 2012), cardiovascular disease (Harismendy et al., 2011; Zhao et al., 2012), inflammatory lung disease (Han et al., 2012), and multiple sclerosis (Alcina et al., 2013). Alteration of regulatory sequences frequently underlie development of diseases as is evident in cancer, where many molecular and biological changes that are characteristic of disease are associated with changes in regulatory regions (Maurano et al., 2012; Sakabe et al., 2012; Sur et al., 2013).

Many disease-associated SNPs have cell type specific effects and are associated with disease in a specific cell type (Ernst et al., 2011). Enhancers operate in a cell-type specific manner and can explain how genetic variants such as SNPs contribute to tissue specific disease phenotypes (Maurano et al., 2012).

Mutations in transcription factors have also been known to promote tumorigenesis. As an example, *MYC* is one of the most frequently mutated genes found in human and animal cancers and overexpression of this transcription factor is associated with aggressive pathology and poor clinical outcomes (Cole, 1986;

Henriksson and Luscher, 1996; Marcu et al., 1992). Many tumor cells rely on the transcription factor MYC for their growth and proliferation (Littlewood et al., 2012). In tumor cells with overexpressed MYC, the transcription factor accumulates in the promoter regions of its target genes and recruits the P-TEFb elongation factor, producing increased levels of transcripts, known as transcriptional amplification (Lin et al., 2012; Nie et al., 2012).

In addition, alterations in co-activator complexes that interact with multiple transcription factors and facilitate DNA looping can contribute to carcinogenesis and other diseases. For instance, mutations in the Mediator coactivator complex has been found in various tumors, particularly in uterine leiomyomas and leiosarcomas and prostate cancer (Barbieri et al., 2012; Makinen et al., 2011).

Dysregulation of chromatin remodelers and epigenetic control are critical contributors to cancer pathogenesis. Loss-of-function mutations in genes coding for nucleosome remodeling proteins such as the SWI/SNF complex proteins and ARID1A have been associated with various cancer types, suggesting that defects in mobilizing nucleosomes are involved in cancer processes (Hargreaves and Crabtree, 2011; Tsai and Baylin, 2011; Wilson and Roberts, 2011).

Further research is needed to advance our understanding of transcriptional regulatory circuits that incorporate information about regulatory sequences and key molecular factors working together at regulatory sites to control gene transcription. Knowledge of sequence variation that contributes to disease is important in understanding mechanisms of disease, which can lead to improved diagnostic and therapeutic approaches to disease.

CHAPTER TWO:

IDENTIFYING FUNCTIONAL SNPS

Note to reader:

Parts of this section have been published in Buckley & Gjyshi et al. 2016 (Nat Protoc. 2016 Jan;11(1):46-60). These articles are Open Access and, when cited, can be reproduced for non-commercial use.

A second manuscript (Gjyshi et al.) has also been submitted for review and includes portions of this chapter.

Introduction

To investigate the mechanism by which SNP allele variation contribute to ovarian cancer pathogenesis, we performed a comprehensive functional analysis of the 8q24.21 ovarian cancer susceptibility locus. The most highly associated SNP (rs10088218) that tagged this region to be associated with epithelial ovarian cancer (EOC) was reported in a previous study, where the effect allele [A] was associated with decreased risk of ovarian cancer, especially in the high grade serous subtype (OR = 0.76; 95% CI 0.70-0.81; $p = 8 \times 10^{-15}$) (Goode et al., 2010). Two additional SNPs (rs1516982 and rs10098821, r^2 with rs10088218 = 0.64 and 0.80, respectively) also reached genome wide significance ($p \leq 5 \times 10^{-8}$). These SNPs map to a gene desert region that is devoid

of protein-coding DNA, located approximately 800 kb telomeric to the proto-oncogene *MYC*, a well-known transcription factor that is involved in many cancer processes including growth and proliferation, differentiation, transformation and apoptosis (Adhikary and Eilers, 2005).

Interestingly, the 8q24 chromosomal region is also associated with other types of cancers including prostate, breast, colorectal and bladder cancers (Easton et al., 2007; Eeles et al., 2008; Ghoussaini et al., 2008; Kiemeney et al., 2008; Tenesa et al., 2008; Thomas et al., 2008; Tomlinson et al., 2007; Yeager et al., 2007). These loci are located centromeric to *MYC* at distances between 30 kb to 800 kb. Growing evidence indicates that these loci contain transcription regulatory regions that influence *MYC* expression (Ahmadiyeh et al., 2010; Jia et al., 2009; Pomerantz et al., 2009; Sotelo et al., 2010; Wright et al., 2010).

In this study, we performed fine mapping of the 8q24.21 ovarian susceptibility locus followed by detailed functional analysis using progressively stringent criteria to select variants that contribute functionally and are most likely to drive association with risk. Considering that all the SNPs are located in non-coding regions, we hypothesized that these SNPs function within enhancer elements that modify transcriptional regulation of a gene involved in ovarian cancer pathogenesis (Freedman et al., 2011; Monteiro and Freedman, 2013). We developed a new method called Enhancer Scanning (Buckley et al., 2016) and also employed various methods to select functional SNPs with allele-specific effects on their respective enhancer elements and target genes, testing in OSE and fallopian tube epithelial (FTE) cells since both cell types may represent the cell of origin for high grade serous ovarian cancer.

Results

Fine Mapping Analysis and Identification of Candidate Functional SNPs

Note to reader: This part of the analyses was performed by Paul Pharoah and Jonathan Tyrer as part of COGS.

Genotyped data from four genotyping projects were used for the analyses: the COGS project and three genome-wide association studies (see methods). A dense set of fine-mapping SNPs spanning 2.06 Mb were selected for genotyping in COGS to cover the association for ovarian, breast and prostate cancer. Additional coverage was provided by imputation into a reference panel from the 1000 Genomes Project.

Overall, 43 studies from 11 countries provided data on 15,437 women diagnosed with invasive EOC, 9,627 of whom were diagnosed with serous EOC, and 30,845 controls from the general population. The final data set for analysis comprised genotypes for 14,043 SNPs of which 2,418 had been genotyped for the COGS samples. Genotype calls of genotyped and imputed SNPs were then tested for association with ovarian cancer (see methods).

The strongest association was for the genotyped SNP, rs1400482, with the major allele [G] being associated with an increased risk (OR = 1.18, 95% CI 1.13 – 1.23; p = 2.5×10^{-13}). The association was even stronger for serous ovarian cancer (OR = 1.29, 95% CI 1.22 – 1.36; p = 1.9×10^{-20}) so subsequent analyses were restricted to this

subtype. The relative risk for high-grade serous (OR = 1.27, 95% CI 1.19 – 1.35) was slightly smaller than that for low-grade serous (OR = 1.51, 95% CI 1.21 – 1.88) and low malignant potential serous tumors (OR = 1.53, 95% CI 1.30 – 1.80), a difference that was of borderline significance (p = 0.050). The regional association plot for serous ovarian cancer is shown in Figure 5A. The current study did not investigate association with survival. However, the initial study that identified the 8q24 locus for association with ovarian cancer reports no evidence of association of the tag SNP (rs10088218) with overall survival in all ovarian cancer cases or the serous subgroup (Goode et al., 2010).

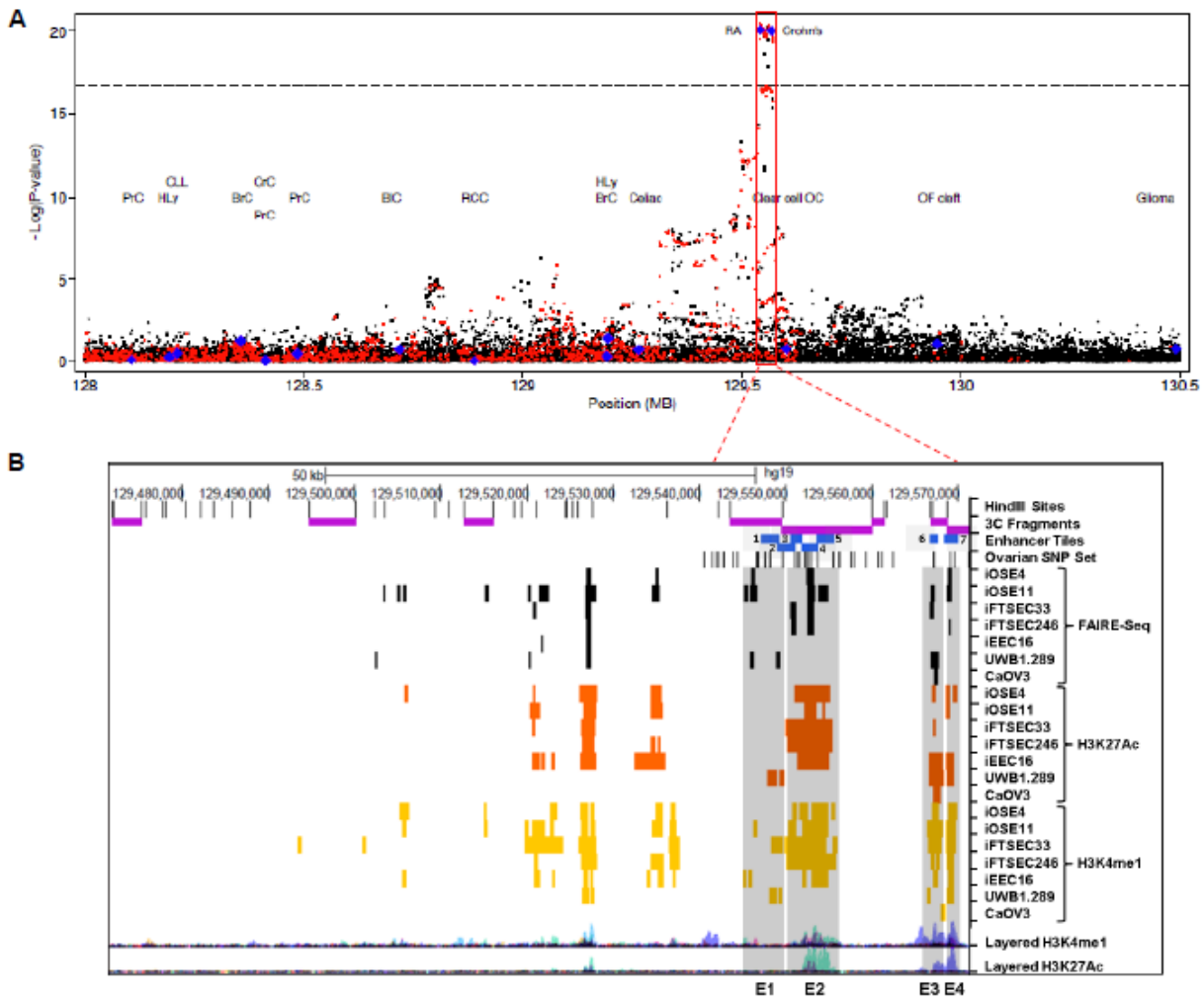


Figure 5. Fine mapping analysis delineates a 31 kb region of association with ovarian cancer in the 8q24.21 locus. ***A.*** Genotyped (red) and imputed (black) SNPs plotted by their position on chromosome 8 (x-axis) and $-Log_{10}$(p values) of association to ovarian cancer for the serous subtype (y-axis). The dashed black horizontal line represents the threshold for odds 100:1 for causal variant. Dashed red vertical lines represent the most highly associated SNPs delineating a 31 kb region. SNPs in blue are associated with the indicated traits in the GWAS catalog ($p \leq 1 \times 10^{-8}$). PrC, prostate cancer; CLL, chronic lymphocytic leukemia; HLy, Hodgkin's lymphoma; BrC, breast cancer; CrC, colorectal cancer; BlC, bladder cancer; RCC, renal cell carcinoma, Celiac disease; RA, rheumatoid arthritis; Crohn's disease; OF, orofacial cleft; Glioma. Note: The fine mapping analysis and figure were generated by Paul Pharoah. ***B.*** Regulatory landscape of the fine mapped 95 kb region. Tracks are indicated on the right. Ovarian SNP set contains 52 SNPs with odds 100:1 or better of being the true causal variant underlying this signal. *Hind*III sites were used to generate 3C fragments (purple). Tiles (blue) for enhancer scanning analysis were designed to cover regulatory biofeatures (FAIRE-seq, H3K27Ac, H3K4Me1) containing candidate functional SNPs (highlighted in gray).

Based on the log likelihood statistics from the association analyses there were 52 SNPs with odds of 100:1 or better of being the true causal variant underlying this signal (Table 2). These included rs10088218 reported by the original GWAS (Goode et al., 2010). This set of SNPs narrows the association signal down from a ~500kb region to a ~31kb region (chr8:129,540,464-129,571,140, Human Genome Browser hg19 assembly).

Table 2. Candidate functional SNP set

SNP ID[1]	Chr	Position (hg19)	100:1 Odds[2]	EAF[3]	R^2_iCOGS
rs1400482	8	129541931	1	0.132749	1
rs10088218	8	129543949	1.262379	0.132693	1
rs7814937	8	129541475	1.294664	0.132526	1
rs7010594	8	129542834	1.367521	0.132757	1
rs7010880	8	129542869	1.368137	0.132765	1
rs73374998	8	129561323	1.402351	0.132693	1
rs73374987	8	129557952	1.590266	0.134054	1
rs28399026	8	129544408	1.838685	0.132877	1
rs73376904	8	129564944	1.974172	0.133806	1
rs73375000	8	129561866	1.985561	0.133886	1
rs77247401	8	129561216	2.055666	0.134051	1
***rs10108517**	8	129549562	2.081732	0.13331	1
***rs13328411**	8	129551089	2.177986	0.133301	1
rs10089868	8	129562430	2.187478	0.133965	1
rs10089972	8	129562458	2.192189	0.133964	1
***rs17807628**	8	129551311	2.197346	0.133301	1
rs10678821	8	129560728	2.233789	0.134144	1
rs6982716	8	129557592	2.41996	0.134228	1
rs6982966	8	129557523	2.457394	0.134183	1
rs1516971	8	129542100	2.495778	0.133784	1
rs10098765	8	129559311	2.516322	0.134067	1
rs1400483	8	129547537	2.667526	0.133306	1
***rs6651252**	8	129567181	2.856799	0.133746	1
***rs1516974**	8	129548134	2.90766	0.13334	1
rs16903080	8	129546865	2.931012	0.133267	1
rs16903078	8	129546735	2.939388	0.133259	1
rs7839493	8	129546651	2.947629	0.133247	1
***rs10088755**	8	129551633	3.148106	0.133334	1
***rs10095481**	8	129549582	3.153455	0.133346	1

Table 2 (Continued). Candidate functional SNP set

SNP ID[1]	Chr	Position (hg19)	100:1 Odds[2]	EAF[3]	R²_iCOGS
***rs1516976**	8	129548258	3.164988	0.133343	1
***rs10113762**	8	129552202	3.422089	0.133345	1
rs1400484	8	129547636	3.433046	0.133312	1
***rs1516975**	8	129548193	3.572978	0.133327	1
***rs16903081**	8	129548309	3.640781	0.132963	1
***rs938650**	8	129552540	3.690645	0.133305	0.999
***rs6651253**	8	129567292	4.040629	0.133752	1
***rs938648**	8	129552491	4.236547	0.133352	1
rs57593539	8	129567515	4.281485	0.133569	1
rs16903097	8	129556356	4.327115	0.134047	1
***rs938649**	8	129552534	4.3754	0.133281	0.999
***rs13328404**	8	129552456	4.546244	0.133339	1
***rs201242438**	8	129552855	4.585047	0.137634	0.975
rs6470637	8	129556163	4.80016	0.134122	1
***rs1561925**	8	129569033	5.379514	0.132855	1
***rs2392944**	8	129552856	5.80866	0.132756	0.875
***rs2011527**	8	129555532	6.095788	0.134128	1
***rs10103637**	8	129553697	6.104329	0.134207	1
***rs10103640**	8	129553703	6.577806	0.134195	1
***rs938651**	8	129555443	6.751749	0.134143	1
rs28455755	8	129571140	7.248904	0.132516	0.979
***rs2165806**	8	129569551	7.30311	0.132827	1
rs16903065	8	129540464	8.631336	0.132642	1

[1]Asterisk (*) and **bold** font indicate the 24 SNPs that overlap with ovarian biofeatures.
[2]SNPs with 100 to 1 odds or better of being the causal variant determined by log likelihood statistics by comparing the likelihood of each SNP from the association analysis with the likelihood of the most strongly associated SNP (rs1400482).
[3]EAF: effect allele frequency

The association analyses was repeated for each SNP adjusting for rs1400482 in order to identify additional, independent signals in the region at $p < 10^{-5}$. Two further independent signals were identified (Figure 6): rs74559819 (RAF = 0.037; OR_{adj} = 1.26, 95%CI 1.15 – 1.38; $p = 4.4 \times 10^{-7}$) and rs4524749 (RAF = 0.68; OR_{adj} = 1.09, 95%CI 1.05 – 1.13; $p = 7.9 \times 10^{-5}$). These are uncorrelated (absolute correlation < 0.02) with the top hit rs1400482 and with each other. Rs74559819 is located within the *PVT1*

spanning region and lies in a DNAse I hypersensitive cluster (in 125 cells from ENCODE) and H3K4me1 region in OSE cells. However the absence of H3K27Ac marks suggests that it is not an active enhancer (Figure 6A). Rs4524749 is located proximal to the *MYC* promoter (Figure 6B).

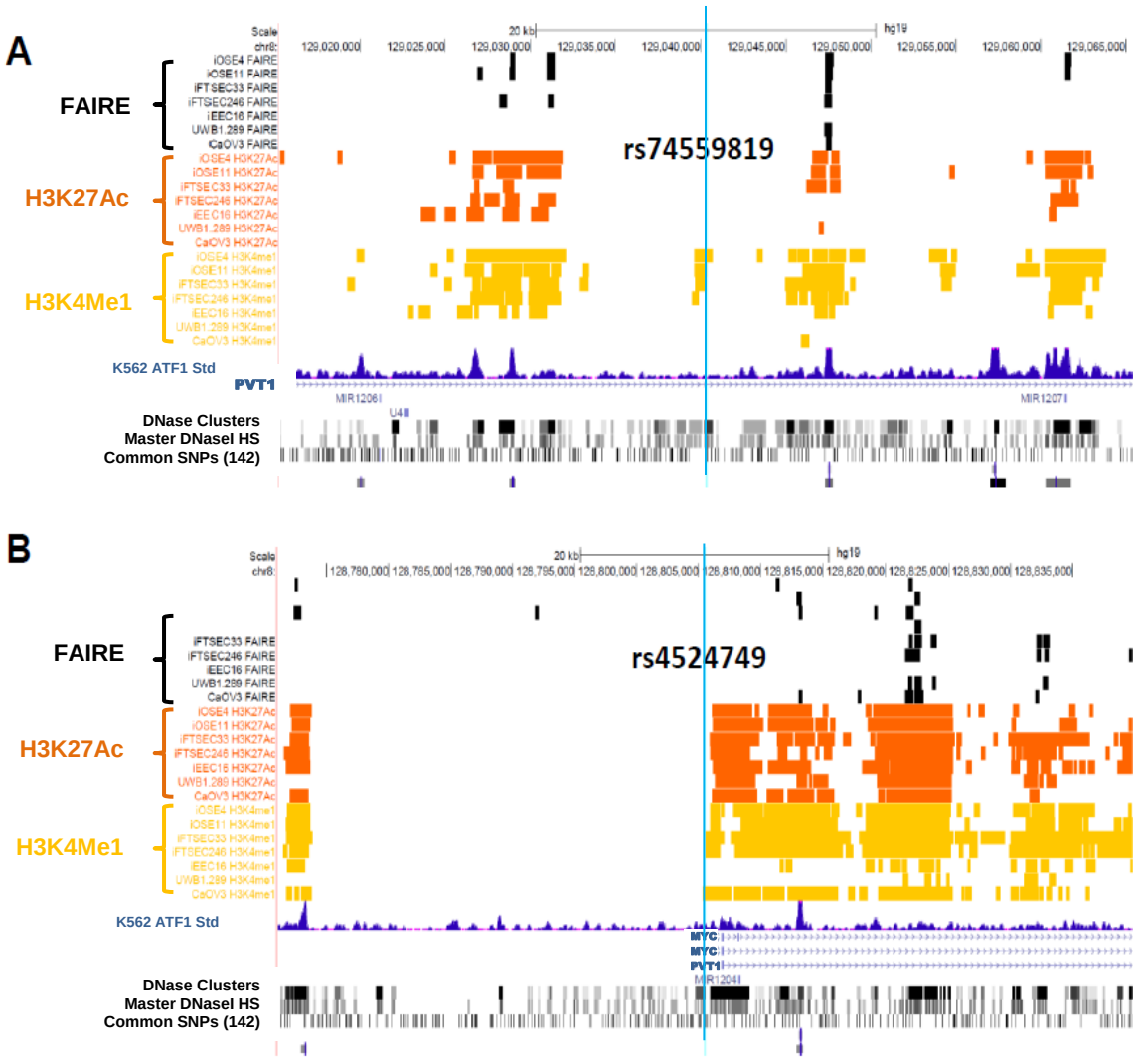


Figure 6. Independent signals revealed by conditional analysis. Chromatin landscape of SNPs with independent signals in the region at $P < 10^{-5}$. ***A.*** rs74559819 (RAF = 0.037; OR_{adj} = 1.26, 95%CI 1.15 – 1.38; p = 4.4 x 10^{-7}). ***B.*** rs4524749 (RAF = 0.68; OR_{adj} = 1.09, 95%CI 1.05 – 1.13; p = 7.9 x 10^{-5}). SNP location is indicated with blue vertical line. Note: Conditional analysis was performed by Jonathan Tyrer and Paul Pharoah.

The association of genotyped and imputed SNPs in the region with the clear cell and endometrioid subtypes of epithelial ovarian cancer was also evaluated. There was little evidence for association with endometrioid ovarian cancer. There was a single SNP, rs2165805, that achieved genome-wide significance for association with clear cell ovarian cancer (Figure 7: OR = 1.46, 95% CI 1.28 – 1.66; $p = 8.2 \times 10^{-9}$). This SNP was not correlated with the nearby SNPs associated with serous ovarian cancer and represents an independent association signal for the region. Additionally, the clear cell SNP is not located in a region with any specific chromatin features.

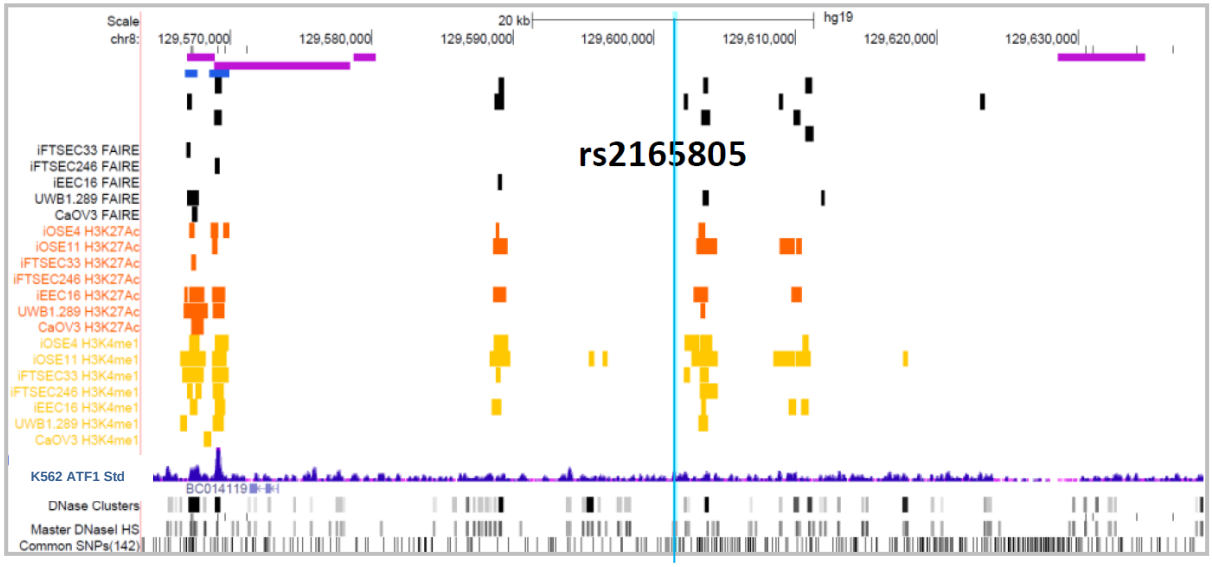


Figure 7. Independent signals revealed by histological subtype analysis. rs2165805 is associated with clear cell ovarian cancer (OR = 1.46, 95% CI 1.28 – 1.66; $p = 8.2 \times 10^{-9}$). SNP location is indicated with blue vertical line. Note: Subtype analysis was performed by Jonathan Tyrer and Paul Pharoah.

Mapping SNPs to Enhancer Elements

The 52 candidate causal SNPs for serous OC risk are all located in non-protein coding regions. We evaluated the overlap between these SNPs and regions displaying features of active regulatory elements using Formaldehyde Assisted Isolation of Regulatory Elements followed by sequencing (FAIRE-seq; indicative of open chromatin), and histone modifications H3K27Ac and H3K4Me1 in a panel of ovarian normal and cancer cell lines (Coetzee et al., 2015). This analysis identified four regions (Figure 5B; labeled E1-E4) in which 24 candidate causal SNPs overlap with recognized chromatin features in immortalized normal ovarian surface epithelial (iOSE4, iOSE11) (Lawrenson et al., 2009) and fallopian tube surface epithelial cells (iFTSEC33, iFTSEC246) (Coetzee et al., 2015) (Figure 5B &Table 2). Very few differences are observed in the overlapping of SNPs with regulatory biofeatures in ovarian versus fallopian tube cells, with no apparent segregation of SNPs according to cell type, giving no indication whether one of these cell types is most likely involved in ovarian cancer compared to the other.

Functional Analysis of SNPs

Identifying Functional SNPs by Enhancer Scanning

To validate the predicted enhancer regions in ovarian cells, a reporter assay was used to scan the four regions for enhancer activity (Buckley et al., 2016; Pharoah et al., 2013). First, seven genomic tiles were generated (Figure 5B; blue boxes) and cloned in forward and reverse orientations upstream of a basal SV40 promoter driving luciferase

expression. Each construct was transfected into iOSE4^{cMYC} ovarian cells (*Tert*-immortalized and partially transformed with *MYC*) (Lawrenson et al., 2009) and luciferase levels were measured 24 h post transfection. Four tiles (T3, T4, T6, and T7), in regions E2, E3 and E4 containing a total of 15 candidate functional SNPs, displayed significant enhancer activity in at least one orientation (Figure 8).

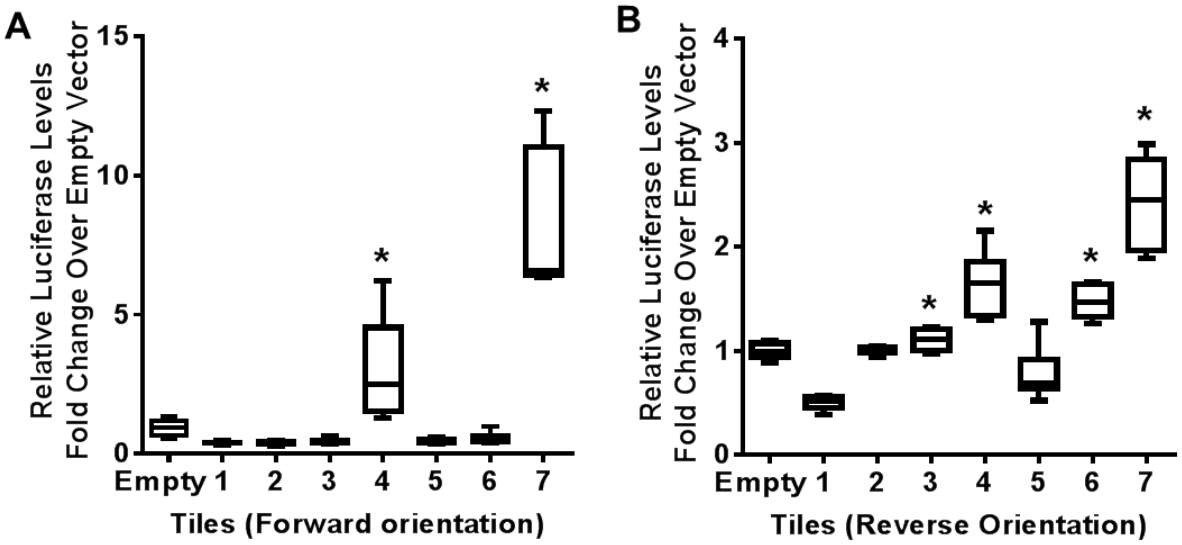


Figure 8. Enhancer scanning in iOSE4^{cMYC} ovarian cells. Box and whisker plots of luciferase activity for tiles in forward (***A***) and reverse (***B***) orientation containing the reference allele for each SNP. ***** denotes significant (unpaired t-test; $p \leq 0.05$) differences in relation to the control (empty vector).

Identifying Functional SNPs with Allele-Specific Enhancer Activity

To assess the effect of both alleles for each of the 15 SNPs, first we generated tiles containing the minor allele for each SNP by using site-directed mutagenesis. The enhancer activity of tiles containing the major versus the minor allele was tested in a similar set up as the enhancer scanning assay. Eleven candidate functional SNPs

displayed significant differences between major and minor allele (unpaired t-test; $p \leq 0.05$) (Figure 9) (Table 3).

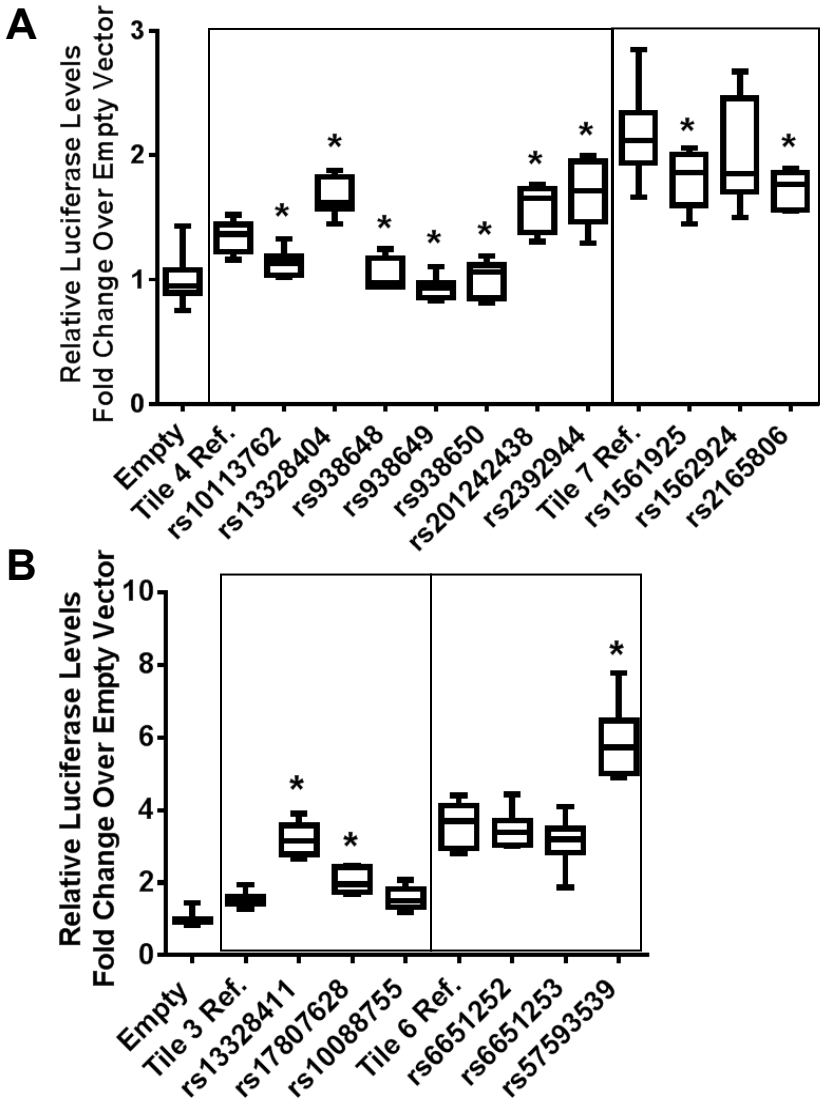


Figure 9. Allele-specific activity in iOSE4^{cMYC} ovarian cells. Box and whisker plots of luciferase activity for tiles that were tested for allele-specific activity in forward (***A***) and reverse (***B***) orientation. ***** denotes significant (unpaired t-test; $p \leq 0.05$) differences in relation to the reference allele. Boxes separate SNPs according to the tiles they are located in for comparison with the effect of the reference allele in their respective tiles.

Table 3. Proposed functional SNPs at the 8q24.21 locus

Region	SNP	Major Allele FWD[a]	Minor Allele	MAF[b]	Serous X^2	Enhancer Scanning	EMSA[c]	3C[d]
E1	rs13328411	T	A	0.0889	84.8	+	–	–
E1	rs17807628	T	C	0.0895	84.8	+	+	–
E1	rs10088755	G	A	0.1124	83.9	–	–	–
E2	rs10113762	T	C	0.1122	83.5	+	+	+
	rs13328404	A	G	0.1120	83.2	+	+	+
	rs938648	T	G	0.1118	83.4	+	+	+
	rs938649	G	T	0.1098	83.2	+	+	+
	rs938650	G	A	0.1120	83.3	+	+	+
	rs201242438	C	del	NA	?	+	+	+
	rs2392944*	C	T	0.1713	62.7	+	–	+
E3	rs6651252	T	C	0.1524	83.9	–	–	+
	rs6651253	G	C	0.1522	84.0	–	–	+
E4	rs1561925*	C	T	0.1508	82.7	+	+	+
	rs1561924*	G	A	0.1508	83.5	–	+	+
	rs2165806*	G	C	0.1516	82.1	+	+	+

(a) All SNP alleles are shown in forward human genome orientation. SNPs marked with (*) are originally described in reverse in dbSNP. Throughout the paper we refer to SNP alleles using their forward orientation. (b) MAF, minor allele frequency; (c) EMSA, electrophoretic mobility shift assay; (d) 3C, chromosome conformation capture, indicates whether the 3C probe that contains the SNP was found to interact with target gene promoters.

Identifying SNPs with Allele-Specific Binding of Nuclear Proteins by EMSA

To further investigate allele-specific effects, we performed Electrophoretic Mobility Shift Assays (EMSA) to investigate differential binding of nuclear proteins of major versus minor allele for each of the 15 SNPs. We used nuclear extracts of iOSE4^{cMYC} ovarian cells incubated with DNA probes containing either major or minor

allele for each SNP. Ten SNPs exhibited differential binding of nuclear proteins between the reference and effect allele (Table 3) (Figure 10). Altogether, we identified nine SNPs in regions E1, E2, and E4 that displayed reproducible allele-specific activity in both assays (enhancer scanning and EMSA). Taken together with the fine mapping data these results indicate that these nine SNPs are the most likely candidate causal SNPs at this locus (Table 3).

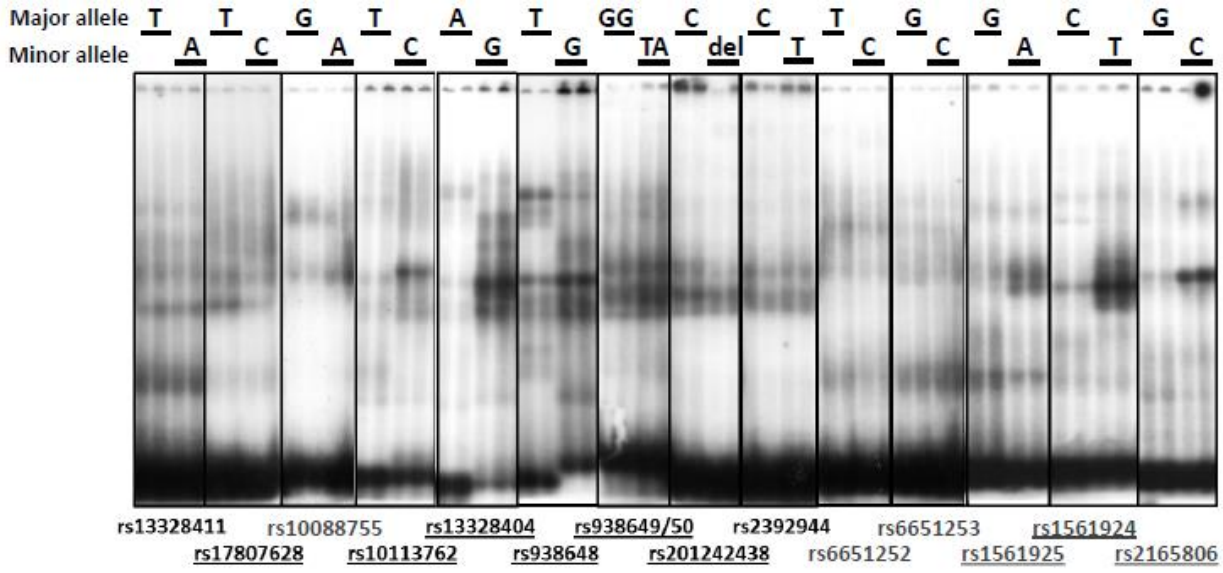


Figure 10. Protein binding profiles of candidate functional SNPs at 8q24.21 by EMSA in iOSE4^{cMYC} cells. Assays were performed in technical duplicates (run in two lanes) with probes containing either the reference or effect allele for SNPs within tiles with enhancer activity. Underlining indicates SNPs with allele-specific binding of nuclear proteins in both technical duplicates and in two independent experiments.

Network of Distant Enhancers at the 8q24 Locus

In order to determine the cell type specificity of different enhancer regions in the 8q24 locus that have been associated with various cancers, we generated tiles for the enhancer scanning assay containing SNPs associated with prostate (Prostate region 1;

rs4242382), breast (rs13281615), and colorectal (rs6983267) cancer risk, with previous data for functional interaction with *MYC* (Ahmadiyeh et al., 2010; Pomerantz et al., 2009; Sotelo et al., 2010; Wright et al., 2010). We also included ovarian SNPs identified in this study (tiles 3, 4, 6, and 7 covering regions E2-4) (Figure 11A). Tiles were generated for each cancer locus and enhancer activities were tested in iOSE4^{cMYC} (ovarian), Cal51 (breast), HCT116 (colorectal), and LNCaP (prostate) cell lines for comparison (Figure 11B).

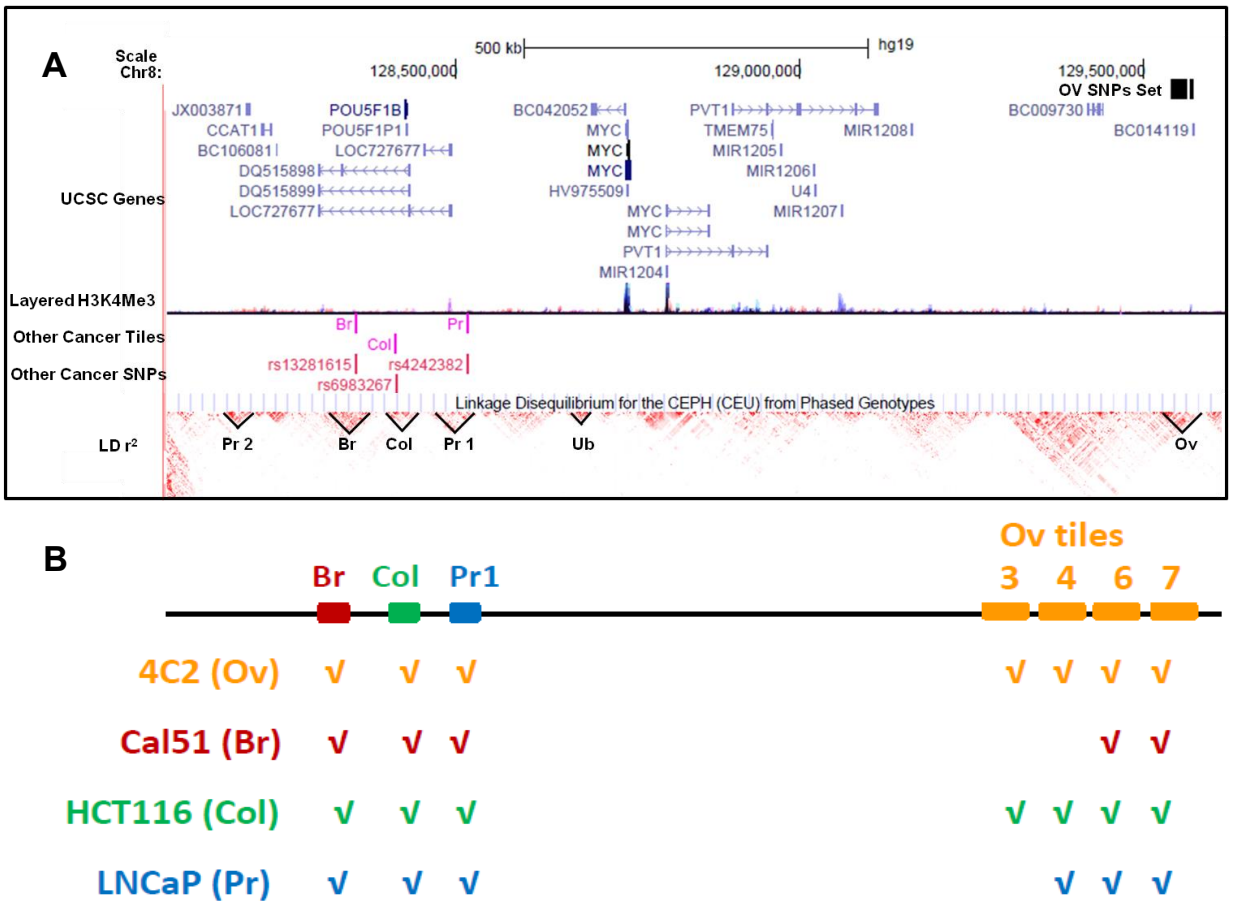


Figure 11. 8q24 region including different cancer susceptibility loci. *A*. Six linkage disequilibrium blocks are highlighted indicating the location of association to Prostate (regions 1 and 2 (Pr1 and Pr2), Breast (Br), Colorectal (Col), urinary bladder (Ub), and Ovarian (Ov) cancer risk. Genomic tiles (pink) containing the three SNPs (red) that were tested in luciferase assay have been shown to have functional activity in previous studies (Ahmadiyeh et al., 2010; Pomerantz et al., 2009; Sotelo et al., 2010; Wright et al., 2010). ***B*.** Summary of enhancer activity by transfection in iOSE4^{cMYC}. Check mark indicates enhancer activity.

Excepting ovarian tiles 3 and 4 (Region E2) all tiles displayed enhancer activity in all three cell lines (Figure 11B and Figure 12). Ovarian tile 3 was negative in Cal51 breast cells and LNCaP prostate cells. Ovarian tile 4 was negative in Cal51 breast cells, but positive in the other three cancer cell lines. Interestingly, regions defined by SNPs associated with colorectal, breast, and prostate do not present a chromatin environment, as judging by presence of FAIRE-Seq, H3K4me1 and H3K27Ac marks in

ovarian cells. Taken together, these results indicate that different tissues express factors required to activate these enhancer regions and suggest that tissue specificity is conferred by the cell type specific chromatin microenvironment, which dictates the accessibility of the underlying DNA containing the regulatory sequence by the appropriate TF that binds that particular motif.

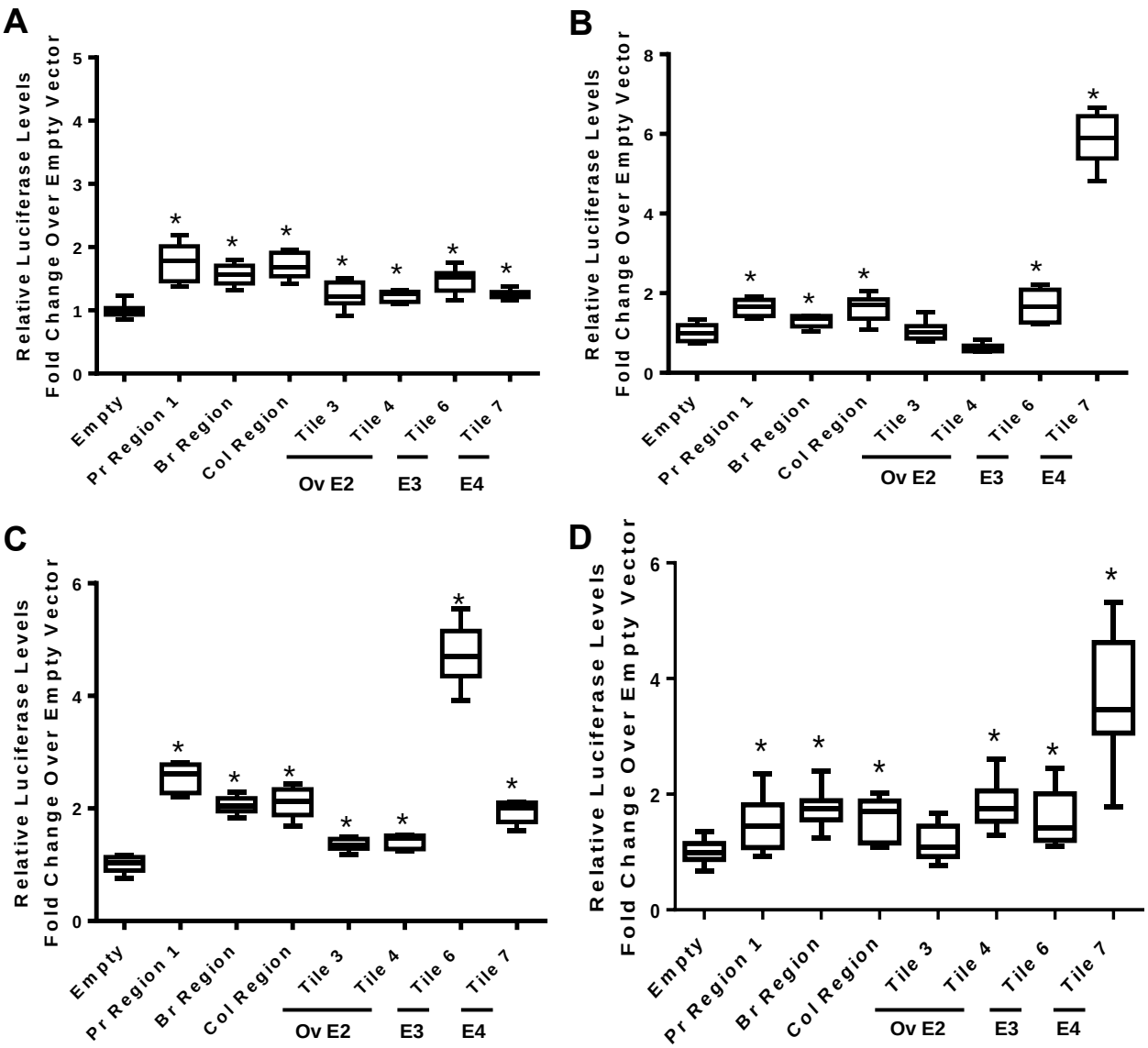


Figure 12. Enhancer activity of different 8q24 cancer regions in ovarian cells. Two replicates of experiments represented in box and whisker plots of luciferase activity for tiles in reverse orientation transfected in iOSE4^{cMYC} (***A***), Cal51 (***B***), HCT116 (***C***), and LnCap (***D***) cell lines. * denotes significant (unpaired t-test; $p \leq 0.05$) differences in relation to the control (empty vector).

Summary

In this chapter, we started from fine mapping analysis of the 8q24.21 ovarian cancer susceptibility locus, which identified a 95 kb region of association (chr8:129,474,065-129,569,064, Human Genome Browser hg19 assembly) (Figure 5). Additional statistical analyses identified 52 SNPs with odds of 100:1 or better of being the candidate functional SNP (Table 1 and Figure 5A), narrowing down the signal of association to a ~32 kb region. Considering that all these SNPs fall in non-coding regions, we hypothesized that these SNPs exert their functions through enhancer elements, regulating transcription of distant target genes that are involved in ovarian cancer pathogenesis.

To test this hypothesis, we identified 24 SNPs, residing within four distinct enhancer regions (E1-E4, Figure 5B and Table 3), spanning approximately 11 kb. Tiles designed to cover the predicted enhancer regions, containing 24 SNPs, were tested for enhancer activity in our enhancer scanning assay, which identified four tiles that displayed enhancer activity in ovarian cells (Tile 3, 4, 6 & 7, Figure 5B). These tiles contained 15 candidate causal SNPs that are most likely implicated in ovarian cancer processes. Luciferase reporter assays and EMSAs were performed to investigate the allele-specific activity of each of those SNPs, narrowing down the number of candidate causal SNPs to 9 variants that displayed allele-specific activity in both assays (Table 3).

Furthermore, we also tested the tissue specificity of various enhancer regions that have been associated with colorectal, breast, and prostate cancer by designing tiles (Figure 11A) to test in the reporter assay along with the four active ovarian tiles we identified in enhancer scanning. All tiles were tested in all four cancer

cell lines (Figure 11 and 12). With the exception of ovarian tiles 3 and 4 (Region E2), all tiles displayed enhancer activity in all three cell lines (Figure 11B and Figure 12) despite the lack regulatory biofeatures of the other cancer types in ovarian cells. This shows that different tissues express factors required to activate these enhancer regions.Tissue specificity is conferred by the pool of available TFs that are present in the cell as well as the specific chromatin microenvironment (i.e. epigenetic modifications), which controls the accessibility of the regulatory DNA sequences by the appropriate transcription factors.

Materials and Methods

Note to reader: The first two parts in this section have been analyzed by Paul Pharoah and Jonathan Tyrer as part of COGS.

Fine Mapping and SNP Selection

Fine-mapping of the region was carried out as part of the Collaborative Oncology Gene-Environment Study (COGS) with the aims of identifying the SNPs most likely to mediate the causal associations at the 8q24 locus harboring multiple association signals for breast, ovarian and prostate cancer and to identify novel association signals within the same region. The association signals that had been reported at the time the fine-mapping SNPs were selected were: rs10088218 for ovarian cancer, rs13281615, rs13262406 and rs1562430 for breast cancer; rs12543663, rs10086908, rs1016343, rs13252298, rs6983561, rs620861, rs6983267, rs10090154, rs16901979, rs13254738 and rs7000448 for prostate cancer.

All SNPs within a 2.06 Mb interval at 8q24.21 (chr8: 127630906 -129693334) were identified from the 1000 Genomes Project (1000G) CEU (April 2010) (Genomes Project et al., 2010) and Hapmap III (International HapMap, 2003). Only variants with the minor allele called at least twice in the 1000GP data and with an Illumina Design score > 0.8 were included. All variants correlated with the tagging SNPs ($r^2 > 0.1$) were selected for genotyping. Additional tagging SNPs ($r^2 > 0.9$) were selected to capture the remaining SNPs in the region. Genotyping was done using a customized Illumina iSelect genotyping array (iCOGS) designed for the Collaborative Oncology Gene-Environment Study.

The iCOGS array was used to genotype cases and controls from 43 constituent studies of OCAC, as previously described (Pharoah et al., 2013). Two thousand four hundred and eighteen SNPs across the 8q24 region were successfully genotyped. In addition to the OCAC iCOGS data, genotype data were used from previously conducted ovarian cancer genome-wide association studies. We excluded samples if they were not of European ancestry, if they had a genotyping call rate of <95%, if they showed low or high heterozygosity, if they were not female or had ambiguous sex or if they were duplicates (cryptic or intended). One individual was excluded from each pair of samples found to be first-degree relatives, and duplicate samples between the iCOGS stage and any of the GWAS were excluded from the iCOGS data. SNPs were excluded if they were mono-morphic, had a call rate of < 95%, showed evidence of deviation from Hardy-Weinberg equilibrium or had low concordance between duplicate pairs. For the GWAS, we also excluded rare SNPs (MAF < 1% or allele count < 5, respectively). The final data set comprised genotype data for 11,069 cases and 21,722 controls from

COGS ('OCAC-iCOGS'), 2,165 cases and 2,564 controls from a GWAS from North America ('US GWAS') (Permuth-Wey et al., 2011), 1,762 cases and 6,118 controls from a UK-based GWAS ('UK GWAS') (Song et al., 2009), and 441 cases and 441 controls from the Mayo Clinic. All subjects included in this analysis provided written informed consent as well as data and blood samples under ethically approved protocols. Overall, studies from 11 countries provided data on 15,437 women diagnosed with invasive EOC, 9,627 of whom were diagnosed with serous EOC, and 30,845 controls.

In order to improve coverage we imputed genotype data for SNPs that had not been genotyped. We performed imputation separately for the iCOGS samples and each of the GWAS. We imputed variants from 1000 Genomes Project data using the v3 April 2012 release (Genomes Project et al., 2010) as the reference panel using the IMPUTE2 software (Howie et al., 2009) without pre-phasing. We changed some of the default parameters in the imputation procedure in order to increase imputation accuracy. These included an increase in the MCMC iterations to 90 (out of which the first 15 were used as burn-in), an increase in the buffer region to 500 kb and an increase in the number of haplotypes used as templates when phasing observed genotypes to 100. SNPs were included for analysis if the minor allele frequency was greater than 1% and the imputation r^2 greater than 0.3. The final data set comprised genotypes for 8,022 SNPs of which 2,342 had been genotyped for the COGS samples.

Association Analysis

We evaluated the association between genotype and disease using logistic regression analyses for each SNP. The analysis was adjusted for study and for

population substructure by including the eigenvectors of the first five ancestry-specific principal components as covariates. We used the same approach to evaluate SNP associations with serous ovarian cancer after excluding all cases with any other or unknown tumor subtype. For imputed SNPs, we used expected dosages in the logistic regression model to estimate SNP effect sizes and P values. We carried out analyses separately for OCAC-iCOGS samples and the three GWAS and pooled data thereafter using a fixed-effects meta-analysis. All results are based on the combined data from iCOGS and the three GWAS. In order to identify a set of potentially causal variants, we excluded SNPs with a likelihood of being causal of less than 1:100, by comparing the likelihood of each SNP from the association analysis with the likelihood of the most strongly associated SNP (rs1400482). We used custom written software for the analysis.

Genome Browser

The Genome Browser (https://genome.ucsc.edu/) is a publically available database developed and maintained by the University of California Santa Cruz (UCSC) that contains the reference sequence of the human genome and also provides portals to the ENCODE data. For the *in silico* analysis we utilized the GRCh37/hg19 human assembly available in this database. A personal session was created containing the region of interest on chromosome 8q24. Datasets for ovarian biofeatures including ChiP-Seq for H3K4me1 and H3K27ac, as well as FAIRE-seq were made available on the Genome Browser by the Coetzee laboratory (Coetzee et al., 2015). Custom tracks were generated for the candidate SNP set, enhancer scanning tiles, as well as the

expected 3C fragments (used in Chapter 3). An example of a bed file used to generate a custom track for the tiles used in the enhancer analysis is shown in Table 4. A bed file contains information about the chromosome of interest, genome browser position that will be displayed on the browser window, name and description for the custom track with an option to add color specification in RGB format, as well as SNP/tile coordinates and names included in each custom track.

Table 4. Example of a bed file used to generate a custom track for enhancer tiles

```
browser position chr8:129,471,565-129,571,564
track name="Tiles" description="Amplicons for enhancer analysis"
visibility=2 color=32,91,229
chr8      129547233 129549217 1
chr8      129549146 129551041 2
chr8      129550538 129551890 3
chr8      129551871 129553687 4
chr8      129553668 129555687 5
chr8      129566784 129567658 6
chr8      129568495 129569898 7
```

Bed files were uploaded to the genome browser by clicking on the following tabs: 1) manage custom tracks, 2) add custom tracks, 3) chose file, 4) submit. The custom tracks are added to the genome browser window along with all the other default tracks, thus enabling us to visualize overlapping features.

Cell Lines

Association analysis identified the most significant SNP, rs1400482, which had a stronger association when the analysis was restricted to high-grade EOC tumors; therefore, the subsequent analyses were restricted to this subtype. Based on the

hypothesis that serous EOC may originate both from ovarian surface and fallopian tube epithelial cells, we used cell lines and datasets that were derived from these cell types.

Experiments for the functional analyses (reporter assays and EMSAs) were conducted in normal epithelial ovarian cell line immortalized with human telomerase (*hTERT*) and partially transformed with *MYC* (iOSE4^{cMYC}) as a model for human serous EOC (Lawrenson et al., 2009; Lawrenson et al., 2010). Experiments for the chromatin conformation capture (3C) assay (see Chapter 3) were conducted in normal ovarian surface epithelial cells (iOSE11) and *hTERT*- immortalized normal fallopian tube surface epithelial cells (iFTSEC283) (Lawrenson et al., 2009; Lawrenson et al., 2010). Cells were cultured in a base medium composed of MCDB105 and Medium 199 (Sigma-Aldrich) at a 1:1 ratio, supplemented with 15% Fetal Bovine Serum (FBS) (Sigma-Aldrich), 34 µg/mL Bovine Pituitary Extract (BPE) (Thermo-Fisher Scientific), 10 ng/mL Epidermal Growth Factor (EGF) (Thermo-Fisher Scientific), 5 µg/mL insulin, and 0.5 µg/mL hydrocortisone (Sigma-Aldrich) to make the complete growth medium. The iOSE4^{cMYC} cells were cultured in complete growth medium containing 2 µg/mL Blasticidin (Thermo-Fisher Scientific).

Datasets for ovarian biofeatures, including FAIRE-Seq, H3K4me1 and H3K27ac27 (Coetzee et al., 2015), contain two immortalized normal ovarian surface epithelial cell lines (iOSE4 and iOSE11), two immortalized normal fallopian tube surface epithelial cells (iFTSEC33 and iFTSEC246), and two ovarian cancer cell lines, CaOV3 exhibiting molecular profiling characteristic of high-grade serous carcinoma (Domcke et al., 2013), and UWB1.289, which is a BRCA1-null cell line (DelloRusso et al., 2007).

Reporter assays for the network analysis of distant enhancers were conducted in iOSE4^{cMYC} cells described above, as well as in HCT116, a colorectal carcinoma cell line; Cal51, a breast cancer cell line; and LNCaP, a prostate carcinoma cell line, all of which were obtained from ATCC and were cultured according to manufacturer's specifications.

Enhancer Scanning

Genomic tiles were PCR amplified from genomic DNA obtained from iOSE4^{cMYC} cells using primers containing att B recombination sites (Table 5). PCR amplification was performed with HotStar Taq polymerase (Qiagen) followed by gel purification (Qiagen). Gateway cloning technology (Gateway BP Clonase II from Life Tech) was used to clone the PCR amplicons into pDONR 221. Inserts were screened by restriction digest. LR reactions (Gateway LR Clonase II from Life Tech) were used to clone the genomic tiles in forward and reverse orientation upstream of the SV40 promoter driving firefly luciferase expression in the pGL3-Promoter-LR vector (Promega) which was obtained through the Gateway Vector Conversion System (Life Tech). For the reporter assay, we used iOSE4^{cMYC} cells plated at 5 x 10^3 per well in 96-well plates. pGL3 vectors containing individual tiles were co-transfected with the pRL-CMV vector serving as internal control using Fugene HD (Promega) at a 3:1 ratio of Fugene HD (volume in ul) to DNA (ng).

For the network analysis of the distant enhancers at the 8q24 locus, Cal 51, LNCaP and HCT 116 cells were plated at 1 x 10^4 per well in 96-well plates. pGL3 vectors containing individual tiles were co-transfected with the pRL-CMV vector using

Lipofectamine 3000 (Life Tech) at a 2:1 ratio of Lipofectamine 3000 (volume in ul) to DNA (ng).

Luciferase readings were obtained 24 h post transfection using the Dual-Glo Luciferase Assay System (Promega). Luciferase activity was calculated by normalizing against the internal control and level of significance was determined by comparing against luciferase activity in the empty vector (unpaired t test; p value ≤ 0.05).

Table 5. Primers for enhancer analysis

Primer Name	Sequence (5' to 3')
Tile 1 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTCACGGCTATGAAGAAATACC
Tile 1 R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGCATGCAAAGGTGCTGTATTA
Tile 2 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTGTTCACCTGAACCATGACTT
Tile 2 R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGATGGGAAGGGAATAAGATGT
Tile 3 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTCCAGTTTCATGCAGCTTTCA
Tile 3 R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGTAAGTGTGGAAGGTGGTGCA
Tile 4 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTTGCACCACCTTCCACACTTA
Tile 4 R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGGAGCACATTGCAGTTGGAAA
Tile 5 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTTTTCCAACTGCAATGTGCTC
Tile 5 R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGAAAACAGCTGGGAGGAAGGT
Tile 6 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTTATGCCTCGGTTTCCTCATC
Tile 6 R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGCCCAGGTAGAGGGAATAGCC
Tile 7 F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTCCTGCTGTATGCCGAGTTTT
Tile 7 R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGTGTCCTGAGAGTGGAGGCTT
BrRs1328165F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTAACCGGTCTTTTCAGTTTATGG
BrRs1328165R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGTTCACCATCCTTGTCCTTGG
ColRs6983267F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTAGACACCAAGAGGGAGGTATCA
ColRs6983267R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGTTACAGCCTGCTGGGAAAGT
Pr1Rs4242385F	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCGTCAGTCACATCCCACCCAACT
Pr1Rs4242385R	GGGGACCACTTTGTACAAGAAAGCTGGGCTCGAGACCAGTCCTGTCCTGTCTCC

Site-Directed Mutagenesis

To test the allele-specific activity of SNPs with the reporter assay, we used site-directed mutagenesis to switch the SNPs from major to minor allele. This was achieved using the QuickChange XL Site-Directed Mutagenesis kit (Agilent), following the manufacturer's protocol. Primers were designed utilizing the freely accessible QuickChange Primer Design Program (http://www.genomics.agilent.com/primerDesignProgram.jsp) available in the manufacturer's website (Table 6). Vectors containing tile 3, 4, 6 and 7 in reverse orientation were used as templates to generate the minor alleles for their respective SNPs. Luciferase activity was measured as described in the Enhancer Scanning method described above. The allele specific activity of each SNP was compared against the template vector containing the major SNP allele(s) in corresponding tile (unpaired t test; p value ≤ 0.05).

Table 6. Primers for site-directed mutagenesis

Tile	SNP	Primer Name	Strand	Primer Sequence
45	rs13328411	45-1rs13328411t-a Rev		ttggtgagcctcttgactgactgagttgaaggaat
		45-1rs13328411t-a Fwd	+	attccttcaactcagtcagtcaagaggctcaccaa
	rs17807628	45-2rs17807628t-c Rev		tttcccttagatttccagtagcaaattagcccacactgatg
		45-2rs17807628t-c Fwd	+	catcagtgtgggctaatttgctactggaaatctaagggaaa
	rs10088755	45-3rs10088755g-a Rev		gtatgtgtcaataaacagttctttattgtctacataattcatttttttcccagaatt
		45-3rs10088755g-a Fwd	+	aattctgggaaaaaaatgaattatgtagacaataaagaactgtttattgacacatac
46	rs10113762	46rs10113762t-c Fwd	+	aaaggccccagcggttgccttagataatacaaacac
		46rs10113762t-c Rev		gtgtttgtattatctaaggcaaccgctggggccttt
	rs13328404	46rs13328404a-g Fwd	+	agtgtgtagcacacaacaaatagcaaatatcagctgggtttttc
		46rs13328404a-g Rev		gaaaaaacccagctgatatttgctatttgttgtgtgctacacact
	rs938648	46rs938648t-g Fwd	+	ggtttttttcccctctccttcagcattaaatgctatagctttca
		46rs938648t-g Rev		tgaaagctatagcatttaatgctgaaggagaggggaaaaaacc
	rs938649	46rs938649g-t Rev		agtatagaaagaatctagggaaattatattaagttaaaagaatgaaagctaagcattta
		46rs938649g-t Fwd	+	taaatgctatagctttcattcttttaacttaatataatttccctagattcttctatact
	rs938650	46rs938650g-a Fwd	+	tcattcttttaacttaatataattgcccaaattctttctatacttaaaaaaatggaagtg
		46rs938650g-a Rev		cacttccatttttttaagtatagaaagaatttagggcaattatattaagttaaaagaatga
	rs201242438	46-6rs201242438del Fwd	+	gaacgttgaaaattacaaagtagaacctcttt-ttttttttttttacatttttttgagaaaagag
		46-6rs201242438del Rev		ctcttttctcaaaaaaatgtaaaaaaaaaaaaa-aaagaggttctactttgtaattttcaacgttc
	rs2392944*	46-7rs2392944c-t Fwd	-	aaattacaaagtagaacctctttttttttttttttacatttttttgagaaaagagattactaggc
		46-7rs2392944c-t Rev		gcctagtaatctctttttctcaaaaaaatgtaaaaaaaaaaaaaaaaaagaggttctactttgtaattt
59	rs6651252	59-1rs6651252t-c Rev		ctgctcacatttcaaagggcccacttttcctccta
		59-1rs6651252t-c Fwd	+	taggaggaaaagtgggccctttgaaatgtgagcag
	rs6651253	59-2rs6651253g-c Fwd	+	gcctgaaacctggaagcgaaaagtccttgactg
		59-2rs6651253g-c Rev		cagtcaaggacttttcgcttccaggtttcaggc
61	rs1561925*	61rs1561925c-t Fwd	-	aggataaaaagccagaacttactttggtggcgcc
		61rs1561925c-t Rev		ggcgccaccaaagtaagttctggctttttatcct
	rs1561924*	61rs1561924g-a Fwd	-	gtaccttgcacaagtaataatcactaccattgatctggc
		61rs1561924g-a Rev		gccagatcaatggtagtgattattacttgtgcaaggtac
	rs2165806*	61rs2165806g-c Rev		actttggacagcatgaacgtcagtgaagtttatctgg
		61rs2165806g-c Fwd	-	ccagataaacttcactgacgttcatgctgtccaaagt

Primer sequences listed are represented in the 5' to 3' direction.

Electrophoretic Mobility Shift Assay

Extraction of nuclear proteins from iOSE4^{cMYC} cells was performed in absence of detergents as they can interfere with the binding activity of the isolated proteins. We prepared Dignam based nuclear extracts (Dignam et al., 1983). Briefly, nuclear extracts were prepared using a hypotonic lysis buffer (10mM HEPES pH 7.9, 1.5mM $MgCl_2$, 10mM KCl, supplemented with DTT and protease inhibitors) followed by cell disruption

with a syringe (gauge No. 27) and isolation of the nuclear fraction with an Extraction Buffer (20mM HEPES pH 7.9, 1.5mM $MgCl_2$, 0.42M NaCl, 0.2mM EDTA, 25% v/v glycerol, supplemented with DTT and protease inhibitors). The DNA probes were designed to contain either the reference or effect allele for each SNP positioned between 20 oligonucleotides on each side (Table 7).

Table 7. DNA probes used in EMSA

SNP	Probe Name	Sequence (5' to 3')
rs13328411	rs13328411m F	GGAATTCCTTCAACTCAGTCAGTCAAGAGGCTCACCAATCC
	rs13328411M F	GGAATTCCTTCAACTCAGTCTGTCAAGAGGCTCACCAATCC
	rs13328411m R	GGATTGGTGAGCCTCTTGACTGACTGAGTTGAAGGAATTCC
	rs13328411M R	GGATTGGTGAGCCTCTTGACAGACTGAGTTGAAGGAATTCC
rs17807628	rs17807628m F	CATCAGTGTGGGCTAATTTGCTACTGGAAATCTAAGGGAAA
	rs17807628M F	CATCAGTGTGGGCTAATTTGTTACTGGAAATCTAAGGGAAA
	rs17807628m R	TTTCCCTTAGATTTCCAGTAGCAAATTAGCCCACACTGATG
	rs17807628M R	TTTCCCTTAGATTTCCAGTAACAAATTAGCCCACACTGATG
rs10088755	rs10088755m F	GAAAAAAATGAATTATGTAGACAATAAAGAACTGTTTATTG
	rs10088755M F	GAAAAAAATGAATTATGTAGGCAATAAAGAACTGTTTATTG
	rs10088755m R	CAATAAACAGTTCTTTATTGTCTACATAATTCATTTTTTTC
	rs10088755M R	CAATAAACAGTTCTTTATTGCCTACATAATTCATTTTTTTC
rs10113762	rs10113762m F	GGCAAAGGCCCCAGCGGTTGCCTTAGATAATACAAACACAA
	rs10113762M F	GGCAAAGGCCCCAGCGGTTGTCTTAGATAATACAAACACAA
	rs10113762m R	TTGTGTTTGTATTATCTAAGGCAACCGCTGGGGCCTTTGCC
	rs10113762M R	TTGTGTTTGTATTATCTAAGACAACCGCTGGGGCCTTTGCC
rs13328404	rs13328404m F	TGTGTAGCACACAACAAATAGCAAATATCAGCTGGGTTTTT
	rs13328404M F	TGTGTAGCACACAACAAATAACAAATATCAGCTGGGTTTTT
	rs13328404m R	AAAAACCCAGCTGATATTTGCTATTTGTTGTGTGCTACACA
	rs13328404M R	AAAAACCCAGCTGATATTTGTTATTTGTTGTGTGCTACACA
rs938648	rs938648m F	GTTTTTTCCCCTCTCCTTCAGCATTAAATGCTATAGCTTTC
	rs938648M F	GTTTTTTCCCCTCTCCTTCATCATTAAATGCTATAGCTTTC
	rs938648m R	GAAAGCTATAGCATTTAATGCTGAAGGAGAGGGGAAAAAAC
	rs938648M R	GAAAGCTATAGCATTTAATGATGAAGGAGAGGGGAAAAAAC
rs938649 & rs938650	rs938649/50m F	TCTTTTAACTTAATATAATTTCCCTAAATTCTTTCTATACTTAAAAA
	rs938649/50M F	TCTTTTAACTTAATATAATTGCCCTAGATTCTTTCTATACTTAAAAA
	rs938649/50m R	TTTTTAAGTATAGAAAGAATTTAGGGAAATTATATTAAGTTAAAAGA
	rs938649/50M R	TTTTTAAGTATAGAAAGAATCTAGGGCAATTATATTAAGTTAAAAGA

Table 7 (Continued). DNA probes used in EMSA

SNP	Probe Name	Sequence (5' to 3')
rs201242438	rs201242438m F	TTACAAAGTAGAACCTCTTTTTTTTTTTTTTTTACATTTT
	rs201242438M F	TTACAAAGTAGAACCTCTTTCTTTTTTTTTTTTTACATTTT
	rs201242438m R	AAAATGTAAAAAAAAAAAAAAAAGAGGTTCTACTTTGTAA
	rs201242438M R	AAAATGTAAAAAAAAAAAAAGAAAGAGGTTCTACTTTGTAA
rs2392944	rs2392944m F	AAAATGTAAAAAAAAAAAAAAAAAGAGGTTCTACTTTGTAA
	rs2392944M F	AAAATGTAAAAAAAAAAAAAGAAAGAGGTTCTACTTTGTAA
	rs2392944m R	TTACAAAGTAGAACCTCTTTTTTTTTTTTTTTACATTTT
	rs2392944M R	TTACAAAGTAGAACCTCTTTCTTTTTTTTTTTTTACATTTT
rs6651252	rs6651252m F	ACATAGGAGGAAAAGTGGGCCCTTTGAAATGTGAGCAGAGC
	rs6651252M F	ACATAGGAGGAAAAGTGGGCTCTTTGAAATGTGAGCAGAGC
	rs6651252m R	GCTCTGCTCACATTTCAAAGGGCCCACTTTTCCTCCTATGT
	rs6651252M R	GCTCTGCTCACATTTCAAAGAGCCCACTTTTCCTCCTATGT
rs6651253	rs6651253m F	TGGAGCCTGAAACCTGGAAGCGAAAAGTCCTTGACTGGAGT
	rs6651253M F	TGGAGCCTGAAACCTGGAAGGGAAAAGTCCTTGACTGGAGT
	rs6651253m R	ACTCCAGTCAAGGACTTTTCGCTTCCAGGTTTCAGGCTCCA
	rs6651253M R	ACTCCAGTCAAGGACTTTTCCCTTCCAGGTTTCAGGCTCCA
rs57593539	rs57593539m F	AGGAGCATTTTGAGCTTGCCAACAGGCTCTCCCATGGATTC
	rs57593539M F	AGGAGCATTTTGAGCTTGCCGACAGGCTCTCCCATGGATTC
	rs57593539m R	GAATCCATGGGAGAGCCTGTTGGCAAGCTCAAAATGCTCCT
	rs57593539M R	GAATCCATGGGAGAGCCTGTCGGCAAGCTCAAAATGCTCCT
rs1561925	rs1561925m F	TCTGCAGGCGCCACCAAAGTAAGTTCTGGCTTTTTATCCTC
	rs1561925M F	TCTGCAGGCGCCACCAAAGTGAGTTCTGGCTTTTTATCCTC
	rs1561925m R	GAGGATAAAAAGCCAGAACTTACTTTGGTGGCGCCTGCAGA
	rs1561925M R	GAGGATAAAAAGCCAGAACTCACTTTGGTGGCGCCTGCAGA
rs1561924	rs1561924m F	AGGAGGTGCCAGATCAATGGTAGTGATTATTACTTGTGCAA
	rs1561924M F	AGGAGGTGCCAGATCAATGGCAGTGATTATTACTTGTGCAA
	rs1561924m R	TTGCACAAGTAATAATCACTACCATTGATCTGGCACCTCCT
	rs1561924M R	TTGCACAAGTAATAATCACTGCCATTGATCTGGCACCTCCT
rs2165806	rs2165806m F	TGACTTTGGACAGCATGAACGTCAGTGAAGTTTATCTGGAA
	rs2165806M F	TGACTTTGGACAGCATGAACCTCAGTGAAGTTTATCTGGAA
	rs2165806m R	TTCCAGATAAACTTCACTGACGTTCATGCTGTCCAAAGTCA
	rs2165806M R	TTCCAGATAAACTTCACTGAGGTTCATGCTGTCCAAAGTCA

Forward and reverse probe pairs were first annealed (10μM each), followed by labeling with ATP [γ-32P] (Perkin Elmer) using T4 polynucleotide kinase (NEB) and then cleanup by utilizing the QIAquick Nucleotide Removal Kit (Qiagen). The clean, labeled probes were subsequently incubated with the previously isolated nuclear protein extracts using LightShift Poly(dI-dC) (Thermo-Fisher Scientific) in a binding buffer solution (10mM Tris, 50mM KCl, 1mM DTT, pH 7.4). The reactions were subjected to electrophoresis on a 6% polyacrylamide gel at 83V, overnight, at 4°C. Gels were dried and films exposed for 5-24 h. EMSA were performed in at least two independent experiments (biological replicates) and each probe was run in duplicate (technical replicates) in each gel. Changes in banding patterns were assessed by visual inspection and only SNPs in probes showing reproducible changes (present in all biological and technical replicates) are further analyzed.

CHAPTER THREE:

IDENTIFYING CANDIDATE TARGET GENES

Note to reader:

A manuscript (Gjyshi et al.) currently under review includes portions of this chapter.

Introduction

Since the 9 candidate causal SNPs are located in three enhancer regions, we hypothesize that they exert their functions through the control of a target gene. Thus, the next step in the analysis is to identify target gene(s) regulated by the three enhancer regions. The candidate causal SNPs are found in a gene desert region in chromosome 8q24 that is devoid of coding DNA. The only well annotated gene in the area is the *MYC* proto-oncogene, located approximately 800 kb centromeric to the region of association that we are investigating. The initial study that identified this locus for association with ovarian cancer reported significant increase of *MYC* expression levels in 24 ovarian cancer cell lines compared to 48 normal primary human surface epithelium cell cultures ($p = 0.0011$); however, no evidence of association was observed between SNP allele of the tag SNP and gene expression (Goode et al., 2010).

In this chapter, we perform *in silico* annotation of putative target genes followed with experimental approaches to identify genes that are active in our panel of ovarian

cancer and normal cells and are the targets of the enhancer regions we identified in the previous chapter.

Results

Identification of Candidate Target Genes, *MYC* and *PVT1*

We confined our search for promoters of candidate target genes to a 1 Mb span on each side of the 31 kb region of association specified in the previous chapter, totaling a span of 2.03 Mb (Figure 13A). This distance was determined based on the observation that most interactions between enhancers and target genes are found within 1 Mb of each other (Jin et al., 2013). To identify candidate target genes we used the University of California Santa Cruz (UCSC) Genes prediction track in the Genome Browser. The only protein coding genes in this region are the *MYC* proto-oncogene, located approximately 800 kb away, and Transmembrane Protein 75 (*TMEM75*) coding for a protein of unknown function, located over 600 kb away, both centromeric to the ovarian cancer susceptibility region. The 2.03 Mb span also includes eight long non-coding RNAs and five micro RNAs (Figure 13A).

We inspected the region and identified promoters located near the transcription start site (TSS), marked by H3K4me3 histone modifications as indicated in the ENCODE Regulation track (Consortium, 2012). Next, we examined which of these genes/transcripts were expressed in ovarian cells using H3K27ac histone modifications as markers of active promoters, and RNA-seq data in ovarian cells. This analysis suggested that *MYC* and *PVT1* were actively expressed in ovarian cells (Figure 13B).

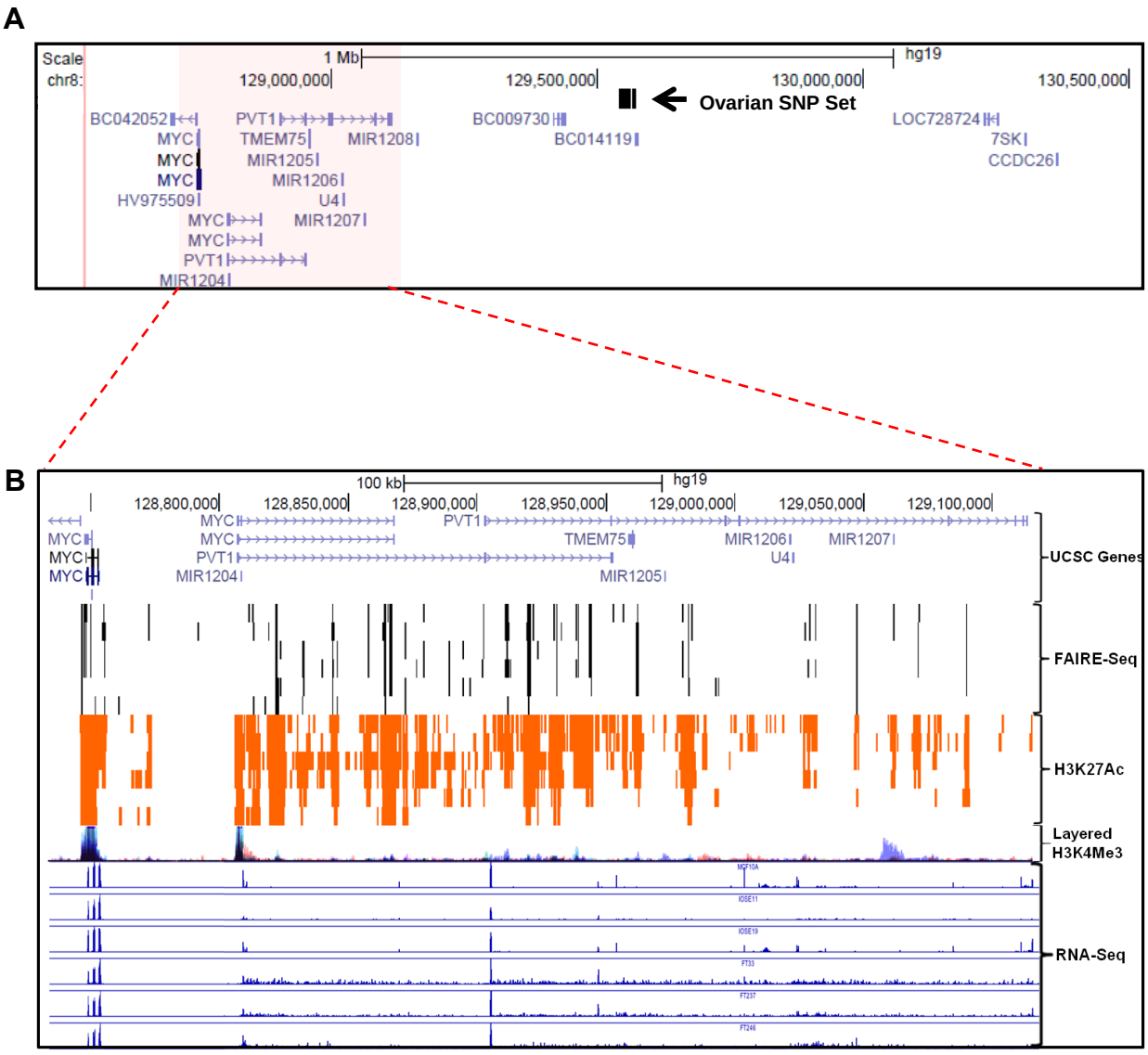


Figure 13. Target genes in the 8q24 locus. *A.* Region surrounding the fine mapped region showing all the UCSC genes found within this region. Location of ovarian SNP set (odds 100:1 or better) (black box) is indicated with an arrow. ***B.*** Region containing the *MYC* and *PVT1* genes showing location of chromatin features (FAIRE-seq and H3K27AC) for ovarian cells, and RNA-seq data (for iOSE11, iOSE19, iFTSEC33, iFTSEC237, and iFTSEC246). Note: RNA-seq data were generated by Kate Lawrenson.

Both have been implicated in ovarian cancer (Guan et al., 2007; Karst et al., 2011; Lawrenson et al., 2011) and are frequently amplified in 135 (44%) of 311 patients/cases

(TCGA Provisional, Ovarian Cystadenocarcinoma). Consequently, we tested whether *MYC* or *PVT1* interacted with enhancer regions in the locus.

Enhancer Regions Display Physical Interactions with Promoters of *MYC* and *PVT1*

We utilized Chromosome Conformation Capture (3C) to investigate physical interactions between the promoter of candidate target genes, *MYC* and *PVT1,* and four enhancer regions containing the set of 15 candidate SNPs (Regions E1-E4). In our analyses of iFTSEC283 and iOSE11 cells, the chromatin fragment containing the *MYC* promoter shows increased interaction frequencies with regions E2, E3, and E4, compared to other surrounding fragments in the 8q24 ovarian cancer susceptibility locus which is located approximately 800 kb away (Figure 14). These regions contain eight SNPs with allele-specific activity in both functional assays (enhancer scanning assay and EMSA) (Table 3). The *PVT1* promoter region also displays peaks of interaction with regions E3 and E4, but at a lower frequency compared to the *MYC* interactions (Figure 15). These data suggest that promoters of both of these genes physically interact with enhancer regions identified in the 8q24.21 ovarian cancer susceptibility locus and are most likely their targets of transcriptional regulation.

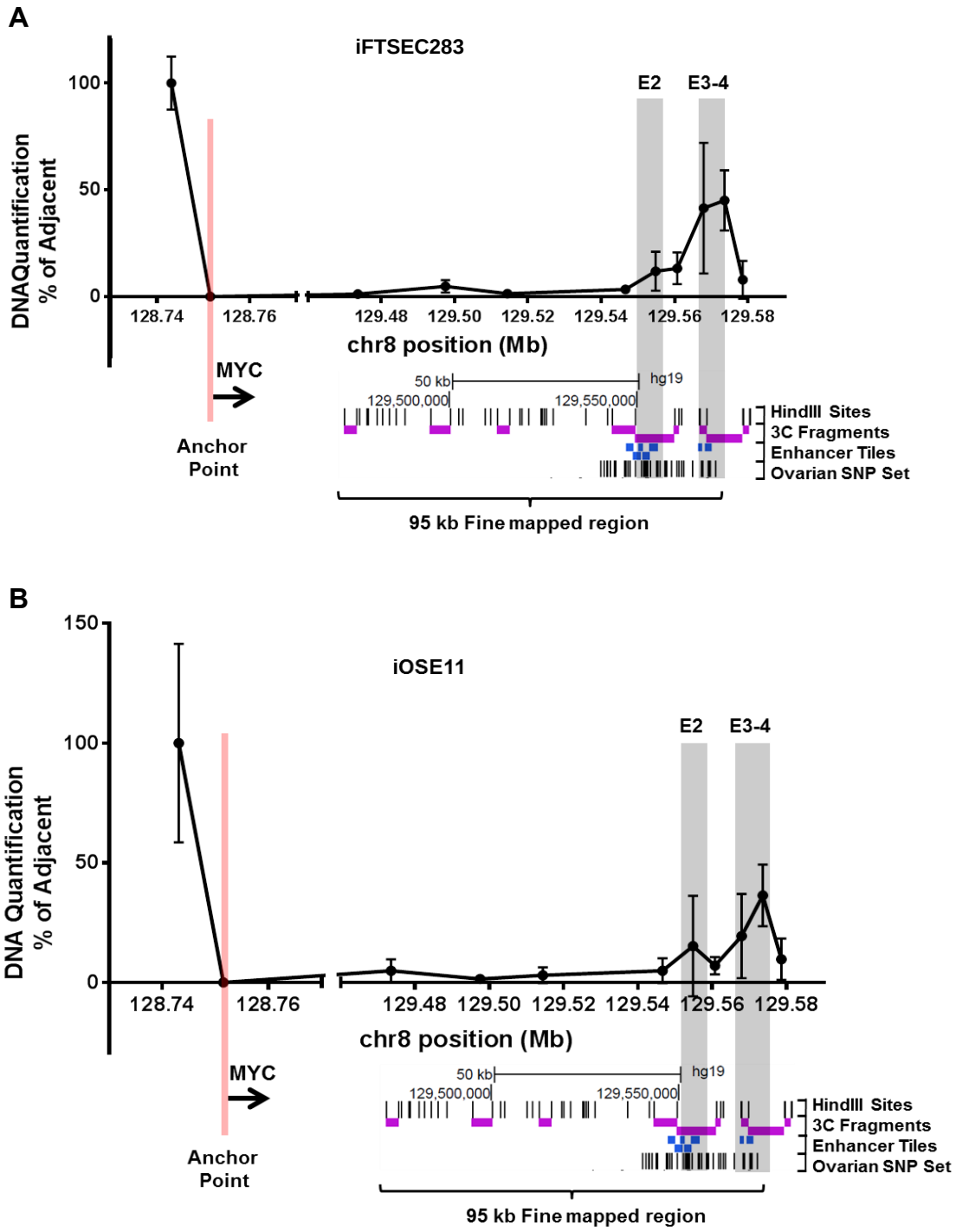


Figure 14. Chromosome conformation capture (3C) analysis of enhancer regions with *MYC* at the 8q24.21 locus. Graphs display 3C interactions in iFTSEC283 (***A***) and in iOSE11 (***B***) cells. The bait is anchored at the *MYC* promoter region (red) which is tested for physical interaction with 3C fragments (purple bars) containing enhancer regions (highlighted in gray). 3C interaction with a fragment adjacent to *MYC* is used as a reference.

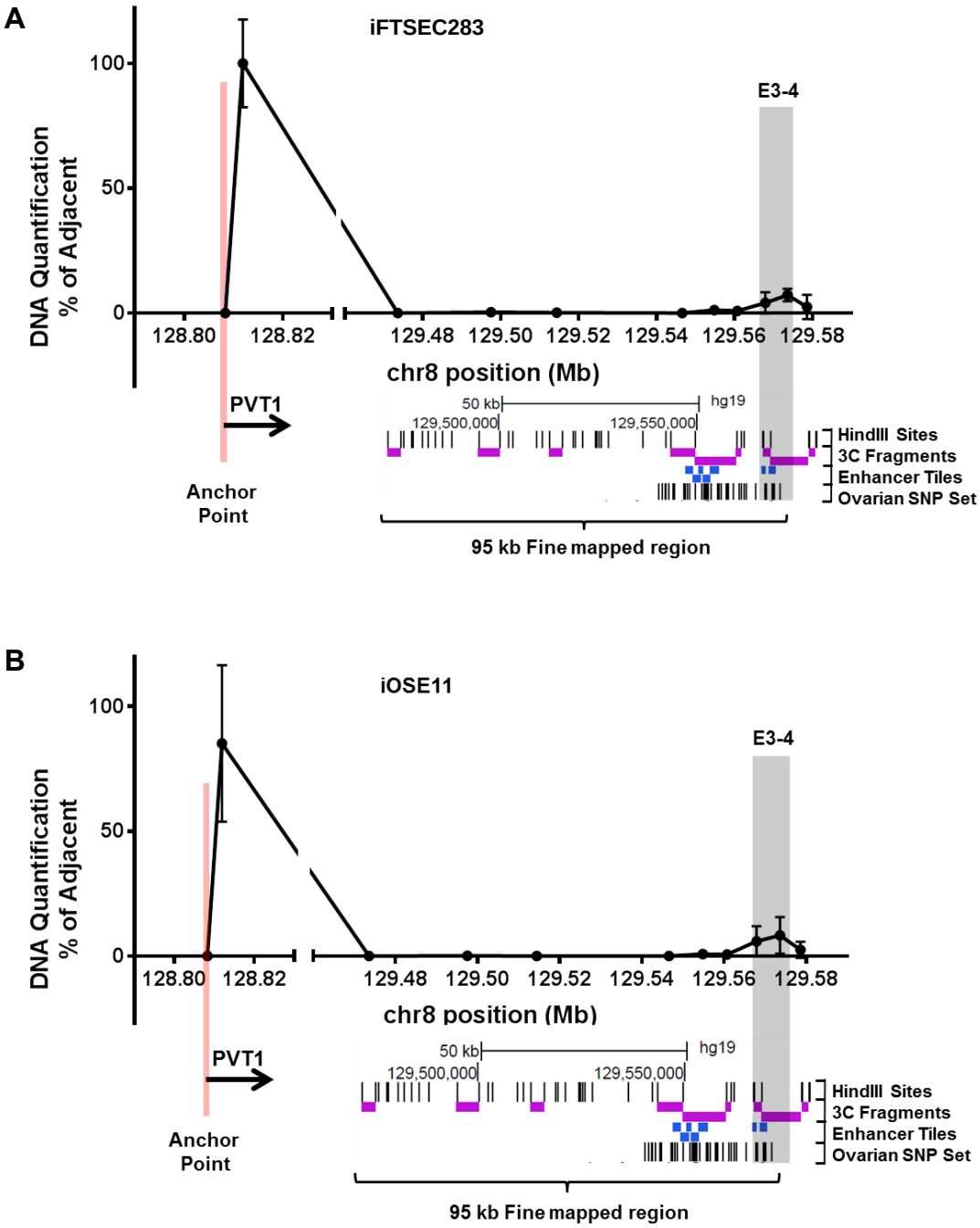


Figure 15. Chromosome conformation capture (3C) analysis of enhancer regions with *PVT1* at the 8q24.21 locus. Graphs display 3C interactions in iFTSEC283 (***A***) and in iOSE11 (***B***) cells. The bait is anchored at the *PVT1* promoter region (red) which is tested for physical interaction with 3C fragments (purple bars) containing enhancer regions (highlighted in gray). 3C interaction with a fragment adjacent to *MYC* or *PVT1* is used as a reference.

Summary

In this chapter, we identified *MYC* and *PVT1* as target genes that are being regulated by the three enhancer regions identified in the previous chapter. These genes displayed transcription activity at their promoters as indicated byH3K4me3 marks in ENCODE and FAIRE-seq landscape in our panel of ovarian cells (Table 8). Additionally, the *MYC* and *PVT1* transcripts were evaluated in our panel of ovarian cells by RNA-seq analysis. Finally, direct physical interactions between promoter regions of each of the two genes and enhancer regions in the 8q24.21 ovarian cancer susceptibility locus confirm *MYC* and *PVT1* as their targets of transcriptional regulation. An illustration of the proposed interaction is presented in Figure 16.

Table 8. Summary of candidate target genes

Genes in 2.03 MB region	ENCODE panel	Ovarian panel	
	H3K4me3	RNA-Seq	FAIRE-Seq
CASC8	-	-	-
CASC11	-	-	-
MYC	+	+	+
PVT1	+	+	+
TMEM75	-	-	-
LINC00824	-	-	-
LINC00977	-	-	-

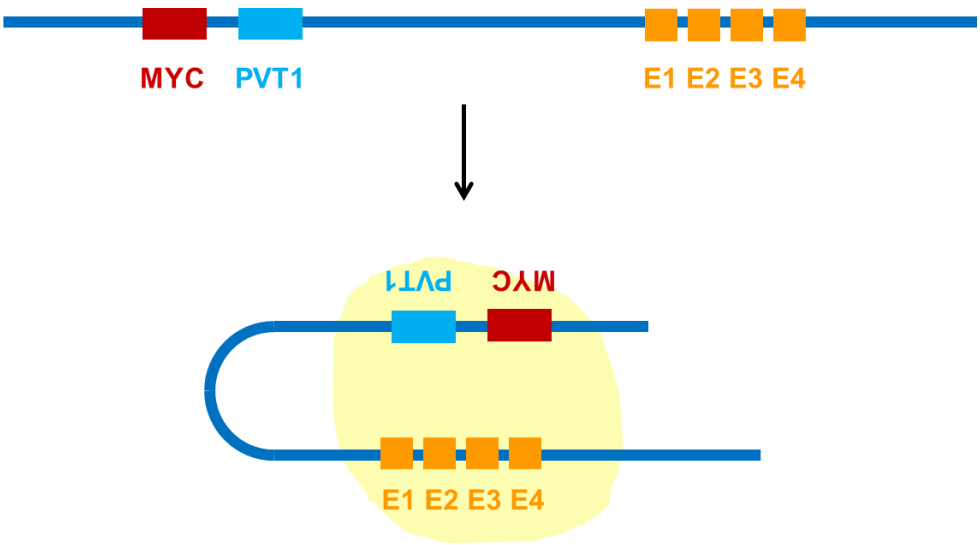


Figure 16. Illustration of looping structure between promoters of target genes and enhancer regions. Linear chromatin is represented in the top configuration. Looping structure that brings the enhancer regions and their target genes in close proximity to each other is represented in the bottom configuration.

Materials and Methods

In Silico Annotation of Candidate Target Genes

The search for candidate target genes extended 1Mb on each side of the 31 Kb region identified by fine mapping and statistical analysis (totaling approximately 2.03 Mb). We used the USCS Genes prediction track within the Genome Browser (https://genome.ucsc.edu/) to obtain a comprehensive list of well-annotated human protein-coding and non-coding genes which comprise our set of candidate target genes.

Chromosome Conformation Capture (3C) Assay

iFTSEC283 and iOSE11 cells were grown to 80% confluence and cross-linked by treatment with 1% formaldehyde (Sigma-Aldrich). Cells were subjected to treatment with homemade lysis buffer (10mMTris HCl pH8, 10mM NaCl, 0.2% NP40 (Sigma-Aldrich) and 1X protease inhibitor cocktail (Roche). The nuclear extracts were digested with HindIII (NEB) followed by ligation with T4 DNA Ligase (NEB), which is carried out in dilute conditions in order to promote intramolecular ligation of the cross-linked fragments. Ligated and unligated (control) samples were examined on an agarose gel before proceeding to the next step. The samples were consequently de-cross-linked by treatment with Proteinase K (Qiagen). The DNA was isolated through phenol-chloroform extraction and ethanol precipitation prior to quantification by qPCR. The ligated 3C library was tested using control primers described previously (Lieberman-Aiden et al., 2009). Test primers were designed utilizing the Primer 3 online tool (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). A list of primers used to quantify the interactions is included in Table 9. Quantitative PCR of 3C library fragments was performed as described previously (Tan-Wong et al., 2008). The qPCR reactions were performed using Hotstart Taq polymerase (Qiagen) and SYTO9 (Life Technologies) on an ABI 7900HT Real-Time PCR System. Artificial 3C libraries, obtained from BAC DNA (RPC11-440N18, RP11-55J15, RPCI-11-1142F3 from Empire Genomics; 2034C18, 96012 from Invitrogen) covering the regions under investigation, were used to test primer efficiency and to generate standard curves for PCR quantification of 3C fragments.

Table 9. 3C Primers used to quantify physical interactions

Primer Name	Sequence (5' to 3')	Distance from Hindiii site (bp)
3F	CTAATTATGAGCTGAACGCTTTACG	92
12F	TGTCATTGTTATGGGAGAATAAAAG	69
18F	GCTGGGTCCAGTCTGACAAA	81
29R	GTCACTTGATTCTCCTTTTCCTTTT	99
30R	CATGAAAACCCTATTAAGCAGAAAA	108
31R	TAATTCATGTGATTGCATAGTCCAG	50
35R	TAGCTCAGGAAGATAAACTGAAATG	70
36R	CCCACCCTCCACTCTAAACC	75
37R	AGGGTGACAGAGCTTTAGTGAAGTA	45
MYC Bait	GAGAACCGGTAATGGCAAAC	83
MYC Adj	TGCATGGTGTTTCATAGTGAGTT	80
PVT1 Bait	ATCTTGGAGGTGAGGACGTG	45
PVT1 Adj	GTCAGGGAGCTGAGGAGTGT	38

CHAPTER FOUR:

TRANSCRIPTION FACTOR BINDING ANALYSIS

Note to reader:

A manuscript (Gjyshi et al.) currently under review includes portions of this chapter.

Introduction

The next step in the analysis of regulatory regions is focused on detection of potential transcription factor binding sites that are involved in regulation of transcription of the target genes. We started with an *in silico* analysis to identify transcription factors binding sites (TFBS) that are disrupted by SNPs and followed up with experiments to verify transcription factor binding and specificity.

Results

Prediction of TFBS Disrupted by SNPs

For the 8 SNPs that displayed allele-specific binding in the reporter and EMSA assays, we used the Match weight matrix-based program (http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi) to predict transcription factor

binding sites (TFBS) for the DNA sequences that were used in the EMSA probes (Table7). A list of predicted transcription factors was generated for each SNP allele. Interestingly, exclusive binding of Activating Transcription Factor 1 (ATF1) to the minor allele [C] in the forward orientation) of rs2165806 (region E4; serous p = 1.32 x 10^{-19}) was predicted. This predicted binding is completely eliminated upon switching from the minor [C] to the major [G] allele (Table 10), consistent with EMSA results (Figure 10).

Isolation of ATF1 Transcription Factor Binding rs2165806 by FPLC and EMSA

To validate this finding experimentally we utilized Fast Protein Liquid Chromatography (FPLC) combined with EMSA to isolate specific SNP binding proteins for rs2165806 (Figure 17). Nuclear extracts of 293FT cells were subjected to a first round of FPLC through a size exclusion column followed by a screen for SNP specific binding proteins by EMSA. We identified three fractions (fractions 20-22) that displayed a signal in the EMSA assay. Those fractions were subjected to a second round of FPLC through an ion-exchange column, followed by EMSA. The two fractions (D4 and D5) that displayed activity in the second EMSA were separated by electrophoresis and all six regions (Figure 17; Step 3) were submitted for mass spectrometry analysis. When proteins matching the gene ontology term “DNA binding” (GO:0003677) were selected, ATF1 emerged as a significant hit that also had a predicted allele-specific binding difference in the context of SNP rs2165806 (Table 10). From these experiments we confirmed ATF1 (Figure 18) to be one of the transcription factors that was isolated from our DNA probe containing the effect allele [C] for rs2165806. These results were

confirmed in iOSE4^{cMYC} cells (Figure 19B). This provides evidence that this functional SNP, harbored within the enhancer region E4, recruits the ATF1 transcription factor in an allele-specific manner.

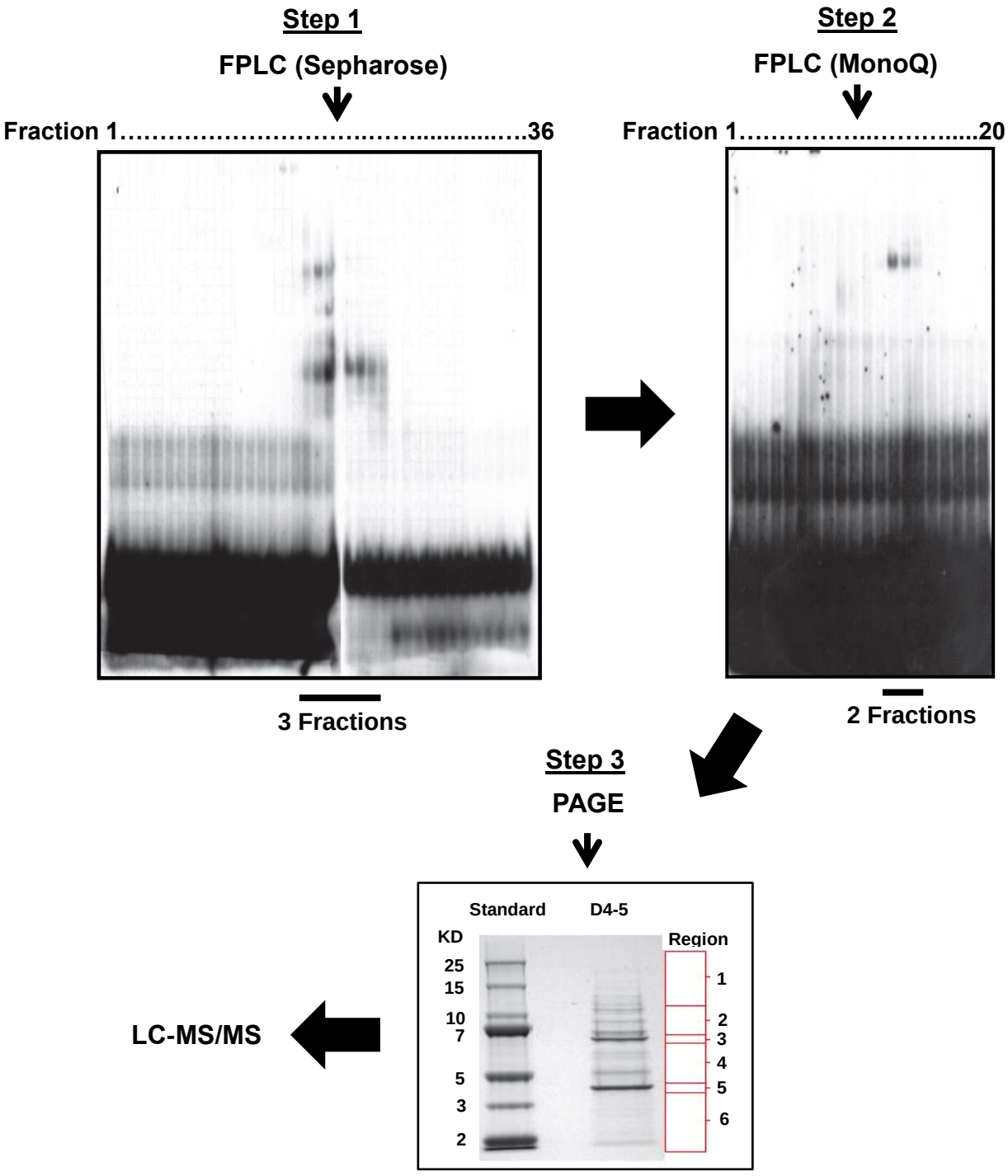
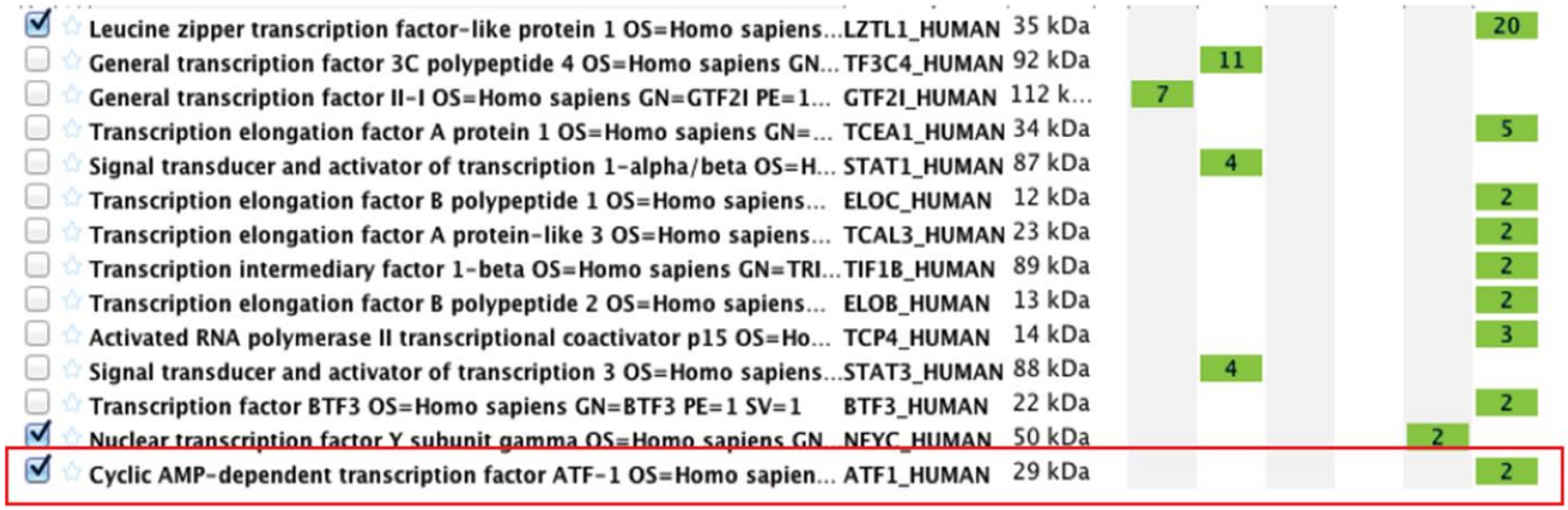


Figure 17. Isolation of ATF1 transcription factor binding to rs2165806 by FPLC and EMSA. Analysis includes two rounds of FPLC (see supplemental figures) followed by EMSA screens for SNP specific binding proteins. For the last step, two fractions with positive EMSA signal were subjected to PAGE and separated into six regions that were processed for proteomic analysis. Note: FPLC analysis was performed by Nicholas Woods and the EMSA assay included in this picture was performed by Gustavo Mendoza.

Table 10. TFBS predicted by the MATCH program

SNP ID	Matrix identifier	Position	Strand	Core match	Matrix match	Sequence (+) strand	Factor name	Present in minor?	Present in major?
	V$OCT1_Q6	1	(-)	0.824	0.709	tgacTTTGGacagca	Oct-1	Yes	
	V$PAX6_01	3	(-)	0.842	0.577	actttggacagCATGAacctc	Pax-6	Yes	
	V$PAX4_01	3	(-)	1	0.711	actttggacagCATGAacctc	Pax-4	Yes	
	V$PAX4_01	9	(-)	0.881	0.589	gacagcatgaaCCTCAgtgaa	Pax-4	Yes	
	V$PAX6_01	21	(+)	0.757	0.609	ctcagTGAAGtttatctggaa	Pax-6	Yes	
	V$PAX4_01	21	(+)	0.888	0.66	ctcagTGAAGtttatctggaa	Pax-4	Yes	
rs2165806_majorG	V$FREAC7_01	25	(-)	1	0.861	gtgaagTTTATctgga	Freac-7	Yes	
	V$GATA_C	27	(-)	1	0.957	gaagtTTATCt	GATA-X	Yes	
	V$GATA1_03	28	(-)	1	0.936	aagtTTATCtggaa	GATA-1	Yes	
	V$LMO2COM_02	30	(-)	1	0.955	gttTATCTg	Lmo2_complex	Yes	
	V$GATA1_05	30	(-)	1	0.959	gtTTATCtgg	GATA-1	Yes	
	V$GATA1_06	30	(-)	1	0.952	gtTTATCtgg	GATA-1	Yes	
	V$GATA2_02	30	(-)	1	0.959	gtTTATCtgg	GATA-2	Yes	
	V$OCT1_Q6	1	(-)	0.824	0.709	tgacTTTGGacagca	Oct-1		Yes
	V$PAX6_01	3	(-)	0.842	0.577	actttggacagCATGAacgtc	Pax-6		Yes
	V$PAX4_01	3	(-)	1	0.712	actttggacagCATGAacgtc	Pax-4		Yes
	V$PAX4_01	9	(-)	0.879	0.588	gacagcatgaaCGTCAgtgaa	Pax-4		Yes
	V$ATF_01	14	(-)	1	0.974	catgaaCGTCAgtg	ATF		No
	V$CREB_Q2	15	(-)	1	0.903	atgaaCGTCAgt	CREB		No
	V$CREBP1_Q2	15	(-)	1	0.873	atgaaCGTCAgt	CRE-BP1		No
	V$AP1_Q2	16	(-)	0.967	0.92	tgaaCGTCAgt	AP-1		No
	V$AP1_Q4	16	(-)	0.935	0.856	tgaaCGTCAgt	AP-1		No
	V$CREB_01	17	(-)	1	0.937	gaaCGTCA	CREB		No
rs2165806_minorC	V$CREB_02	17	(-)	1	0.924	gaaCGTCAgtga	CREB		No
	V$ER_Q6	17	(-)	0.846	0.792	gaaCGTCAgtgaagtttat	ER		No
	V$PAX6_01	21	(+)	0.757	0.609	gtcagTGAAGtttatctggaa	Pax-6		Yes
	V$PAX4_01	21	(+)	0.888	0.66	gtcagTGAAGtttatctggaa	Pax-4		Yes
	V$FREAC7_01	25	(-)	1	0.861	gtgaagTTTATctgga	Freac-7		Yes
	V$GATA_C	27	(-)	1	0.957	gaagtTTATCt	GATA-X		Yes
	V$GATA1_03	28	(-)	1	0.936	aagtTTATCtggaa	GATA-1		Yes
	V$LMO2COM_02	30	(-)	1	0.955	gttTATCTg	Lmo2_complex		Yes
	V$GATA1_05	30	(-)	1	0.959	gtTTATCtgg	GATA-1		Yes
	V$GATA1_06	30	(-)	1	0.952	gtTTATCtgg	GATA-1		Yes
	V$GATA2_02	30	(-)	1	0.959	gtTTATCtgg	GATA-2		Yes



☑	Leucine zipper transcription factor-like protein 1 OS=Homo sapiens...	LZTL1_HUMAN	35 kDa						20
☐	General transcription factor 3C polypeptide 4 OS=Homo sapiens GN...	TF3C4_HUMAN	92 kDa		11				
☐	General transcription factor II-I OS=Homo sapiens GN=GTF2I PE=1...	GTF2I_HUMAN	112 k...	7					
☐	Transcription elongation factor A protein 1 OS=Homo sapiens GN=...	TCEA1_HUMAN	34 kDa						5
☐	Signal transducer and activator of transcription 1-alpha/beta OS=H...	STAT1_HUMAN	87 kDa		4				
☐	Transcription elongation factor B polypeptide 1 OS=Homo sapiens...	ELOC_HUMAN	12 kDa						2
☐	Transcription elongation factor A protein-like 3 OS=Homo sapiens...	TCAL3_HUMAN	23 kDa						2
☐	Transcription intermediary factor 1-beta OS=Homo sapiens GN=TRI...	TIF1B_HUMAN	89 kDa						2
☐	Transcription elongation factor B polypeptide 2 OS=Homo sapiens...	ELOB_HUMAN	13 kDa						2
☐	Activated RNA polymerase II transcriptional coactivator p15 OS=Ho...	TCP4_HUMAN	14 kDa						3
☐	Signal transducer and activator of transcription 3 OS=Homo sapiens...	STAT3_HUMAN	88 kDa		4				
☐	Transcription factor BTF3 OS=Homo sapiens GN=BTF3 PE=1 SV=1	BTF3_HUMAN	22 kDa						2
☑	Nuclear transcription factor Y subunit gamma OS=Homo sapiens GN...	NFYC_HUMAN	50 kDa					2	
☑	Cyclic AMP-dependent transcription factor ATF-1 OS=Homo sapien...	ATF1_HUMAN	29 kDa						2

ATF1_HUMAN (100%), 29,232.4 Da
Cyclic AMP-dependent transcription factor ATF-1 OS=Homo sapiens GN=ATF1 PE=1 SV=2
2 exclusive unique peptides, 2 exclusive unique spectra, 2 total spectra, 35/271 amino acids (13% coverage)

M E D S H K S T T S E T A P Q P G S A V Q G A H I S H I A Q Q V S S L S E S E E
S Q D S S D S I G S S Q K A H G I L A R R P S Y R K I L K D L S S E D T R G R K
G D G E N S G V S A A V T S M S V P T P I Y Q T S S G Q Y I A I A P N G A L Q L
A S P G T D G V Q G L Q T L T M T N S G S T Q Q G T T I L Q Y A Q T S D G Q Q I
L V P S N Q V V V Q T A S G D M Q T Y Q I R **T T P S A T S L** **P Q T V V M T S P V**
T L T S Q T T K T D D P Q L K R E I R L M K N R E A A R E C R R K K K E Y V K C
L E N R **V A V L E N** **Q N K** T L I E E L K T L K D L Y S N K S V

Figure 18. Proteomics results isolating ATF1 by FPLC and EMSA. *A*. List of DNA binding proteins screed against CrapOme (found < 2.5% in CrapOme). ATF1 is marked with a red box. ***B***. 35/271 amino acids were isolated by LC-MS/MS (highlighted in yellow). Amino acid affected by rs2165806 is highlighted in green.

ATF1 Transcription Factor Binds Exclusively to Effect Allele of rs2165806

To confirm ATF1 binding to the rs2165806 site, we performed EMSA super shift and competition assays with nuclear extracts of 293FT cells. In the competition assays, the EMSA signal that is detected in the radioactively labeled DNA probe containing the minor (effect) allele is completely abolished upon competition with unlabeled probe (Figure 19 A). No signal was detected in the DNA probe containing the major SNP allele [G]. Furthermore, we incubated nuclear extracts of 293FT and iOSE4^{cMYC} cells transfected with pNTAP-ATF1, expressing a fusion of Calmodulin-binding protein (CBP) and ATF1, mixed with anti-CBP antibody and labeled reference or effect allele DNA probes. We detected a super shift signal in the sample containing the effect allele [C] probe for rs2165806, but not on the one containing the reference allele [G] probe (Figure 19 B). This experiment confirms binding of ATF1 to a regulatory element in region E4 that contains the effect allele of rs2165806.

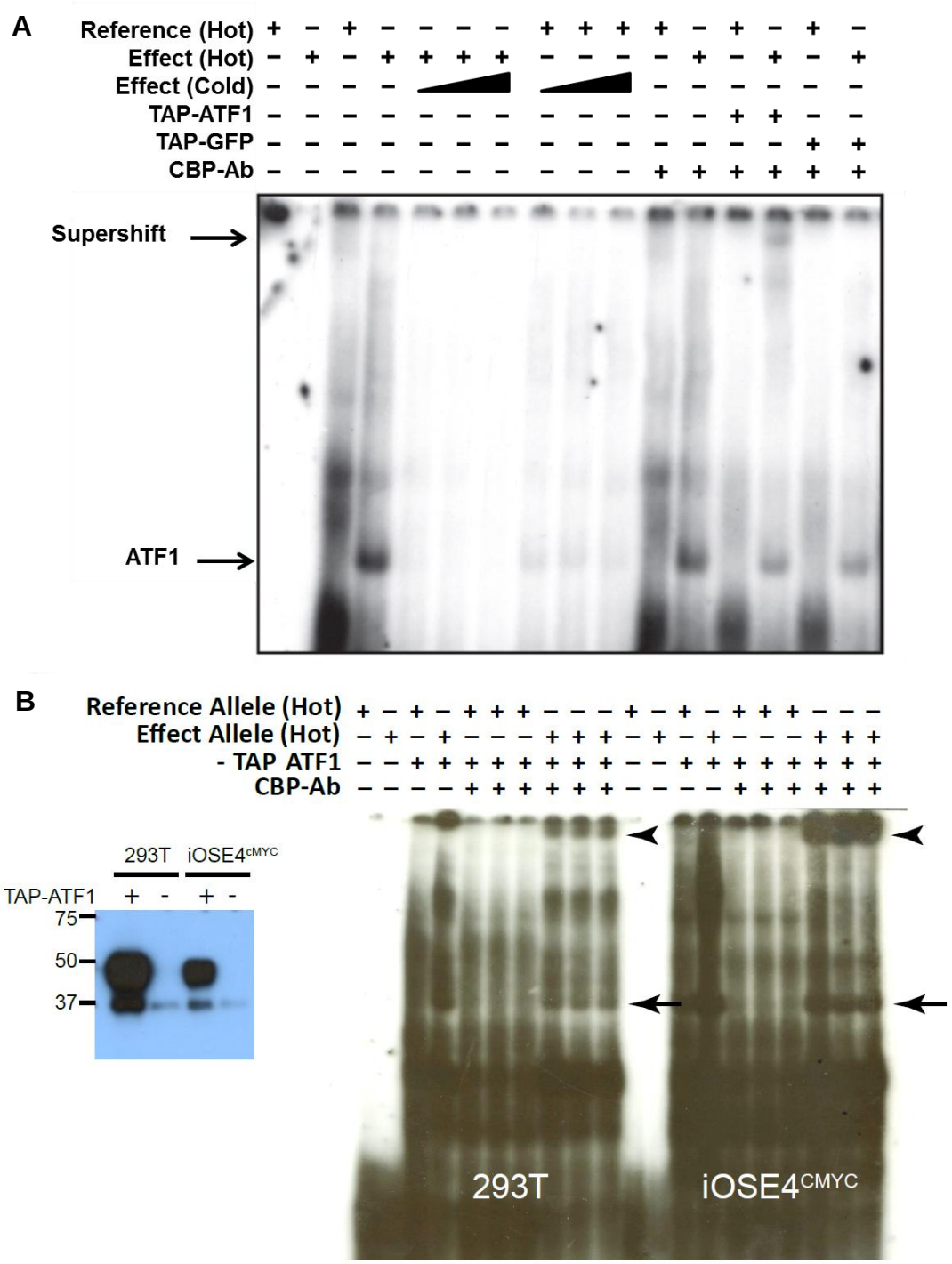


Figure 19. Confirmation of ATF1 binding rs2165806 region. ***A.*** ATF1 binding (lane 4, arrow) is displaced by competition with cold probe containing the Effect allele (lanes 5-7). Super shift is obtained with the addition of an antibody that recognizes a tag (CBP) on the ectopic copy of ATF1 (lane 14, upper arrow). ***B.*** Replicate EMSA also including nuclear extracts from iOSE4^{cMYC} cells. ATF1 binding (arrow) and super shift (arrowhead) are shown.

To assess the effect of ATF1 on the regulatory region, we co-transfected iOSE4^{cMYC} cells with a luciferase reporter vector with the rs2165806 region cloned as an enhancer and an ATF1 expression vector. Ectopic expression of ATF1 increased expression of a reporter gene driven by a region containing the minor [C] allele of rs2165806 but not when the reporter is driven by a region with the major [G] allele (Figure 20). Taken together these experiments indicate that ATF1 binds a region containing the minor allele of rs2165806 and is able to regulate gene expression.

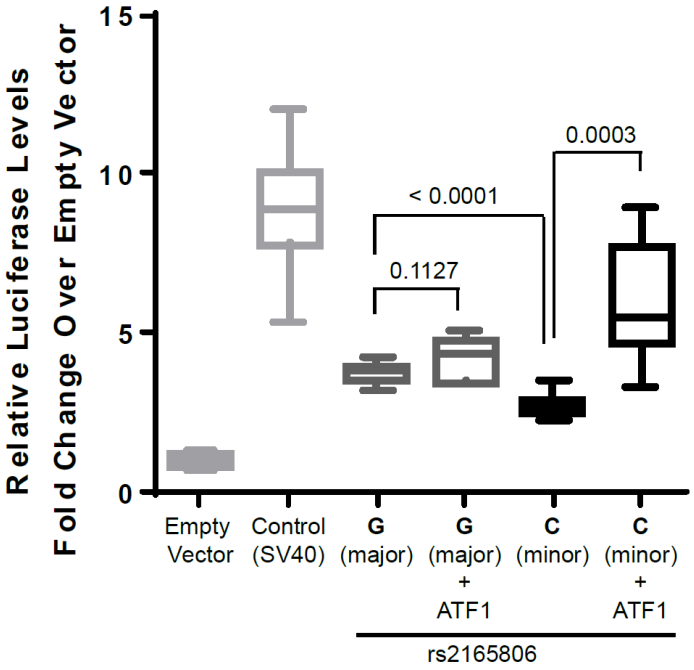


Figure 20. ATF1 affects transcriptional activity of downstream gene reporter. Ectopic expression of ATF1 increases expression of a reporter gene driven by a region containing the effect (minor) [C] allele of rs2165806 but not when the reporter is driven by a region with the reference (major) [G] allele.

Homozygous Deletion of ATF1 Binding Region Impacts the Expression of *MYC*, *PVT1*, and *TMEM75*

To assess the functional relevance of the ATF1 binding region we generated homozygous deletions by CRISPR-Cas9 in iOSE4^{CMYC}. Clones ΔCC1 and ΔCC2 presented homozygous deletion of the rs2165806 region (Figure 21), while CC3 retained the intact region and was used as a reference. Compared to CC3, a significant decrease in expression levels were observed for all three genes in the region (*MYC*, *PVT1*, and *TMEM75*). However, only *MYC* expression was reduced consistently in both clones and in both replicates.

To rule out that changes in gene expression were due to large scale chromosome structural changes due to the CRISPR-mediated deletion we conducted 3C anchoring at *MYC* or *PVT1.* Physical interactions between *MYC* or *PVT1* promoters and regions E3 or E4 in CRISPR clones were similar to the interaction pattern obtained in unmodified cells (Figure 14 and 15). In summary, deletion of the region containing SNP rs2165806 led to consistent changes in *MYC* expression, and in the other two genes in the region without large effects on chromatin structure.

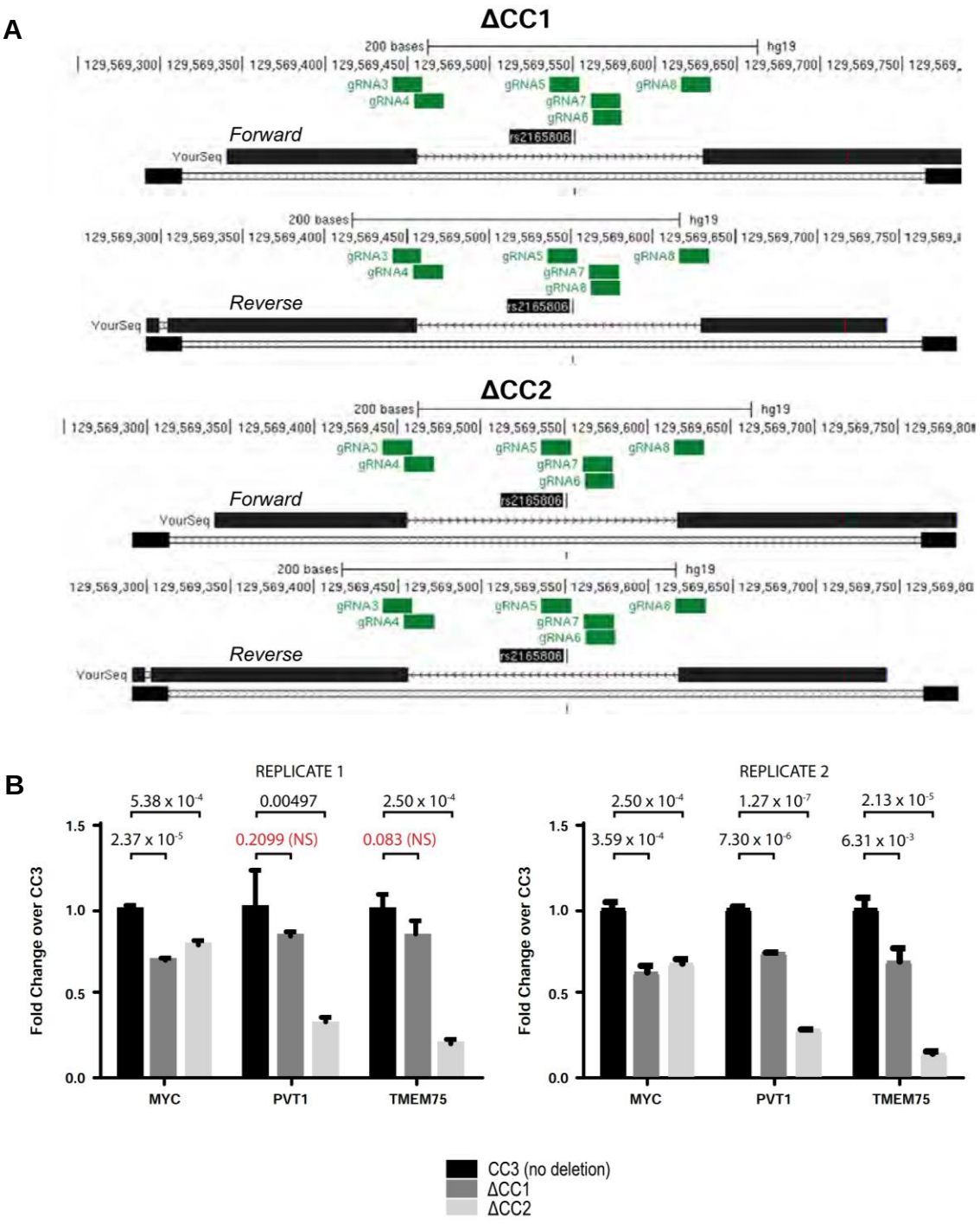


Figure 21. CRISPR-edited clones removing region containing SNP rs2165806.
***A*.** Diagram showing the region removed in two independent iOSE4CMYC CRISPR clones. Positions of gRNAs used are shown in green boxes. ***B*.** RT-qPCR results depicting the expression of MYC, PVT1, and TMEM75 relative to the expression in a CRISPR clone with the region intact (CC3). P values are shown (t-test) and non-significant values are shown in red font.

MYC Transcription Factor Binding Sites by ChIP-seq from ENCODE

Furthermore, a search of the ENCODE ChIP-Seq signals for ATF1 also revealed that this transcription factor as well as BRCA1 have been found to bind to the MYC promoter (Figure 22). This information opens the possibility of BRCA1 involvement and also further supports our suggested mechanism of the identified regulatory elements regulating transcription of distant target genes by recruitment of ATF1.

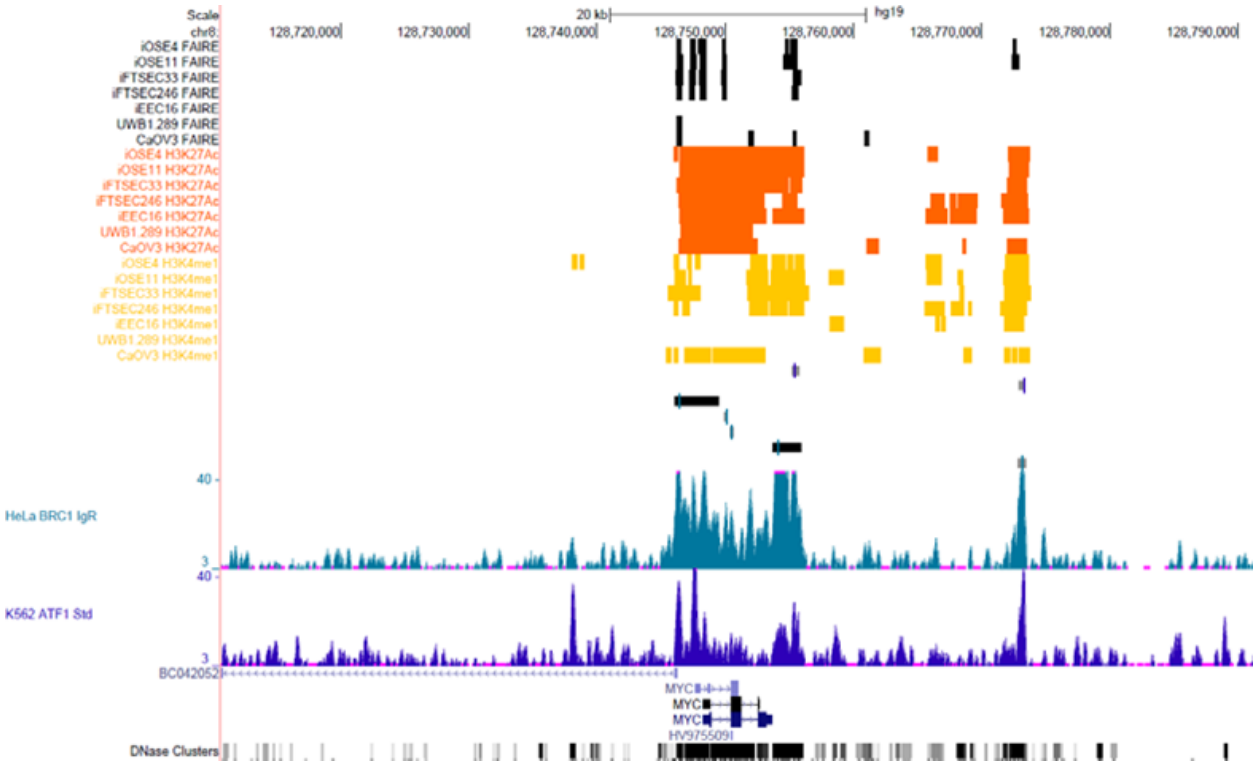


Figure 22. MYC region illustrating ChIP-Seq signals for BRCA1 (in HeLa cells) and for ATF1 (in K562 cells)*;* SYDH TFBS (Transcription Factor Binding Sites by ChIP-seq from ENCODE/Stanford/Yale/USC/Harvard) available through the human genome browser as part of ENCODE.

Summary

In this chapter we identified and validated specific and exclusive binding of the ATF1 transcription factor to the enhancer region (E4) containing the effect allele [C] of the rs2165806 SNP. The predicted ATF1 binding to this region was confirmed by mass spectrometry and further validated by EMSA competition and super shift assays. ATF1 binding to the E4 enhancer region was also shown to regulate gene expression as demonstrated by luciferase assays. Furthermore, deletion of the ATF1 binding region by CRISPR-Cas9 mediated genome editing resulted in decreased levels of gene expression of *MYC*, *PVT1* and *TMEM75*. Additional evidence of ATF1 and BRCA1 binding at the promoter region of *MYC* in two cell lines suggests possible involvement of BRCA1 in transcriptional regulation of *MYC*.

Materials and Methods

Fast Protein Liquid Chromatography (FPLC) Coupled with EMSA

Note to reader: FPLC was performed by Nicholas Woods.

We started with Dignam-based nuclear extracts of 293FT cells (Dignam et al., 1983). Nuclear extracts (10mg) were loaded onto FPLC using size exclusion column Sepharose 6 10/300 GL, and fractions were screened for SNP-specific binding proteins by EMSA. Fractions displaying EMSA signals were combined and subjected to buffer exchange and concentration. Samples were then subjected to a second round of FPLC

using ion-exchange column MonoQ 5/10 GL. These fractions were again screened by EMSA. The two fractions displaying EMSA signals were combined and concentrated and then subjected to polyacrylamide gel electrophoresis (PAGE). The gel was separated into six regions, processed for in-gel trypsin digestion, and submitted to the Moffitt Proteomics core for mass spectrometry analysis on the Orbitrap instrument.

EMSA Competition and Super Shift Assays

EMSA samples were prepared as described in Chapter 2. For the competition assays a cold (unlabeled) probe was used in addition to the labeled probe containing the effect allele [C] for rs2165806. The super shift assay was performed with cells overexpressing ATF1 fused to CBP and CBP antibody was added to the reaction.

CRISPR-Cas9 Mediated Genome Editing in iOSE4^{cMYC} Cells

We used the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 system to remove the region containing the ATF1 binding site, which includes rs2165806, in iOSE4^{CMYC} cells, as previously described (Storici, 2014). Briefly, six gRNA sequences (Table 11) were designed and cloned into the gRNA cloning vector (Addgene, plasmid ID 41824) following the Mali lab protocol option B (http://arep.med.harvard.edu/human_crispr/). Incorporation of gRNA sequences was confirmed by Sanger sequencing using the M13 Reverse universal sequencing primer. iOSE4^{CMYC} cells were co-transfected with six gRNA expression constructs or the empty gRNA plasmid and hCas9 (Addgene, plasmid ID 41815) and pBABE-puro (Addgene, plasmid ID 1764) using FugeneHD transfection reagent (Promega, E2311) at a 3:1 ratio

of FugeneHD volume (µL) to DNA (µg). Forty eight hours post transfection, cells were harvested, plated at different densities, and treated with 1 µg/mL puromycin. Individual clones were isolated and expanded in cell culture. Excision of the AFT1 binding site was confirmed by sequencing (Table 11). Two clones (ΔCC1 and ΔCC2) were confirmed to have the desired CRISPR-mediated deletions at Chr8:129,569,456-129,569,629 and Chr8:129,569,456-129,569,618. One CRISPR clone presented with a small deletion (Chr8:129,569,577-129,569,618) but with the ATF1 binding region intact (CC1) and was retained as a negative control.

Table 11. gRNA and sequencing primers used in CRISPR experiments

Oligos	Primer Name	Sequence (5' to 3')
gRNA1	rs216Guide3Fwd	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGTTCTCAAAACAAGTTTGCC
gRNA1	rs216Guide3Rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGCAAACTTGTTTTGAGAAC
gRNA2	rs216Guide4Fwd	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGTTTGCCAGGCAAGTATTAT
gRNA2	rs216Guide4Rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACATAATACTTGCCTGGCAAAC
gRNA3	rs216Guide5Fwd	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGAACCTCAGTGAAGTTTATC
gRNA3	rs216Guide5Rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGATAAACTTCACTGAGGTTC
gRNA4	rs216Guide6Fwd	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGTCCAAAGTCACACAGCGCT
gRNA4	rs216Guide6Rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGCGCTGTGTGACTTTGGAC
gRNA5	rs216Guide7Fwd	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGTCCAAAGTCACACAGCGC
gRNA5	rs216Guide7Rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGCGCTGTGTGACTTTGGACC
gRNA6	rs216Guide8Fwd	TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGTAGGTGTGCAGTAATGCAT
gRNA6	rs216Guide8Rev	GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACATGCATTACTGCACACCTAC
Sequencing primer F	CRISPRfwd	TGGTCCCACTACATACCTAGCA
Sequencing primer R	CRISPRrev	GAGAGCCTGGCACAAAGTAAG

Gene Expression

RNA from CRISPR clones was extracted using the Qiagen RNeasy Mini-Prep Kit and reverse transcription was performed using the Qiagen QuantiTect Reverse Transcription Kit per manufacturer's specifications. Quantitative PCR was performed

using TaqMan gene expression assays for *MYC* (Hs00153408_m1), *PVT1* (Hs00413039_m1) and *TMEM75* (Hs02597353_s1). Eukaryotic 18S rRNA assay was used as internal control (Thermo Fisher, 4319413E). The delta-delta method was used to determine levels of expression compared to the negative CRISPR clone in two biological replicates.

DISCUSSION AND FUTURE WORK

We present a comprehensive fine mapping and functional analysis of the 8q24.21 ovarian cancer locus to identify biological mechanisms underlying disease susceptibility. Fine-mapping revealed rs1400482 to have the most significant association with ovarian cancer ($p = 2.5x10^{-13}$), especially with the serous subtype ($p = 1.9 x10^{-20}$). Log likelihood statistics identified 52 SNPs with odds of 100:1 or better of being true causal variants narrowing down the signal of association from a ~500 kb to a ~31 kb region. Three additional independent signals were found, including rs4524749 located proximal to *MYC* and rs74559819 located inside the *PVT1* region, and rs2165805 which reached genome-wide significance for association with clear cell ovarian cancer.

For the most significant association, overlaying chromatin features data identified eight functional SNPs (Table 3, Figure 5) that are in strong LD ($r^2 = 1$) with each other as well as the most highly associated SNP from fine-mapping (rs1400482) and the initial tagSNP (rs10088218). The identified SNPs mark two regulatory regions (E2 and E4) that modulate transcription of *MYC* and/or *PVT1* genes *via*, at least partially, the recruitment of the ATF1 transcription factor (Figure 23).

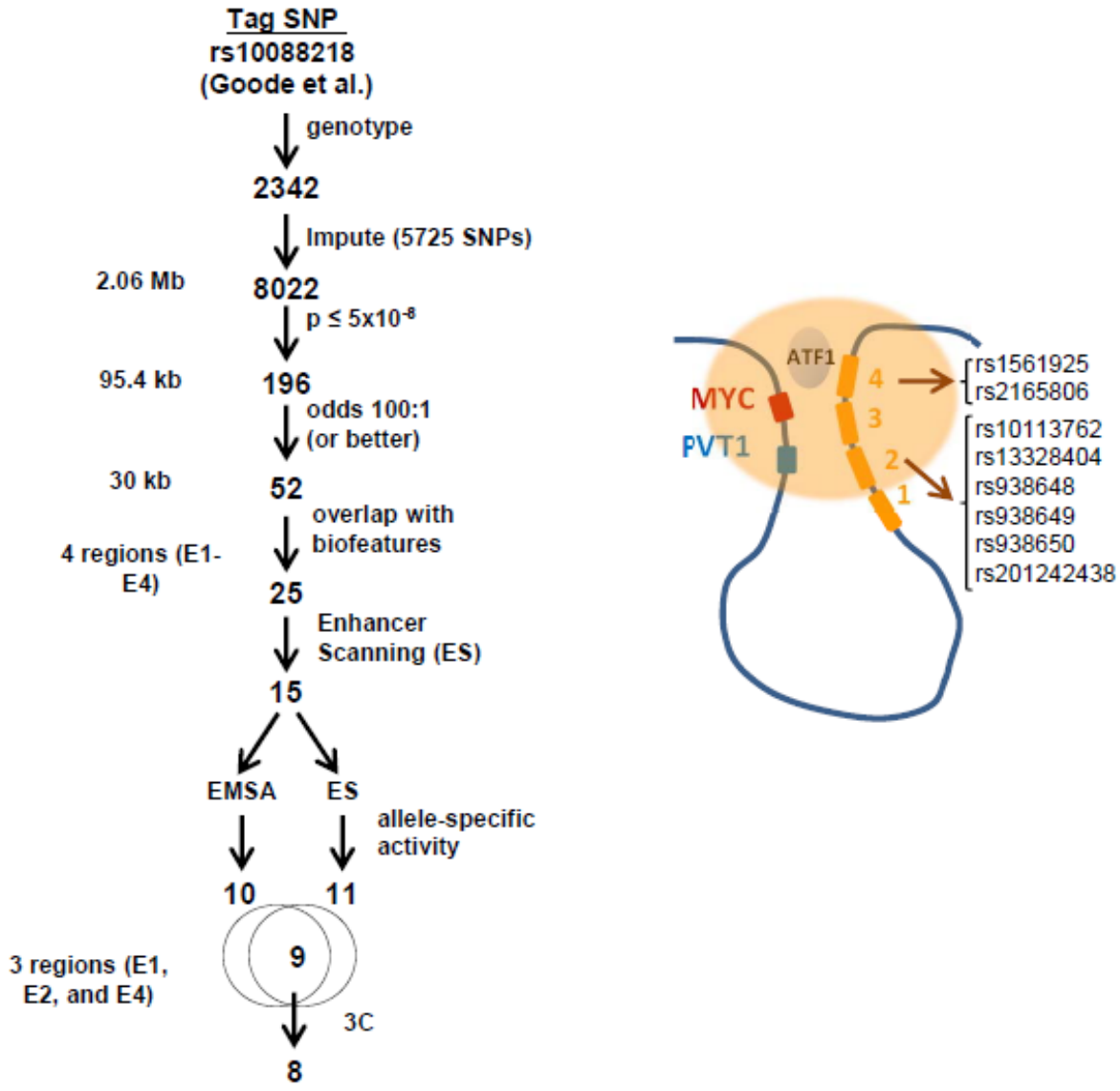


Figure 23. Summary of functional analyses of ovarian cancer susceptibility at the 8q24 locus. Rules guiding SNP selection and prioritization of causal SNPs. The study identified two enhancer regions with 8 causal SNPs that interact with *MYC* and, to a lesser extent, *PVT1* promoter regions.

The decreased risk haplotype (defined by [A] allele of rs10088218) contains the ATF1 binding site (rs2165806 effect allele, which is [C] in forward orientation) suggesting that the absence of binding of ATF1 leads to deregulated higher *MYC*

expression. This is consistent with enhancer scanning data in which a tile with the [C] allele shows decreased activity when compared to the major [G] allele (Figure 5D). Plausibly, a relatively higher basal level of *MYC* expression could lead to oncogenesis.

Several lines of evidence support the notion that deregulation of *MYC* and *PVT1* lead to ovarian oncogenesis. *MYC* is a transcription factor known to play a central role in carcinogenesis (Hermeking et al., 2000) and can cooperate with other somatic alterations to transform ovarian and fallopian cells when ectopically overexpressed (Lawrenson et al., 2010; Xing and Orsulic, 2006). Although *MYC* is tightly regulated and subtle changes in its protein levels may not be detectable by experimental approaches, it is possible that even a slight fluctuation in its protein levels may have a major impact on various cellular processes (Li et al., 2003). Interestingly, *MYC* may be regulated through a positive feedback type of interaction wherein *MYC* activates transcription of *PVT1,* which in turn promotes stability of the MYC protein (Carramusa et al., 2007; Tseng et al., 2014). Of note, conditional analysis revealed an independent peak of association that was not correlated with the top SNP but located proximal to the *MYC* promoter and inside the *PVT1* region (Figure 6).

Over-expression or amplification of *MYC* occurs in up to 65% of EOCs (Chen et al., 2005; Wisman et al., 2003), and amplification of 8q24 (including *MYC* and *PVT1*) has been found to be correlated with poorer survival in 380 stage I to III ovarian tumors when compared to patients with tumors without 8q24 amplification (Guan et al., 2007). The region commonly shows focal amplification in high grade serous tumors (Cancer Genome Atlas Research, 2011). In addition to correlative studies, *in vitro* models of ovarian cancer progression also support a role for *MYC* and *PVT1* in ovarian cancer.

Immortalized ovarian cells transformed with *MYC* show expression signatures strongly correlated with those of invasive high-grade, but not low malignant potential, EOCs (Lawrenson et al., 2011). Also, siRNA-mediated silencing of *PVT1* expression has been seen to result in decreased cell proliferation and increased apoptosis in breast and ovarian cancer cell lines, an effect restricted to cells in which *PVT1* was amplified (Guan et al., 2007).

Although ATF1 has been reported to interact with and to be co-activated by BRCA1 (Houvras et al., 2000) a direct role for ATF1 in ovarian cancer has not been previously reported. Interestingly, inspection of the *MYC* region for ENCODE data for BRCA1 and ATF1 ChIP-Seq revealed a peak of BRCA1 binding in HeLa cells which overlaps with ATF1 binding in K562 cells (the only cell line with data for ATF1 binding) (Figure 22), raising the possibility of BRCA1 involvement.

This locus has also been associated with several other traits at genome-wide significance levels. In particular, the cluster of functional SNPs identified contains a SNP that has been associated with increased risk of two chronic inflammatory diseases, rheumatoid arthritis (rs1516971) (Okada et al., 2014) and Crohn's disease (rs6651252) (Franke et al., 2010). Rs6651252 is contained in Region E3 which showed significant interactions with the promoter of *MYC* but the SNP failed to show allele-specific activity in ovarian cells (Table 3). Given previously reported associations between inflammation and ovarian cancer risk (Charbonneau et al., 2013; Risch and Howe, 1995) further dissection of these interactions is warranted.

In terms of clinical implications of our findings, further work is needed to evaluate the cumulative effects of common genetic variants in development of disease. We are

just beginning to unveil the mechanisms by which these variants could influence disease. At this stage, given their small genetic effect, genetic testing would not be recommended due to our uncertainty in interpreting the results of such testing. Another possibility is to consider MYC and/or PVT1 as therapeutic targets for cancer patients. Myc particularly is an attractive therapeutic target because it is implicated in a wide variety of human cancers, and drugs that specifically target Myc could, therefore, have an impact in the treatment of a broad range of malignancies. There are several strategies that can be explored to inactivate expression or function of Myc at different biological levels by utilizing various novel approaches such as antisense oligonucleotides, siRNA, or small molecule inhibitors, which can be employed alone or in combination with current therapies. However, there are several challenges that limit the application of these approaches including the stability/degradation of these compounds under physiological conditions, delivery methods and uptake by target cell population, drug-related toxicity and resistance, possible side effects, etc. There is tremendous potential that remains to be explored toward clinical applications of these findings, and it is likely that successful interventions will combine attacks at multiple levels to achieve long-term success.

A particular strength of this project is the very large sample size that provided ample power to identify risk associations and to prioritize the most likely causal SNPs for functional analysis. Limitations of this work include the use of data derived from cell lines and from non-chromatinized assays that may not represent the tissue and cell nuclear environments of the cells from which ovarian neoplasms originate. The use of a panel of normal ovarian and cancer cell lines, multiple functional assays, and the

integration of publically available data reduces those limitations and provides a framework for more in-depth functional analyses. Our data implicate *MYC* as the nexus of a regulatory network in the 8q24.21 locus that drives susceptibility to many cancers including ovarian cancer. Further analysis will be necessary to understand the general and cell type-specific pleiotropic contributions of individual enhancer elements to risk of ovarian and other cancers and to other diseases.

LIST OF REFERENCES

Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nature reviews Genetics *13*, 720-731.

Adhikary, S., and Eilers, M. (2005). Transcriptional regulation and transformation by Myc proteins. Nature reviews Molecular cell biology *6*, 635-645.

Ahmadiyeh, N., Pomerantz, M.M., Grisanzio, C., Herman, P., Jia, L., Almendro, V., He, H.H., Brown, M., Liu, X.S., Davis, M.*, et al.* (2010). 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proceedings of the National Academy of Sciences of the United States of America *107*, 9742-9746.

Alcina, A., Fedetz, M., Fernandez, O., Saiz, A., Izquierdo, G., Lucas, M., Leyva, L., Garcia-Leon, J.A., Abad-Grau Mdel, M., Alloza, I.*, et al.* (2013). Identification of a functional variant in the KIF5A-CYP27B1-METTL1-FAM119B locus associated with multiple sclerosis. Journal of medical genetics *50*, 25-33.

Altaf, M., Auger, A., Monnet-Saksouk, J., Brodeur, J., Piquet, S., Cramet, M., Bouchard, N., Lacoste, N., Utley, R.T., Gaudreau, L.*, et al.* (2010). NuA4-dependent acetylation of nucleosomal histones H4 and H2A directly stimulates incorporation of H2A.Z by the SWR1 complex. The Journal of biological chemistry *285*, 15966-15977.

Amundadottir, L.T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnarsson, B.A., Sigurdsson, A., Benediktsdottir, K.R., Cazier, J.B., Sainz, J.*, et al.* (2006). A common variant associated with prostate cancer in European and African populations. Nature genetics *38*, 652-658.

Andersson, R., Sandelin, A., and Danko, C.G. (2015). A unified architecture of transcriptional regulatory elements. Trends in genetics : TIG.

Antoniou, A.C., Pharoah, P.D., McMullan, G., Day, N.E., Stratton, M.R., Peto, J., Ponder, B.J., and Easton, D.F. (2002). A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. British journal of cancer *86*, 76-83.

Ashworth, A. (2008). A synthetic lethal therapeutic approach: poly(ADP) ribose polymerase inhibitors for the treatment of cancers deficient in DNA double-strand break repair. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *26*, 3785-3790.

Auersperg, N., Maines-Bandiera, S.L., and Dyck, H.G. (1997). Ovarian carcinogenesis and the biology of ovarian surface epithelium. Journal of cellular physiology *173*, 261-265.

Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature *226*, 1209-1211.

Bancaud, A., Lavelle, C., Huet, S., and Ellenberg, J. (2012). A fractal model for nuclear organization: current evidence and biological implications. Nucleic acids research *40*, 8783-8792.

Barbieri, C.E., Baca, S.C., Lawrence, M.S., Demichelis, F., Blattner, M., Theurillat, J.P., White, T.A., Stojanov, P., Van Allen, E., Stransky, N., *et al.* (2012). Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nature genetics *44*, 685-689.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bast, R.C., Jr., Hennessy, B., and Mills, G.B. (2009). The biology of ovarian cancer: new opportunities for translation. Nature reviews Cancer *9*, 415-428.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387-396.

Bell, A.C., West, A.G., and Felsenfeld, G. (2001). Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. Science *291*, 447-450.

Bell, J., Brady, M.F., Young, R.C., Lage, J., Walker, J.L., Look, K.Y., Rose, G.S., Spirtos, N.M., and Gynecologic Oncology, G. (2006). Randomized phase III trial of three versus six cycles of adjuvant carboplatin and paclitaxel in early stage epithelial ovarian carcinoma: a Gynecologic Oncology Group study. Gynecologic oncology *102*, 432-439.

Benedet, J.L., Bender, H., Jones, H., 3rd, Ngan, H.Y., and Pecorelli, S. (2000). FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. FIGO Committee on Gynecologic Oncology. International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics *70*, 209-262.

Berns, E.M., and Bowtell, D.D. (2012). The changing view of high-grade serous ovarian cancer. Cancer research *72*, 2701-2704.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M.*, et al.* (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899-905.

Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E.*, et al.* (2013). Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. Nature genetics *45*, 371-384, 384e371-372.

Bolton, K.L., Ganda, C., Berchuck, A., Pharaoh, P.D., and Gayther, S.A. (2012). Role of common genetic variants in ovarian cancer susceptibility and outcome: progress to date from the Ovarian Cancer Association Consortium (OCAC). Journal of internal medicine *271*, 366-378.

Bolton, K.L., Tyrer, J., Song, H., Ramus, S.J., Notaridou, M., Jones, C., Sher, T., Gentry-Maharaj, A., Wozniak, E., Tsai, Y.Y.*, et al.* (2010). Common variants at 19p13 are associated with susceptibility to ovarian cancer. Nature genetics *42*, 880-884.

Bondarenko, V.A., Liu, Y.V., Jiang, Y.I., and Studitsky, V.M. (2003). Communication over a large distance: enhancers and insulators. Biochemistry and cell biology = Biochimie et biologie cellulaire *81*, 241-251.

Bonome, T., Lee, J.Y., Park, D.C., Radonovich, M., Pise-Masison, C., Brady, J., Gardner, G.J., Hao, K., Wong, W.H., Barrett, J.C.*, et al.* (2005). Expression profiling of serous low malignant potential, low-grade, and high-grade tumors of the ovary. Cancer research *65*, 10602-10612.

Bray, F., Loos, A.H., Tognazzo, S., and La Vecchia, C. (2005). Ovarian cancer in Europe: Cross-sectional trends in incidence and mortality in 28 countries, 1953-2000. International journal of cancer Journal international du cancer *113*, 977-990.

Brisbin, A.G., Asmann, Y.W., Song, H., Tsai, Y.Y., Aakre, J.A., Yang, P., Jenkins, R.B., Pharoah, P., Schumacher, F., Conti, D.V.*, et al.* (2011). Meta-analysis of 8q24 for seven cancers reveals a locus between NOV and ENPP2 associated with cancer development. BMC medical genetics *12*, 156.

Bristow, R.E., Tomacruz, R.S., Armstrong, D.K., Trimble, E.L., and Montz, F.J. (2002). Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *20*, 1248-1259.

Brown, J.M., Green, J., das Neves, R.P., Wallace, H.A., Smith, A.J., Hughes, J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R.*, et al.* (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. The Journal of cell biology *182*, 1083-1097.

Bryant, H.E., Schultz, N., Thomas, H.D., Parker, K.M., Flower, D., Lopez, E., Kyle, S., Meuth, M., Curtin, N.J., and Helleday, T. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature *434*, 913-917.

Buckley, M., Gjyshi, A., Mendoza-Fandino, G., Baskin, R., Carvalho, R.S., Carvalho, M.A., Woods, N.T., and Monteiro, A.N. (2016). Enhancer scanning to locate regulatory regions in genomic loci. Nature protocols *11*, 46-60.

Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. Trends in genetics : TIG *28*, 276-284.

Bukowski, R.M., Ozols, R.F., and Markman, M. (2007). The management of recurrent ovarian cancer. Seminars in oncology *34*, S1-15.

Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., and Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. Proceedings of the National Academy of Sciences of the United States of America *99 Suppl 4*, 16433-16437.

Cairns, B.R. (2007). Chromatin remodeling: insights and intrigue from single-molecule studies. Nature structural & molecular biology *14*, 989-996.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Molecular cell *49*, 825-837.

Cancer Genome Atlas Research, N. (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609-615.

Carden, C.P., Yap, T.A., and Kaye, S.B. (2010). PARP inhibition: targeting the Achilles' heel of DNA repair to treat germline and sporadic ovarian cancers. Current opinion in oncology *22*, 473-480.

Cardon, L.R., and Abecasis, G.R. (2003). Using haplotype blocks to map human complex trait loci. Trends in genetics : TIG *19*, 135-140.

Cardon, L.R., and Bell, J.I. (2001). Association study designs for complex diseases. Nature reviews Genetics *2*, 91-99.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N.*, et al.* (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. Nature genetics *22*, 231-238.

Carlson, C.S., Eberle, M.A., Kruglyak, L., and Nickerson, D.A. (2004). Mapping complex disease loci in whole-genome association studies. Nature *429*, 446-452.

Carramusa, L., Contino, F., Ferro, A., Minafra, L., Perconti, G., Giallongo, A., and Feo, S. (2007). The PVT-1 oncogene is a Myc protein target that is overexpressed in transformed cells. Journal of cellular physiology *213*, 511-518.

Castera, L., Krieger, S., Rousselin, A., Legros, A., Baumann, J.J., Bruet, O., Brault, B., Fouillet, R., Goardon, N., Letac, O.*, et al.* (2014). Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. European journal of human genetics : EJHG *22*, 1305-1313.

Charbonneau, B., Goode, E.L., Kalli, K.R., Knutson, K.L., and Derycke, M.S. (2013). The immune system in the pathogenesis of ovarian cancer. Critical reviews in immunology *33*, 137-164.

Chen, C.H., Shen, J., Lee, W.J., and Chow, S.N. (2005). Overexpression of cyclin D1 and c-Myc gene products in human primary epithelial ovarian cancer. International journal of gynecological cancer : official journal of the International Gynecological Cancer Society *15*, 878-883.

Chen, K., Ma, H., Li, L., Zang, R., Wang, C., Song, F., Shi, T., Yu, D., Yang, M., Xue, W.*, et al.* (2014). Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women. Nature communications *5*, 4682.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J.*, et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Chiaffarino, F., Parazzini, F., Decarli, A., Franceschi, S., Talamini, R., Montella, M., and La Vecchia, C. (2005). Hysterectomy with or without unilateral oophorectomy and risk of ovarian cancer. Gynecologic oncology *97*, 318-322.

Cho, K.R., and Shih Ie, M. (2009). Ovarian cancer. Annual review of pathology *4*, 287-313.

Chung, C.C., Magalhaes, W.C., Gonzalez-Bosquet, J., and Chanock, S.J. (2010). Genome-wide association studies in cancer--current and future directions. Carcinogenesis *31*, 111-120.

Coetzee, S.G., Shen, H.C., Hazelett, D.J., Lawrenson, K., Kuchenbaecker, K., Tyrer, J., Rhie, S.K., Levanon, K., Karst, A., Drapkin, R.*, et al.* (2015). Cell-type-specific enrichment of risk-associated regulatory elements at ovarian cancer susceptibility loci. Human molecular genetics *24*, 3595-3607.

Cole, M.D. (1986). The myc oncogene: its role in transformation and differentiation. Annual review of genetics *20*, 361-384.

Coleman, M.P., Esteve, J., Damiecki, P., Arslan, A., and Renard, H. (1993). Trends in cancer incidence and mortality. IARC scientific publications, 1-806.

Consortium, E.P. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636-640.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T.*, et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nature genetics *46*, 1311-1320.

Corvin, A., Craddock, N., and Sullivan, P.F. (2010). Genome-wide association studies: a primer. Psychological medicine *40*, 1063-1077.

Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nature reviews Genetics *14*, 572-584.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences of the United States of America *107*, 21931-21936.

Crick, F. (1970). Central dogma of molecular biology. Nature *227*, 561-563.

Crick, F.H. (1958). On protein synthesis. Symposia of the Society for Experimental Biology *12*, 138-163.

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R.C., and Croce, C.M. (1982). Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proceedings of the National Academy of Sciences of the United States of America *79*, 7824-7827.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS biology *8*, e1000384.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes & development *26*, 11-24.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature reviews Genetics *14*, 390-403.

DelloRusso, C., Welcsh, P.L., Wang, W., Garcia, R.L., King, M.C., and Swisher, E.M. (2007). Functional characterization of a novel BRCA1-null ovarian cancer cell line in response to ionizing radiation. Molecular cancer research : MCR *5*, 35-45.

Diesinger, P.M., and Heermann, D.W. (2009). Depletion effects massively change chromatin properties and influence genome folding. Biophysical journal *97*, 2146-2153.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic acids research *11*, 1475-1489.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F.*, et al.* (2012). Landscape of transcription in human cells. Nature *489*, 101-108.

Domcke, S., Sinha, R., Levine, D.A., Sander, C., and Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nature communications *4*, 2126.

Durst, M., Croce, C.M., Gissmann, L., Schwarz, E., and Huebner, K. (1987). Papillomavirus sequences integrate near cellular oncogenes in some cervical carcinomas. Proceedings of the National Academy of Sciences of the United States of America *84*, 1070-1074.

Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R.*, et al.* (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature *447*, 1087-1093.

Edelman, L.B., and Fraser, P. (2012). Transcription factories: genetic programming in three dimensions. Current opinion in genetics & development *22*, 110-114.

Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J.*, et al.* (2008). Multiple newly identified loci associated with prostate cancer susceptibility. Nature genetics *40*, 316-321.

Egan, K.M. (2006). Commentary: sunlight, vitamin D, and the cancer connection revisited. International journal of epidemiology *35*, 227-230.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M.*, et al.* (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43-49.

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., Holstege, F., and Werner, M. (2008). Mediator-dependent recruitment of TFIIH modules in preinitiation complex. Molecular cell *31*, 337-346.

Fairfield, K.M., Hunter, D.J., Fuchs, C.S., Colditz, G.A., and Hankinson, S.E. (2002). Aspirin, other NSAIDs, and ovarian cancer risk (United States). Cancer causes & control : CCC *13*, 535-542.

Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C.*, et al.* (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature *434*, 917-921.

Fasching, P.A., Gayther, S., Pearce, L., Schildkraut, J.M., Goode, E., Thiel, F., Chenevix-Trench, G., Chang-Claude, J., Wang-Gohrke, S., Ramus, S.*, et al.* (2009). Role of genetic polymorphisms and ovarian cancer susceptibility. Molecular oncology *3*, 171-181.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Fong, P.C., Boss, D.S., Yap, T.A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M.J.*, et al.* (2009). Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. The New England journal of medicine *361*, 123-134.

Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R.*, et al.* (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nature genetics *42*, 1118-1125.

Freedman, M.L., Monteiro, A.N., Gayther, S.A., Coetzee, G.A., Risch, A., Plass, C., Casey, G., De Biasi, M., Carlson, C., Duggan, D.*, et al.* (2011). Principles for the post-GWAS functional characterization of cancer risk loci. Nature genetics *43*, 513-518.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186-192.

Fujita, M., Enomoto, T., Yoshino, K., Nomura, T., Buzard, G.S., Inoue, M., and Okudaira, Y. (1995). Microsatellite instability and alterations in the hMSH2 gene in human ovarian cancer. International journal of cancer Journal international du cancer *64*, 361-366.

Gaj, P., Paziewska, A., Bik, W., Dabrowska, M., Baranowska-Bik, A., Styczynska, M., Chodakowska-Zebrowska, M., Pfeffer-Baczuk, A., Barcikowska, M., Baranowska, B.*, et al.* (2012). Identification of a late onset Alzheimer's disease candidate risk variant at 9q21.33 in Polish patients. Journal of Alzheimer's disease : JAD *32*, 157-168.

Gayther, S.A., Russell, P., Harrington, P., Antoniou, A.C., Easton, D.F., and Ponder, B.A. (1999). The contribution of germline BRCA1 and BRCA2 mutations to familial ovarian cancer: no evidence for other ovarian cancer-susceptibility genes. American journal of human genetics *65*, 1021-1029.

Genomes Project, C., Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061-1073.

Genomes Project, C., Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56-65.

Gerchman, S.E., and Ramakrishnan, V. (1987). Chromatin higher-order structure studied by neutron scattering and scanning transmission electron microscopy. Proceedings of the National Academy of Sciences of the United States of America *84*, 7802-7806.

Gerstein, M.B., Bruce, C., Rozowsky, J.S., Zheng, D., Du, J., Korbel, J.O., Emanuelsson, O., Zhang, Z.D., Weissman, S., and Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. Genome research *17*, 669-681.

Ghoussaini, M., Song, H., Koessler, T., Al Olama, A.A., Kote-Jarai, Z., Driver, K.E., Pooley, K.A., Ramus, S.J., Kjaer, S.K., Hogdall, E.*, et al.* (2008). Multiple loci with different cancer specificities within the 8q24 gene desert. Journal of the National Cancer Institute *100*, 962-966.

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Molecular cell *49*, 773-782.

Gien, L.T., and Mackay, H.J. (2010). The Emerging Role of PARP Inhibitors in the Treatment of Epithelial Ovarian Cancer. Journal of oncology *2010*, 151750.

Gilbert, L., Basso, O., Sampalis, J., Karp, I., Martins, C., Feng, J., Piedimonte, S., Quintal, L., Ramanakumar, A.V., Takefman, J.*, et al.* (2012). Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. The Lancet Oncology *13*, 285-291.

Gilks, C.B. (2004). Subclassification of ovarian surface epithelial tumors based on correlation of histologic and molecular pathologic data. International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists *23*, 200-205.

Giresi, P.G., Kim, J., McDaniell, R.M., Iyer, V.R., and Lieb, J.D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome research *17*, 877-885.

Giresi, P.G., and Lieb, J.D. (2009). Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). Methods *48*, 233-239.

Goff, B.A., Mandel, L., Muntz, H.G., and Melancon, C.H. (2000). Ovarian carcinoma diagnosis. Cancer *89*, 2068-2075.

Goff, B.A., Mandel, L.S., Drescher, C.W., Urban, N., Gough, S., Schurman, K.M., Patras, J., Mahony, B.S., and Andersen, M.R. (2007). Development of an ovarian cancer symptom index: possibilities for earlier detection. Cancer *109*, 221-227.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X.*, et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678-691.

Goldgar, D.E., Easton, D.F., Cannon-Albright, L.A., and Skolnick, M.H. (1994). Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. Journal of the National Cancer Institute *86*, 1600-1608.

Goode, E.L., Chenevix-Trench, G., Song, H., Ramus, S.J., Notaridou, M., Lawrenson, K., Widschwendter, M., Vierkant, R.A., Larson, M.C., Kjaer, S.K.*, et al.* (2010). A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. Nature genetics *42*, 874-879.

Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell stem cell *14*, 762-775.

Gorringe, K.L., Jacobs, S., Thompson, E.R., Sridhar, A., Qiu, W., Choong, D.Y., and Campbell, I.G. (2007). High-resolution single nucleotide polymorphism array analysis of epithelial ovarian cancer reveals numerous microdeletions and amplifications. Clinical cancer research : an official journal of the American Association for Cancer Research *13*, 4731-4739.

Grisanzio, C., and Freedman, M.L. (2010). Chromosome 8q24-Associated Cancers and MYC. Genes & cancer *1*, 555-559.

Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annual review of biochemistry *57*, 159-197.

Guan, Y., Kuo, W.L., Stilwell, J.L., Takano, H., Lapuk, A.V., Fridlyand, J., Mao, J.H., Yu, M., Miller, M.A., Santos, J.L., *et al.* (2007). Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research *13*, 5745-5755.

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L.T., Gudbjartsson, D., Helgason, A., Rafnar, T., Bergthorsson, J.T., Agnarsson, B.A., Baker, A., *et al.* (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nature genetics *39*, 631-637.

Gunderson, C.C., and Moore, K.N. (2015). Olaparib: an oral PARP-1 and PARP-2 inhibitor with promising activity in ovarian cancer. Future oncology *11*, 747-757.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., *et al.* (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223-227.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nature structural & molecular biology *11*, 394-403.

Haiman, C.A., Le Marchand, L., Yamamato, J., Stram, D.O., Sheng, X., Kolonel, L.N., Wu, A.H., Reich, D., and Henderson, B.E. (2007a). A common genetic risk factor for colorectal and prostate cancer. Nature genetics *39*, 954-956.

Haiman, C.A., Patterson, N., Freedman, M.L., Myers, S.R., Pike, M.C., Waliszewska, A., Neubauer, J., Tandon, A., Schirmer, C., McDonald, G.J., *et al.* (2007b). Multiple regions within 8q24 independently affect risk for prostate cancer. Nature genetics *39*, 638-644.

Han, Y.J., Ma, S.F., Wade, M.S., Flores, C., and Garcia, J.G. (2012). An intronic MYLK variant associated with inflammatory lung disease regulates promoter activity of the smooth muscle myosin light chain kinase isoform. Journal of molecular medicine *90*, 299-308.

Hankinson, S.E., Colditz, G.A., Hunter, D.J., Willett, W.C., Stampfer, M.J., Rosner, B., Hennekens, C.H., and Speizer, F.E. (1995). A prospective study of reproductive factors and risk of epithelial ovarian cancer. Cancer *76*, 284-290.

Hankinson, S.E., Hunter, D.J., Colditz, G.A., Willett, W.C., Stampfer, M.J., Rosner, B., Hennekens, C.H., and Speizer, F.E. (1993). Tubal ligation, hysterectomy, and risk of ovarian cancer. A prospective study. Jama *270*, 2813-2818.

Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. Cell research *21*, 396-420.

Harismendy, O., Notani, D., Song, X., Rahim, N.G., Tanasa, B., Heintzman, N., Ren, B., Fu, X.D., Topol, E.J., Rosenfeld, M.G., *et al.* (2011). 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. Nature *470*, 264-268.

Hartley, P.D., and Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. Cell *137*, 445-458.

Haverty, P.M., Hon, L.S., Kaminker, J.S., Chant, J., and Zhang, Z. (2009). High-resolution analysis of copy number alterations and associated expression changes in ovarian tumors. BMC medical genomics *2*, 21.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., *et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., *et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics *39*, 311-318.

Henikoff, S. (2008). Nucleosome destabilization in the epigenetic regulation of gene expression. Nature reviews Genetics *9*, 15-26.

Henriksson, M., and Luscher, B. (1996). Proteins of the Myc network: essential regulators of cell growth and differentiation. Advances in cancer research *68*, 109-182.

Hermeking, H., Rago, C., Schuhmacher, M., Li, Q., Barrett, J.F., Obaya, A.J., O'Connell, B.C., Mateyak, M.K., Tam, W., Kohlhuber, F., *et al.* (2000). Identification of CDK4 as a target of c-MYC. Proceedings of the National Academy of Sciences of the United States of America *97*, 2229-2234.

Herrick, J., Conti, C., Teissier, S., Thierry, F., Couturier, J., Sastre-Garau, X., Favre, M., Orth, G., and Bensimon, A. (2005). Genomic organization of amplified MYC genes suggests distinct mechanisms of amplification in tumorigenesis. Cancer research *65*, 1174-1179.

Ho, T.H., Park, I.Y., Zhao, H., Tong, P., Champion, M.D., Yan, H., Monzon, F.A., Hoang, A., Tamboli, P., Parker, A.S., *et al.* (2016). High-resolution profiling of histone h3 lysine 36 trimethylation in metastatic renal cell carcinoma. Oncogene *35*, 1565-1574.

Hobert, O. (2008). Gene regulation by transcription factors and microRNAs. Science *319*, 1785-1786.

Houvras, Y., Benezra, M., Zhang, H., Manfredi, J.J., Weber, B.L., and Licht, J.D. (2000). BRCA1 physically and functionally interacts with the ATF1 transcription factor. J BiolChem *275*, 36230-36237.

Howe, H.L., Wu, X., Ries, L.A., Cokkinides, V., Ahmed, F., Jemal, A., Miller, B., Williams, M., Ward, E., Wingo, P.A.*, et al.* (2006). Annual report to the nation on the status of cancer, 1975-2003, featuring cancer among U.S. Hispanic/Latino populations. Cancer *107*, 1711-1742.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics *5*, e1000529.

Hu, Q., Kwon, Y.S., Nunez, E., Cardamone, M.D., Hutt, K.R., Ohgi, K.A., Garcia-Bassets, I., Rose, D.W., Glass, C.K., Rosenfeld, M.G.*, et al.* (2008). Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. Proceedings of the National Academy of Sciences of the United States of America *105*, 19199-19204.

Hubner, M.R., Eckersley-Maslin, M.A., and Spector, D.L. (2013). Chromatin organization and transcriptional regulation. Current opinion in genetics & development *23*, 89-95.

Huncharek, M., Geschwind, J.F., and Kupelnick, B. (2003). Perineal application of cosmetic talc and risk of invasive epithelial ovarian cancer: a meta-analysis of 11,933 subjects from sixteen observational studies. Anticancer research *23*, 1955-1960.

Huppi, K., Pitt, J.J., Wahlberg, B.M., and Caplen, N.J. (2012). The 8q24 gene desert: an oasis of non-coding transcriptional activity. Frontiers in genetics *3*, 69.

International HapMap, C. (2003). The International HapMap Project. Nature *426*, 789-796.

International HapMap, C., Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P.*, et al.* (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851-861.

Irwin, K.L., Weiss, N.S., Lee, N.C., and Peterson, H.B. (1991). Tubal sterilization, hysterectomy, and the subsequent occurrence of epithelial ovarian cancer. American journal of epidemiology *134*, 362-369.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of molecular biology *3*, 318-356.

Jelovac, D., and Armstrong, D.K. (2011). Recent progress in the diagnosis and treatment of ovarian cancer. CA: a cancer journal for clinicians *61*, 183-203.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., and Thun, M.J. (2009). Cancer statistics, 2009. CA: a cancer journal for clinicians *59*, 225-249.

Jia, L., Landan, G., Pomerantz, M., Jaschek, R., Herman, P., Reich, D., Yan, C., Khalid, O., Kantoff, P., Oh, W.*, et al.* (2009). Functional enhancers at the gene-poor 8q24 cancer-linked locus. PLoS genetics *5*, e1000597.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & development *21*, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nature genetics *41*, 941-945.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290-294.

Johnson, G.C., Esposito, L., Barratt, B.J., Smith, A.N., Heward, J., Di Genova, G., Ueda, H., Cordell, H.J., Eaves, I.A., Dudbridge, F.*, et al.* (2001). Haplotype tagging for the identification of common disease genes. Nature genetics *29*, 233-237.

Jordan, S.J., Whiteman, D.C., Purdie, D.M., Green, A.C., and Webb, P.M. (2006). Does smoking increase risk of ovarian cancer? A systematic review. Gynecologic oncology *103*, 1122-1129.

Juven-Gershon, T., Hsu, J.Y., Theisen, J.W., and Kadonaga, J.T. (2008). The RNA polymerase II core promoter - the gateway to transcription. Current opinion in cell biology *20*, 253-259.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S.*, et al.* (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430-435.

Kar, S.P., Beesley, J., Amin Al Olama, A., Michailidou, K., Tyrer, J., Kote-Jarai, Z., Lawrenson, K., Lindstrom, S., Ramus, S.J., Thompson, D.J.*, et al.* (2016). Genome-Wide Meta-Analyses of Breast, Ovarian, and Prostate Cancer Association Studies Identify Multiple New Susceptibility Loci Shared by at Least Two Cancer Types. Cancer discovery *6*, 1052-1067.

Karst, A.M., Levanon, K., and Drapkin, R. (2011). Modeling high-grade serous ovarian carcinogenesis from the fallopian tube. Proceedings of the National Academy of Sciences of the United States of America *108*, 7547-7552.

Kauff, N.D., Domchek, S.M., Friebel, T.M., Robson, M.E., Lee, J., Garber, J.E., Isaacs, C., Evans, D.G., Lynch, H., Eeles, R.A.*, et al.* (2008). Risk-reducing salpingo-oophorectomy for the prevention of BRCA1- and BRCA2-associated breast and gynecologic cancer: a multicenter, prospective study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *26*, 1331-1337.

Kelemen, L.E. (2015). New mucinous ovarian cancer loci. In preparation.

Kellum, R., and Schedl, P. (1991). A position-effect assay for boundaries of higher order chromosomal domains. Cell *64*, 941-950.

Kiemeney, L.A., Thorlacius, S., Sulem, P., Geller, F., Aben, K.K., Stacey, S.N., Gudmundsson, J., Jakobsdottir, M., Bergthorsson, J.T., Sigurdsson, A.*, et al.* (2008). Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. Nature genetics *40*, 1307-1312.

Kim, J., Coffey, D.M., Creighton, C.J., Yu, Z., Hawkins, S.M., and Matzuk, M.M. (2012). High-grade serous ovarian cancer arises from fallopian tube in a mouse model. Proceedings of the National Academy of Sciences of the United States of America *109*, 3921-3926.

Kim, T.H., Barrera, L.O., Qu, C., Van Calcar, S., Trinklein, N.D., Cooper, S.J., Luna, R.M., Glass, C.K., Rosenfeld, M.G., Myers, R.M.*, et al.* (2005a). Direct isolation and identification of promoters in the human genome. Genome research *15*, 830-839.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005b). A high-resolution map of active promoters in the human genome. Nature *436*, 876-880.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S.*, et al.* (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182-187.

Kim, Y.J., Bjorklund, S., Li, Y., Sayre, M.H., and Kornberg, R.D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. Cell *77*, 599-608.

Kobel, M., Kalloger, S.E., Huntsman, D.G., Santos, J.L., Swenerton, K.D., Seidman, J.D., Gilks, C.B., and Cheryl Brown Ovarian Cancer Outcomes Unit of the British Columbia Cancer Agency, V.B.C. (2010). Differences in tumor type in low-stage versus high-stage ovarian carcinomas. International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists *29*, 203-211.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P.*, et al.* (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome research *17*, 691-707.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C.*, et al.* (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. Nature structural & molecular biology *18*, 956-963.

Koonings, P.P., Campbell, K., Mishell, D.R., Jr., and Grimes, D.A. (1989). Relative frequency of primary ovarian neoplasms: a 10-year review. Obstetrics and gynecology *74*, 921-926.

Kraus, I., Driesch, C., Vinokurova, S., Hovig, E., Schneider, A., von Knebel Doeberitz, M., and Durst, M. (2008). The majority of viral-cellular fusion transcripts in cervical carcinomas cotranscribe cellular sequences of known or predicted genes. Cancer research *68*, 2514-2522.

Krivega, I., and Dean, A. (2012). Enhancer and promoter interactions-long distance calls. Current opinion in genetics & development *22*, 79-85.

Kuchenbaecker, K.B., Ramus, S.J., Tyrer, J., Lee, A., Shen, H.C., Beesley, J., Lawrenson, K., McGuffog, L., Healey, S., Lee, J.M.*, et al.* (2015). Identification of six new susceptibility loci for invasive epithelial ovarian cancer. Nature genetics *47*, 164-171.

Kuo, K.T., Guan, B., Feng, Y., Mao, T.L., Chen, X., Jinawath, N., Wang, Y., Kurman, R.J., Shih Ie, M., and Wang, T.L. (2009). Analysis of DNA copy number alterations in ovarian serous tumors identifies new molecular genetic changes in low-grade and high-grade carcinomas. Cancer research *69*, 4036-4042.

Kurman, R.J., and Shih Ie, M. (2008). Pathogenesis of ovarian cancer: lessons from morphology and molecular biology and their clinical implications. International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists *27*, 151-160.

Kurman, R.J., and Shih Ie, M. (2010). The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. The American journal of surgical pathology *34*, 433-443.

Kuusisto, K.M., Bebel, A., Vihinen, M., Schleutker, J., and Sallinen, S.L. (2011). Screening for BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50, and CDH1 mutations in high-risk Finnish BRCA1/2-founder mutation-negative breast and/or ovarian cancer individuals. Breast cancer research : BCR *13*, R20.

Landen, C.N., Jr., Birrer, M.J., and Sood, A.K. (2008). Early events in the pathogenesis of epithelial ovarian cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *26*, 995-1005.

Lawrenson, K., Benjamin, E., Turmaine, M., Jacobs, I., Gayther, S., and Dafou, D. (2009). In vitro three-dimensional modelling of human ovarian surface epithelial cells. Cell proliferation *42*, 385-393.

Lawrenson, K., Grun, B., Benjamin, E., Jacobs, I.J., Dafou, D., and Gayther, S.A. (2010). Senescent fibroblasts promote neoplastic transformation of partially transformed ovarian epithelial cells in a three-dimensional model of early stage ovarian cancer. Neoplasia *12*, 317-325.

Lawrenson, K., Grun, B., and Gayther, S.A. (2012). Heterotypic three-dimensional in vitro modeling of stromal-epithelial interactions during ovarian cancer initiation and progression. Journal of visualized experiments : JoVE, e4206.

Lawrenson, K., Grun, B., Lee, N., Mhawech-Fauceglia, P., Kan, J., Swenson, S., Lin, Y.G., Pejovic, T., Millstein, J., and Gayther, S.A. (2015a). NPPB is a novel candidate biomarker expressed by cancer-associated fibroblasts in epithelial ovarian cancer. International journal of cancer Journal international du cancer *136*, 1390-1401.

Lawrenson, K., Mhawech-Fauceglia, P., Worthington, J., Spindler, T.J., O'Brien, D., Lee, J.M., Spain, G., Sharifian, M., Wang, G., Darcy, K.M.*, et al.* (2015b). Identification of novel candidate biomarkers of epithelial ovarian cancer by profiling the secretomes of three-dimensional genetic models of ovarian carcinogenesis. International journal of cancer Journal international du cancer *137*, 1806-1817.

Lawrenson, K., Notaridou, M., Lee, N., Benjamin, E., Jacobs, I.J., Jones, C., and Gayther, S.A. (2013). In vitro three-dimensional modeling of fallopian tube secretory epithelial cells. BMC cell biology *14*, 43.

Lawrenson, K., Sproul, D., Grun, B., Notaridou, M., Benjamin, E., Jacobs, I.J., Dafou, D., Sims, A.H., and Gayther, S.A. (2011). Modelling genetic and clinical heterogeneity in epithelial ovarian cancers. Carcinogenesis *32*, 1540-1549.

Lee, K.R., and Young, R.H. (2003). The distinction between primary and metastatic mucinous carcinomas of the ovary: gross and histologic findings in 50 cases. The American journal of surgical pathology *27*, 281-292.

Lee, T.I., and Young, R.A. (2013). Transcriptional regulation and its misregulation in disease. Cell *152*, 1237-1251.

Lee, Y., Miron, A., Drapkin, R., Nucci, M.R., Medeiros, F., Saleemuddin, A., Garber, J., Birch, C., Mou, H., Gordon, R.W., *et al.* (2007). A candidate precursor to serous carcinoma that originates in the distal fallopian tube. The Journal of pathology *211*, 26-35.

Lefkowitz, E.S., and Garland, C.F. (1994). Sunlight, vitamin D, and ovarian cancer mortality rates in US women. International journal of epidemiology *23*, 1133-1136.

Lelli, K.M., Slattery, M., and Mann, R.S. (2012). Disentangling the many layers of eukaryotic transcriptional regulation. Annual review of genetics *46*, 43-68.

Levanon, K., Crum, C., and Drapkin, R. (2008). New insights into the pathogenesis of serous ovarian cancer and its clinical impact. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *26*, 5284-5293.

Levo, M., and Segal, E. (2014). In pursuit of design principles of regulatory sequences. Nature reviews Genetics *15*, 453-468.

Li, Z., Gadue, P., Chen, K., Jiao, Y., Tuteja, G., Schug, J., Li, W., and Kaestner, K.H. (2012). Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. Cell *151*, 1608-1616.

Li, Z., Van Calcar, S., Qu, C., Cavenee, W.K., Zhang, M.Q., and Ren, B. (2003). A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proceedings of the National Academy of Sciences of the United States of America *100*, 8164-8169.

Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., and Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. The New England journal of medicine *343*, 78-85.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.* (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lin, C.Y., Loven, J., Rahl, P.B., Paranal, R.M., Burge, C.B., Bradner, J.E., Lee, T.I., and Young, R.A. (2012). Transcriptional amplification in tumor cells with elevated c-Myc. Cell *151*, 56-67.

Lindor, N.M., Petersen, G.M., Hadley, D.W., Kinney, A.Y., Miesfeldt, S., Lu, K.H., Lynch, P., Burke, W., and Press, N. (2006). Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review. Jama *296*, 1507-1517.

Littlewood, T.D., Kreuzaler, P., and Evan, G.I. (2012). All things to all people. Cell *151*, 11-13.

Loft, A., Lidegaard, O., and Tabor, A. (1997). Incidence of ovarian cancer after hysterectomy: a nationwide controlled follow up. British journal of obstetrics and gynaecology *104*, 1296-1301.

Low, S.K., Kuchiba, A., Zembutsu, H., Saito, A., Takahashi, A., Kubo, M., Daigo, Y., Kamatani, N., Chiku, S., Totsuka, H.*, et al.* (2010). Genome-wide association study of pancreatic cancer in Japanese population. PloS one *5*, e11824.

Luo, Z., Lin, C., Guest, E., Garrett, A.S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M.P., and Shilatifard, A. (2012). The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. Molecular and cellular biology *32*, 2608-2617.

Makinen, N., Mehine, M., Tolvanen, J., Kaasinen, E., Li, Y., Lehtonen, H.J., Gentile, M., Yan, J., Enge, M., Taipale, M.*, et al.* (2011). MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. Science *334*, 252-255.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A.*, et al.* (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.

Marcu, K.B., Bossone, S.A., and Patel, A.J. (1992). myc function and regulation. Annual review of biochemistry *61*, 809-860.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J.*, et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190-1195.

McKay, J.D., Hung, R.J., Han, Y., Zong, X., Carreras-Torres, R., Christiani, D.C., Caporaso, N.E., Johansson, M., Xiao, X., Li, Y.*, et al.* (2017). Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nature genetics *49*, 1126-1132.

Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R.L., Schmidt, M.K., Chang-Claude, J., Bojesen, S.E., Bolla, M.K.*, et al.* (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nature genetics *45*, 353-361, 361e351-352.

Monteiro, A.N., and Freedman, M.L. (2013). Lessons from postgenome-wide association studies: functional analysis of cancer predisposition loci. Journal of internal medicine *274*, 414-424.

Morch, L.S., Lokkegaard, E., Andreasen, A.H., Kruger-Kjaer, S., and Lidegaard, O. (2009). Hormone therapy and ovarian cancer. Jama *302*, 298-305.

Myer, V.E., and Young, R.A. (1998). RNA polymerase II holoenzymes and subcomplexes. The Journal of biological chemistry *273*, 27757-27760.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annual review of genetics *46*, 1-19.

Ness, R.B. (2003). Endometriosis and ovarian cancer: thoughts on shared pathophysiology. American journal of obstetrics and gynecology *189*, 280-294.

Nie, Z., Hu, G., Wei, G., Cui, K., Yamane, A., Resch, W., Wang, R., Green, D.R., Tessarollo, L., Casellas, R.*, et al.* (2012). c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells. Cell *151*, 68-79.

Nobrega, M.A., Ovcharenko, I., Afzal, V., and Rubin, E.M. (2003). Scanning human gene deserts for long-range enhancers. Science *302*, 413.

Obata, K., Morland, S.J., Watson, R.H., Hitchcock, A., Chenevix-Trench, G., Thomas, E.J., and Campbell, I.G. (1998). Frequent PTEN/MMAC mutations in endometrioid but not serous or mucinous epithelial ovarian tumors. Cancer research *58*, 2095-2097.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S.*, et al.* (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376-381.

Olins, A.L., and Olins, D.E. (1974). Spheroid chromatin units (v bodies). Science *183*, 330-332.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nature reviews Genetics *12*, 283-293.

Ong, C.T., and Corces, V.G. (2012). Enhancers: emerging roles in cell fate specification. EMBO reports *13*, 423-430.

Orr, N., and Chanock, S. (2008). Common genetic variation and human disease. Advances in genetics *62*, 1-32.

Ovcharenko, I., Loots, G.G., Nobrega, M.A., Hardison, R.C., Miller, W., and Stubbs, L. (2005). Evolution and functional classification of vertebrate gene deserts. Genome research *15*, 137-145.

Ozols, R.F., Bundy, B.N., Greer, B.E., Fowler, J.M., Clarke-Pearson, D., Burger, R.A., Mannel, R.S., DeGeest, K., Hartenbach, E.M., Baergen, R.*, et al.* (2003). Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group study. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *21*, 3194-3200.

Panne, D. (2008). The enhanceosome. Current opinion in structural biology *18*, 236-242.

Partridge, E., Kreimer, A.R., Greenlee, R.T., Williams, C., Xu, J.L., Church, T.R., Kessel, B., Johnson, C.C., Weissfeld, J.L., Isaacs, C.*, et al.* (2009). Results from four rounds of ovarian cancer screening in a randomized trial. Obstetrics and gynecology *113*, 775-782.

Patel, A.V., Rodriguez, C., Pavluck, A.L., Thun, M.J., and Calle, E.E. (2006). Recreational physical activity and sedentary behavior in relation to ovarian cancer risk in a large cohort of US women. American journal of epidemiology *163*, 709-716.

Patel, J.H., Loboda, A.P., Showe, M.K., Showe, L.C., and McMahon, S.B. (2004). Analysis of genomic targets reveals complex functions of MYC. Nature reviews Cancer *4*, 562-568.

Permuth-Wey, J., Kim, D., Tsai, Y.Y., Lin, H.Y., Chen, Y.A., Barnholtz-Sloan, J., Birrer, M.J., Bloom, G., Chanock, S.J., Chen, Z.*, et al.* (2011). LIN28B polymorphisms influence susceptibility to epithelial ovarian cancer. Cancer research *71*, 3896-3903.

Permuth-Wey, J., Lawrenson, K., Shen, H.C., Velkova, A., Tyrer, J.P., Chen, Z., Lin, H.Y., Chen, Y.A., Tsai, Y.Y., Qu, X.*, et al.* (2013). Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. Nature communications *4*, 1627.

Peter, M., Rosty, C., Couturier, J., Radvanyi, F., Teshima, H., and Sastre-Garau, X. (2006). MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. Oncogene *25*, 5985-5993.

Pharoah, P.D., Dunning, A.M., Ponder, B.A., and Easton, D.F. (2004). Association studies for finding cancer-susceptibility genetic variants. Nature reviews Cancer *4*, 850-860.

Pharoah, P.D., and Ponder, B.A. (2002). The genetics of ovarian cancer. Best practice & research Clinical obstetrics & gynaecology *16*, 449-468.

Pharoah, P.D., Tsai, Y.Y., Ramus, S.J., Phelan, C.M., Goode, E.L., Lawrenson, K., Buckley, M., Fridley, B.L., Tyrer, J.P., Shen, H., *et al.* (2013). GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nature genetics *45*, 362-370, 370e361-362.

Phelan, C.M., Kuchenbaecker, K.B., Tyrer, J.P., Kar, S.P., Lawrenson, K., Winham, S.J., Dennis, J., Pirie, A., Riggan, M.J., Chornokur, G., *et al.* (2017). Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. Nature genetics *49*, 680-691.

Pomerantz, M.M., Ahmadiyeh, N., Jia, L., Herman, P., Verzi, M.P., Doddapaneni, H., Beckwith, C.A., Chan, J.A., Hills, A., Davis, M., *et al.* (2009). The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nature genetics *41*, 882-884.

Press, J.Z., De Luca, A., Boyd, N., Young, S., Troussard, A., Ridge, Y., Kaurah, P., Kalloger, S.E., Blood, K.A., Smith, M., *et al.* (2008). Ovarian carcinomas with genetic and epigenetic BRCA1 loss have distinct molecular abnormalities. BMC cancer *8*, 17.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. Nature *386*, 569-577.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432-445.

Reich, D.E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R., *et al.* (2001). Linkage disequilibrium in the human genome. Nature *411*, 199-204.

Ren, B. (2010). Transcription: Enhancers make non-coding RNA. Nature *465*, 173-174.

Risch, H.A. (1998). Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone. Journal of the National Cancer Institute *90*, 1774-1786.

Risch, H.A., and Howe, G.R. (1995). Pelvic inflammatory disease and the risk of epithelial ovarian cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology *4*, 447-451.

Risch, H.A., Marrett, L.D., and Howe, G.R. (1994). Parity, contraception, infertility, and the risk of epithelial ovarian cancer. American journal of epidemiology *140*, 585-597.

Risch, N.J. (2000). Searching for genetic determinants in the new millennium. Nature *405*, 847-856.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., *et al.* (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nature methods *4*, 651-657.

Rodriguez, G. (2003). New insights regarding pharmacologic approaches for ovarian cancer prevention. Hematology/oncology clinics of North America *17*, 1007-1020, x.

Rodriguez, G.C., Walmer, D.K., Cline, M., Krigman, H., Lessey, B.A., Whitaker, R.S., Dodge, R., and Hughes, C.L. (1998). Effect of progestin on the ovarian epithelium of macaques: cancer prevention through apoptosis? Journal of the Society for Gynecologic Investigation *5*, 271-276.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS letters *579*, 909-915.

Rosa, M.I., Silva, G.D., de Azedo Simoes, P.W., Souza, M.V., Panatto, A.P., Simon, C.S., Madeira, K., and Medeiros, L.R. (2013). The prevalence of human papillomavirus in ovarian cancer: a systematic review. International journal of gynecological cancer : official journal of the International Gynecological Cancer Society *23*, 437-441.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., *et al.* (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature *419*, 832-837.

Sakabe, N.J., Savic, D., and Nobrega, M.A. (2012). Transcriptional enhancers in development and disease. Genome biology *13*, 238.

Salehi, F., Dunfield, L., Phillips, K.P., Krewski, D., and Vanderhyden, B.C. (2008). Risk factors for ovarian cancer: an overview with emphasis on hormonal factors. Journal of toxicology and environmental health Part B, Critical reviews *11*, 301-321.

Sankaranarayanan, R., and Ferlay, J. (2006). Worldwide burden of gynaecological cancer: the size of the problem. Best practice & research Clinical obstetrics & gynaecology *20*, 207-225.

Santos-Rosa, H., and Caldas, C. (2005). Chromatin modifier enzymes, the histone code and cancer. European journal of cancer *41*, 2381-2402.

Sanyal, A., Bau, D., Marti-Renom, M.A., and Dekker, J. (2011). Chromatin globules: a common motif of higher order chromosome structure? Current opinion in cell biology *23*, 325-331.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., *et al.* (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nature genetics *42*, 53-61.

Schreiber, V., Dantzer, F., Ame, J.C., and de Murcia, G. (2006). Poly(ADP-ribose): novel functions for an old molecule. Nature reviews Molecular cell biology *7*, 517-528.

Schulze, S.R., and Wallrath, L.L. (2007). Gene regulation by chromatin structure: paradigms established in Drosophila melanogaster. Annual review of entomology *52*, 171-192.

Seidman, J.D., Horkayne-Szakaly, I., Haiba, M., Boice, C.R., Kurman, R.J., and Ronnett, B.M. (2004). The histologic type and stage distribution of ovarian carcinomas of surface epithelial origin. International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists *23*, 41-44.

SF, A. (2010). SEER Cancer Statistics Review, 1975-2007, National Cancer Institute., K.C. Altekruse SF, Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (eds), ed. (Bethesda, MD).

Shapiro, M.J., and Shapiro, V.S. (2011). Transcriptional repressors, corepressors and chromatin modifying enzymes in T cell development. Cytokine *53*, 271-281.

Shen, H., Fridley, B.L., Song, H., Lawrenson, K., Cunningham, J.M., Ramus, S.J., Cicek, M.S., Tyrer, J., Stram, D., Larson, M.C., *et al.* (2013). Epigenetic analysis leads to identification of HNF1B as a subtype-specific susceptibility gene for ovarian cancer. Nature communications *4*, 1628.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., *et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shih Ie, M., and Kurman, R.J. (2004). Ovarian tumorigenesis: a proposed model based on morphological and molecular genetic analysis. The American journal of pathology *164*, 1511-1518.

Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. (2008). Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. PLoS biology *6*, e65.

Siegel, R., Ma, J., Zou, Z., and Jemal, A. (2014). Cancer statistics, 2014. CA: a cancer journal for clinicians *64*, 9-29.

Simon, J.M., Giresi, P.G., Davis, I.J., and Lieb, J.D. (2012). Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA. Nature protocols *7*, 256-267.

Singer, G., Oldt, R., 3rd, Cohen, Y., Wang, B.G., Sidransky, D., Kurman, R.J., and Shih Ie, M. (2003). Mutations in BRAF and KRAS characterize the development of low-grade ovarian serous carcinoma. Journal of the National Cancer Institute *95*, 484-486.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. Annual review of biochemistry *72*, 449-479.

Smith, E.R., Lin, C., Garrett, A.S., Thornton, J., Mohaghegh, N., Hu, D., Jackson, J., Saraf, A., Swanson, S.K., Seidel, C.*, et al.* (2011). The little elongation complex regulates small nuclear RNA transcription. Molecular cell *44*, 954-965.

Song, H., Ramus, S.J., Tyrer, J., Bolton, K.L., Gentry-Maharaj, A., Wozniak, E., Anton-Culver, H., Chang-Claude, J., Cramer, D.W., DiCioccio, R.*, et al.* (2009). A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nature genetics *41*, 996-1000.

Sotelo, J., Esposito, D., Duhagon, M.A., Banfield, K., Mehalko, J., Liao, H., Stephens, R.M., Harris, T.J., Munroe, D.J., and Wu, X. (2010). Long-range enhancers on 8q24 regulate c-Myc. Proceedings of the National Academy of Sciences of the United States of America *107*, 3001-3005.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nature reviews Genetics *13*, 613-626.

Splinter, E., Grosveld, F., and de Laat, W. (2004). 3C technology: analyzing the spatial organization of genomic loci in vivo. Methods in enzymology *375*, 493-507.

Stadler, Z.K., Thom, P., Robson, M.E., Weitzel, J.N., Kauff, N.D., Hurley, K.E., Devlin, V., Gold, B., Klein, R.J., and Offit, K. (2010). Genome-wide association studies of cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *28*, 4255-4267.

Staples, J., and Goodman, A. (2013). PARP Inhibitors in Ovarian Cancer.

Storici, F. (2014). Gene correction : methods and protocols (New York: Humana Press).

Stratton, J.F., Pharoah, P., Smith, S.K., Easton, D., and Ponder, B.A. (1998). A systematic review and meta-analysis of family history and risk of ovarian cancer. British journal of obstetrics and gynaecology *105*, 493-499.

Struhl, K. (1989). Molecular mechanisms of transcriptional regulation in yeast. Annual review of biochemistry *58*, 1051-1077.

Struhl, K. (1995). Yeast transcriptional regulatory mechanisms. Annual review of genetics *29*, 651-674.

Studies, N.-N.W.G.o.R.i.A., Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D.*, et al.* (2007). Replicating genotype-phenotype associations. Nature *447*, 655-660.

Sur, I., Tuupanen, S., Whitington, T., Aaltonen, L.A., and Taipale, J. (2013). Lessons from functional analysis of genome-wide association studies. Cancer research *73*, 4180-4184.

Svahn, M.F., Faber, M.T., Christensen, J., Norrild, B., and Kjaer, S.K. (2014). Prevalence of human papillomavirus in epithelial ovarian cancer tissue. A meta-analysis of observational studies. Acta obstetricia et gynecologica Scandinavica *93*, 6-19.

Tan-Wong, S.M., French, J.D., Proudfoot, N.J., and Brown, M.A. (2008). Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. Proceedings of the National Academy of Sciences of the United States of America *105*, 5160-5165.

Tangutoori, S., Baldwin, P., and Sridhar, S. (2015). PARP inhibitors: A new era of targeted therapy. Maturitas *81*, 5-9.

Teer, J.K., Yoder, S., Gjyshi, A., Nicosia, S.V., Zhang, C., and Monteiro, A.N.A. (2017). Mutational heterogeneity in non-serous ovarian cancers. Scientific reports *7*, 9728.

Tenesa, A., Farrington, S.M., Prendergast, J.G., Porteous, M.E., Walker, M., Haq, N., Barnetson, R.A., Theodoratou, E., Cetnarskyj, R., Cartwright, N.*, et al.* (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nature genetics *40*, 631-637.

Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A.*, et al.* (2008). Multiple loci identified in a genome-wide association study of prostate cancer. Nature genetics *40*, 310-315.

Thorstenson, Y.R., Roxas, A., Kroiss, R., Jenkins, M.A., Yu, K.M., Bachrich, T., Muhr, D., Wayne, T.L., Chu, G., Davis, R.W.*, et al.* (2003). Contributions of ATM mutations to familial breast and ovarian cancer. Cancer research *63*, 3325-3333.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B.*, et al.* (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75-82.

Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., Penegar, S., Chandler, I., Gorman, M., Wood, W*., et al.* (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nature genetics *39*, 984-988.

Tsai, H.C., and Baylin, S.B. (2011). Cancer epigenetics: linking basic biology to clinical medicine. Cell research *21*, 502-517.

Tseng, Y.Y., Moriarity, B.S., Gong, W., Akiyama, R., Tiwari, A., Kawakami, H., Ronning, P., Reuland, B., Guenther, K., Beadnell, T.C*., et al.* (2014). PVT1 dependence in cancer with MYC copy-number increase. Nature *512*, 82-86.

Urnov, F.D. (2003). Chromatin remodeling as a guide to transcriptional regulatory networks in mammals. Journal of cellular biochemistry *88*, 684-694.

Van Gorp, T., Amant, F., Neven, P., Vergote, I., and Moerman, P. (2004). Endometriosis and the development of malignant tumours of the pelvis. A review of literature. Best practice & research Clinical obstetrics & gynaecology *18*, 349-371.

Vaughan, S., Coward, J.I., Bast, R.C., Jr., Berchuck, A., Berek, J.S., Brenton, J.D., Coukos, G., Crum, C.C., Drapkin, R., Etemadmoghadam, D*., et al.* (2011). Rethinking ovarian cancer: recommendations for improving outcomes. Nature reviews Cancer *11*, 719-725.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A*., et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. American journal of human genetics *101*, 5-22.

Waki, H., Nakamura, M., Yamauchi, T., Wakabayashi, K., Yu, J., Hirose-Yotsuya, L., Take, K., Sun, W., Iwabu, M., Okada-Iwabu, M*., et al.* (2011). Global mapping of cell type-specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation. PLoS genetics *7*, e1002311.

Wallrath, L.L., Nagy, P.L., and Geyer, P.K. (2008). Editorial. Epigenetics of development and human disease. Mutation research *647*, 1-2.

Walsh, T., Casadei, S., Lee, M.K., Pennil, C.C., Nord, A.S., Thornton, A.M., Roeb, W., Agnew, K.J., Stray, S.M., Wickramanayake, A*., et al.* (2011). Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. Proceedings of the National Academy of Sciences of the United States of America *108*, 18032-18037.

Wang, W.Y., Barratt, B.J., Clayton, D.G., and Todd, J.A. (2005). Genome-wide association studies: theoretical and practical concerns. Nature reviews Genetics *6*, 109-118.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nature genetics *40*, 897-903.

Wasserman, N.F., Aneas, I., and Nobrega, M.A. (2010). An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. Genome research *20*, 1191-1197.

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

Wheeler, J.M., Bodmer, W.F., and Mortensen, N.J. (2000). DNA mismatch repair genes and colorectal cancer. Gut *47*, 148-153.

Whittemore, A.S., Harris, R., and Itnyre, J. (1992). Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies. II. Invasive epithelial ovarian cancers in white women. Collaborative Ovarian Cancer Group. American journal of epidemiology *136*, 1184-1203.

Wilson, B.G., and Roberts, C.W. (2011). SWI/SNF nucleosome remodellers and cancer. Nature reviews Cancer *11*, 481-492.

Wisman, G.B., Hollema, H., Helder, M.N., Knol, A.J., Van der Meer, G.T., Krans, M., De Jong, S., De Vries, E.G., and Van der Zee, A.G. (2003). Telomerase in relation to expression of p53, c-Myc and estrogen receptor in ovarian tumours. International journal of oncology *23*, 1451-1459.

Workman, J.L., and Kingston, R.E. (1992). Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. Science *258*, 1780-1784.

Wright, J.B., Brown, S.J., and Cole, M.D. (2010). Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. Molecular and cellular biology *30*, 1411-1420.

Wu, R., Zhai, Y., Fearon, E.R., and Cho, K.R. (2001). Diverse mechanisms of beta-catenin deregulation in ovarian endometrioid adenocarcinomas. Cancer research *61*, 8247-8255.

Xing, D., and Orsulic, S. (2006). A Mouse Model for the Molecular Characterization of Brca1-Associated Ovarian Carcinoma. Cancer research *66*, 8949-8953.

Yan, H., Dobbie, Z., Gruber, S.B., Markowitz, S., Romans, K., Giardiello, F.M., Kinzler, K.W., and Vogelstein, B. (2002). Small changes in expression affect predisposition to tumorigenesis. Nature genetics *30*, 25-26.

Yang, T.H., Kon, M., and DeLisi, C. (2013). Genome-wide association studies. Methods in molecular biology *939*, 233-251.

Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N.*, et al.* (2007). Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nature genetics *39*, 645-649.

Young, R.C., Walton, L.A., Ellenberg, S.S., Homesley, H.D., Wilbanks, G.D., Decker, D.G., Miller, A., Park, R., and Major, F., Jr. (1990). Adjuvant therapy in stage I and stage II epithelial ovarian cancer. Results of two prospective randomized trials. The New England journal of medicine *322*, 1021-1027.

Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556-559.

Zanke, B.W., Greenwood, C.M., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S.M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E.*, et al.* (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nature genetics *39*, 989-994.

Zech, L., Haglund, U., Nilsson, K., and Klein, G. (1976). Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas. International journal of cancer Journal international du cancer *17*, 47-56.

Zhang, M., Wang, Z., Obazee, O., Jia, J., Childs, E.J., Hoskins, J., Figlioli, G., Mocci, E., Collins, I., Chung, C.C.*, et al.* (2016). Three new pancreatic cancer susceptibility signals identified on chromosomes 1q32.1, 5p15.33 and 8q24.21. Oncotarget *7*, 66328-66343.

Zhang, M., Xie, X., Lee, A.H., and Binns, C.W. (2004). Sedentary behaviours and epithelial ovarian cancer risk. Cancer causes & control : CCC *15*, 83-89.

Zhao, J.Y., Yang, X.Y., Gong, X.H., Gu, Z.Y., Duan, W.Y., Wang, J., Ye, Z.Z., Shen, H.B., Shi, K.H., Hou, J.*, et al.* (2012). Functional variant in methionine synthase reductase intron-1 significantly increases the risk of congenital heart disease in the Han Chinese population. Circulation *125*, 482-490.

Ziltener, H.J., Maines-Bandiera, S., Schrader, J.W., and Auersperg, N. (1993). Secretion of bioactive interleukin-1, interleukin-6, and colony-stimulating factors by human ovarian surface epithelium. Biology of reproduction *49*, 635-641.

Zingher, G. (2003). The Corlears Colonial Workshop. School Library Media Activities Monthly *20*, 40.

APPENDIX

During my time in the Ph.D. program, in addition to this dissertation, I worked on numerous other projects that resulted in publications and poster presentations as follows:

Publications

1. **Gjyshi A**, Dash, S, Cen L, Cheng CH, Zhang C, Yoder SJ, Teer J, Armaiz-Pena GN, Monteiro AN (2018) Early transcriptional response of human ovarian and fallopian tube surface epithelial cells to norepinephrine. Scientific Reports 8(1):8291.

As a lead author, I contributed to the above referenced paper by developing and implementing ideas; performing RNA-isolation, gene expression assays and western blot experiments; interpreting results; literature curation and manuscript preparation.

2. Teer JK, Yoder S, **Gjyshi A**, Nicosia SV, Zhang C, Monteiro AN (2017) Mutational heterogeneity in non-serous ovarian cancers. Scientific Reports 7(1):9728.

As a contributing author, I performed literature curation and compiled a comprehensive overview of the current published knowledge on somatic alterations and germline

variation in high grade serous ovarian cancer, which is found in Supplementary Table 1. I also participated in data analyses and manuscript preparation.

3. Buckley MA, Woods NT, Tyrer J, Mendoza-Fandino G, Lawrenson K, Hazelett DJ., **Gjyshi A**, ... (162 authors)… Monteiro ANA (2018) Functional analysis of the 9p22.2 ovarian cancer susceptibility locus reveals a transcription regulatory network mediated by BNC2 in ovarian cells. *In press (Cancer Research)*.

As a contributing author, I performed functional analyses of SNP variants at the 9p22.2 ovarian cancer susceptibility locus, specifically generating tiles, recombinational cloning, luciferase reporter assays as well as library preparation and primer design for chromosome conformation capture (3C) assay. The results of my input are found in Figure 2B, Figure 3H and Supplementary Table 1.

4. Buckley M, **Gjyshi A**, Carvalho RS, Mendoza-Fandiño G, Carvalho MA, Baskin R, Woods NT, Monteiro ANA (2016) An enhancer scanning method to locate regulatory regions in genomic loci. Nature Protocols 11:46-60.

As a co-first author, I made a significant contribution to this paper including conceiving ideas, resource building, designing workflow, primer design, tile generation, cloning, mutagenesis, creating BED files and custom tracks, performing luciferase reporter assays, troubleshooting, generating Figure 4 and Figure 5, and manuscript preparation.

5. Pharoah PD,… (63 authors) … **Gjyshi A**, .. (102 authors).. Sellers TA (2013) GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nature Genetics 45(4):362-70.

My contribution in this paper is in the functional analyses of SNP variants, specifically the functional enhancer mapping at the 8q21 locus, which is found in Figure 2F.

6. Buckley M, **Gjyshi A**, Monteiro ANA (2013) Functional Analysis of Predisposition loci. AACR Annual Meeting 2013 Education Book article. Am Assoc Cancer Res Educ Book 2013: 85-90. doi: 10.1158/AACR.EDB-13-8396.

My contribution to this publication includes developing ideas and workflow, experimental design, and manuscript preparation.

7. **Gjyshi A**, Mendoza-Fandino G, Tyrer J, Woods NT, Lawrenson K, Buckley MA, …(153 authors) … Alvaro N.A. Monteiro (2018) MYC distal enhancers underlie ovarian cancer susceptibility in the 8q24.21 locus. *Submitted to publication (Nature Communications, revised manuscript in preparation)*.

This is my main project during my doctoral training and my contribution is discussed in great detail in this dissertation.

Posters

1. **Gjyshi A**, Mendoza-Fandino G, Tyrer J, Woods NT, Lawrenson K, Buckley MA, Shen HC, Carvalho RS, Seo JH, Phelan C, Freedman ML, Goode EL, Sellers TA, Gayther SA, Pharoah PD, and Monteiro AN on behalf of the Ovarian Cancer Association Consortium (2015) MYC distal enhancers underlie ovarian cancer susceptibility at the 8q24.21 locus. AACR Meeting on Ovarian Cancer, Orlando, FL. Chosen for platform presentation. Clinical Cancer Research 22(2 Supplement):PR15-PR15 · January 2016.

2. Teer JK, Yoder S, **Gjyshi A**, Zhang C, Monteiro AN (2015) Mutational heterogeneity in non-serous ovarian cancers. AACR Meeting on Ovarian Cancer, Orlando, FL. Clinical Cancer Research 22(2 Supplement):B11-B11 · January 2016.

3. Mendoza-Fandino G, Woods N, Baskin R, **Gjyshi A**, Monteiro AN (2015) SNP-FEMS: a method to identify DNA binding proteins interacting with enhancer elements. AACR 106th Annual Meeting, Philadelphia, PA. Cancer Research 75(15 Supplement):841-841 · July 2015

4. Buckley MC, Shen HC, Mendoza-Fandino GA, Woods NT, **Gjyshi A**, French J, Lawrenson K, Song H, Tyrer J, Carvalho RS, Valle A, Chen YA, Yoder S, Bloom G, Tsai YY, Yang A, Hughes TR, Qu X, Cicek M, Larson M, Goode E, Fridley B, Ramus

S, Chenevix-Trench G, Pharoah P, Sellers TA, Gayther, Monteiro AN, Ovarian Cancer Association Consortium (2014) Functional analysis of the 9p22 locus implicates the transcriptional regulation of BNC2 as a mechanism in ovarian cancer predisposition. AACR 105th Annual Meeting, San Diego, CA. Cancer Research 74(19 Supplement):3285-3285 · September 2014

ABOUT THE AUTHOR

Anxhela Gjyshi Gustafson (maiden name Anxhela Gjyshi), born and raised in Korçë, Albania, received a Bachelor of Science degree in Microbiology and Cell Science from the University of Florida in 2004. Afterward, she received a Master of Public Health degree from the University of South Florida in 2008. During graduate school, Anxhela worked as a Biological Scientist in the Florida Department of Health Bureau of Public Health Laboratories in Tampa, Florida. She subsequently pursued her graduate degree in the Cancer Biology Ph.D. program at the University of South Florida and conducted research in the laboratory of Dr. Alvaro Monteiro at the H. Lee Moffitt Cancer Center and Research Institute where she focused on identifying genetic contributions and biological mechanisms associated with ovarian cancer risk. During her doctoral training Anxhela played an active role in the NIH/NCI Genetic Associations and Mechanisms in Oncology (GAME-ON) consortium as part of the Functional Working Group in the emerging field of GWAS functional analysis.