APPLICATION OF WHOLE-EXOME SEQUENCING TO MINE FOR NOVEL GENES ASSOCIATED WITH LYNCH SYNDROME

A Research Project Presented to the Faculty
of
California State University, Stanislaus

In Partial Fulfillment
of the Requirements for the Degree
of Master of Genetic Counseling

By
Jenna S. Guiltinan
July, 2013

CERTIFICATION OF APPROVAL

APPLICATION OF WHOLE-EXOME SEQUENCING TO MINE FOR NOVEL GENES ASSOCIATED WITH LYNCH SYNDROME

by
Jenna S. Guiltinan

Signed Certification of Approval Page is on file with the University Library

Janey Youngblom, Ph.D. Professor of Genetics, California State University, Stanislaus	July 22nd, 2013
Amie Blanco, MS, LCGC Gordon and Betty Moore Endowed Counselor of Hereditary GI Cancer Prevention University of California, San Francisco	July 22nd, 2013
Eliana Hechter, Ph.D. Medical and Population Genetics Platform, Broad Institute, Cambridge, MA	July 22nd, 2013

ACKNOWLEDGEMENTS

First, I would like to thank the study participants and their families. It was a pleasure to speak with so many brave cancer survivors. Secondly, this project would not have been possible without the generous contributions from Ambry Genetics, Illumina, Inc., and SoftGenetics. The genetic counselors at the UCSF Cancer Risk Program also deserve many thanks for allowing me to use their database, their office space, and for entrusting me with their patients. Finally, I would like to thank and acknowledge my three research advisors for their role in this project.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

ABSTRACT

Lynch syndrome is a highly heterogeneous cancer predisposition syndrome characterized by an increased risk for colorectal and endometrial tumors exhibiting microsatellite instability due to a deficiency in the DNA mismatch repair (MMR) system. Most individuals with Lynch syndrome have germline mutations in one of the four main MMR genes, however there is a group of Lynch syndrome patients who display all of the hallmarks of the disease, but lack identifiable MMR gene mutations. We used exome sequencing as a tool to mine for novel genetic causes of Lynch syndrome in a cohort of 32 individuals, 31 of which clinical Lynch syndrome and are MMR mutation negative by Sanger sequencing. The method of variant filtering used for this study excluded variants in all genes except for 809 genes previously identified to have a role in carcinogenesis, and only variants that are predicted damaging by *in silico* models were included. Exome sequencing did not identify new mutations in the four main MMR genes associated with Lynch syndrome in this cohort of 32 individuals, confirming the original Sanger sequencing results. Variants in seven known cancer susceptibility genes were identified, as well as variants in 103 other genes. This study did not produce any definitive link between a novel gene and LS, however the *FOXO3, JAG2,* and *MTUS1* genes were identified as possible candidates based on their molecular role in cancer. Future research will focus on alternative methods of variant filtering in this same data set.

INTRODUCTION

Exome sequencing is a next-generation sequencing technique used to sequence the 1% of the genome that is protein coding; the exons of genes (Choi et al., 2009). Exome sequencing has successfully been used to discover the underlying genetic causes of Mendelian diseases using small cohorts of patients (Ng et al., 2010), and is now utilized more frequently in a clinical setting to search for deleterious mutations in patients when traditional genetic tests are uninformative or clinicians can't pinpoint a specific genetic syndrome (Jamal et al., 2013). Exome sequencing is particularly promising as a diagnostic tool in diseases that are characterized by genetic or phenotypic heterogeneity, especially when all known genes can only explain a portion of the cases (Ku et al., 2012). Lynch syndrome (LS) is a hereditary cancer predisposition syndrome in which patients may benefit from clinical exome sequencing because it can be caused by mutations in any of five known genes, a good example of genetic heterogeneity (Schneider, 2012). Also, there are an additional five genes known to be involved with the underlying mechanism causing LS, but how mutations in these genes affect disease is largely unknown. Furthermore, a portion of cases of LS cannot be molecularly confirmed with Sanger sequencing, the currently available clinical genetic testing method, suggesting that either Sanger sequencing is inadequate or there are other causative genes yet to be discovered (Bansidhar et al., 2012). This study utilized exome sequencing in a cohort of cancer patients who have the clinical picture of LS but for whom all attempts to identify a causative mutation

by Sanger sequencing had failed, leaving the patients without a clear molecular diagnosis. The purpose is two-fold; to provide our research subjects with an answer to their long quest for a molecularly confirmed diagnosis, while assessing the efficacy of exome sequencing to both mine for novel genes and to detect changes in genes that traditional Sanger sequencing may have missed.

Lynch Syndrome Overview

Lynch syndrome is a hereditary cancer predisposition syndrome caused by a deficiency in mismatch repair (MMR) proteins and associated with tumors exhibiting microsatellite instability (MSI) (Vilar et al., 2010). Individuals with LS have up to an 80% lifetime risk of colorectal carcinomas, as well as an increased risk for a variety of other cancers. Women with LS also have an increased risk of both endometrial and ovarian cancers (Robinson et al., 2007). See **Table 1** for the lifetime risks of all cancers associated with LS compared to the general population according to the National Comprehensive Cancer Network (NCCN).

Table 1

Lifetime Cancer Risk in Individuals with Lynch Syndrome Compared to the General Population*

Cancer	General Population Risk	Lynch Syndrome Risk
Colon	5.5%	Up to 80%
Endometrium	2.7%	Up to 60%
Stomach	<1%	Up to 13%
Ovary	1.60%	Up to 24%
Hepatobiliary tract	<1%	Up to 4%
Urinary tract	<1%	Up to 4%
Small bowel	<1%	Up to 6%
Brain/CNS	<1%	Up to 3%
Sebaceous neoplasms	<1%	Up to 9%
Pancreas	<1%	Up to 6%

Note. Risk estimates are from the NCCN Guidelines Version 1.2013 (Burt et al., 2010)
*Up to age 70.

LS is inherited in an autosomal dominant pattern; each child of a person with LS is at a 50% risk of inheriting the condition. LS is the most common hereditary colon cancer syndrome, occurring in approximately 1 in 660 to 1 in 2,000 individuals (Chapelle et al., 2005). It is difficult to estimate a more precise incidence of LS because screening methods designed to differentiate cancer patients with LS from cancer patients with sporadic cancers are not universally implemented, nor are they 100% accurate. Every year an unknown portion of individuals with LS go undiagnosed (Hendricks et al., 2006).

The mismatch repair (MMR) proteins related to Lynch syndrome are responsible for identifying and correcting errors made during DNA replication, such as deletions, insertions, and mismatched base pairs. These MMR proteins function as

tumor suppressors by protecting the integrity of the genome. In the "two-hit" model of cancer formation, tumor suppressor genes only require the presence of a single functional allele in order to effectively inhibit the development of cancer (Sherr, 2004), and cancer only occurs if the both alleles of a tumor suppresser gene are non-functional.

It is traditionally understood that individuals who have a cancer predisposition syndrome such as LS inherit one non-working tumor suppresser gene and develop cancer when they sustain a "second-hit", i.e. a randomly acquired somatic mutation in the other allele of the gene. In contrast, sporadic cancer formation requires two somatic mutational events affecting both alleles of the same tumor suppressor gene in the same cell (Berger et al., 2007).

The main MMR proteins causing LS when deficient are MutL homolog1 (MLH1), MutS protein homolog 2 (MSH2), MutS homolog 6 (MSH6), and Postmeiotic segregation increased 2 (PMS2)**.** The proteins function as heterodimeric complexes; MLH1 complexes with PMS2, while MSH2 complexes with MSH6. PMS2 is unstable without MLH1, therefore loss of MLH1 is almost always is accompanied by loss of PMS2. However, a defective PMS2 protein leads to loss of only PMS2. This same principle applies to the MSH2 and MSH6 proteins; MSH6 is unstable without MSH2 (Yang, 2000). Approximately 90% of LS is caused by germline mutations in *MLH1* or *MSH2*, but mutations in *MSH6* and *PMS2* can also occur. Other proteins involved in the mismatch repair system are MSH3, MSH4,

MSH5, MLH3, and PMS2, but their role is not currently well understood (Bansidhar et al., 2012).

Methods of Identifying Lynch syndrome

Colorectal cancer is the third most common cancer in the United States. In 2012 alone, well over 100,000 new cases of colorectal cancer were diagnosed and Lynch syndrome causes approximately 2-4% of these. In addition to LS, there are several other colorectal cancer predisposition syndromes (Burt et al., 2010), making the diagnostic workup to ascertain who is at risk for LS elaborate and often time consuming, requiring the collaboration of clinicians from many different medical specialties.

An LS workup ideally begins with an inquiry about the family history of cancer. Those with a family history suggestive of LS should be referred to a genetics clinic so that a full and detailed pedigree can be assessed by a genetic counselor. There are two well-established guidelines that outline the criteria recommended for a *clinical* definition of LS: the Amsterdam Criteria I and the Amsterdam Criteria II. Approximately half of individuals with molecularly confirmed LS do not meet the Amsterdam I or II criteria. The Revised Bethesda Guidelines are less stringent and designed to aid in the decision of which patients should have tumor testing, regardless if they meet Amsterdam Criteria. The NCCN recommends that cancer patients meeting either set of guidelines be further evaluated for their risk of LS by screening tumors with immunohistochemistry (IHC) or microsatellite instability (MSI) tests,

and sometimes both (Burt et al., 2010). Increasingly, medical centers are implementing universal IHC and/or MSI testing on all colorectal and endometrial cancers in an effort to identify more patients with Lynch syndrome (Zhang, 2013).

Tumor Analysis: Immunohistochemistry (IHC) and Microsatellite Instability (MSI)

An immunohistochemistry (IHC) test is a relatively inexpensive laboratory test performed on colorectal or endometrial tumor tissue, (though it can be used on other tumors as well), that uses specific antibodies to detect the presence or absence of the four main MMR proteins, (MLH1, MSH2, MSH6, and PMS2). IHC results that demonstrate a loss of one or more MMR proteins suggest that there is a defective MMR gene in the tumor genome (Schneider, 2012). LS is not the only cause for a loss of an MMR protein, but by definition LS tumors should always show a loss of one or more MMR proteins on IHC. IHC also has the convenient advantage of predicting with germline MMR gene harbors a mutation, helping to guide genetic testing (Hendriks et. al, 2006).

Any defects in the mismatch repair system, regardless of etiology, are expected to lead to microsatellite instability (MSI), another characteristic of most LS tumors (De la Chapelle et al., 2010). Microsatellites are repetitive and highly polymorphic sequences in non-coding regions of the genome. These long strands of repeating sequences are especially vulnerable to errors during DNA replication. Functioning MMR proteins maintain genomic stability by correcting errors. One

outcome of this system is that as cells divide, the microsatellites lengths do not differ in successive daughter cells. However, with deficient MMR proteins, the errors are not successfully corrected, and in the rapidly duplicating tumor cells, the lengths of the microsatellites can be widely variable between cells and so are regarded as microsatellite unstable (MSI-High). This microsatellite instability, caused by MMR protein deficits, is a hallmark of LS tumors (Vilar et al, 2010). An estimated 85-92% of LS tumors are MSI-High (Hendriks et al., 2006).

Causes of MMR Protein Deficiency

While the sensitivity of MSI testing for Lynch syndrome is high, approximately 10-15% of all colorectal tumors that occur in the general population are also MSI-High (Hendriks et al., 2006). About 80% of these MSI-High tumors show loss of MLH1 on IHC tests, while the remaining 20% display loss of MSH2 and/or MSH6, or of PMS2 alone. The majority of MSI-High, MLH1 deficient tumors are caused by a randomly acquired mutation in one copy of an individuals *BRAF* proto-oncogene, a common cause of sporadic cancer (Davies et al., 2002). MLH1 can also be lost on IHC due to aberrant methylation of the *MLH1* promoter and subsequent silencing of the *MLH1* gene (Jensen et al., 2009). If a *BRAF* mutation or *MLH1*-promoter methylation is found in tumor cells, a diagnosis of LS can essentially be ruled out in favor of a sporadic etiology.

Tumors with loss of MSH2 were originally assumed to correlate only with mutations in the *MSH2* gene, however now it is known that this loss can also be

caused by germline mutations in the *EpCAM* gene in rare instances. The Epithelial Cell Adhesion Molecule (*EpCAM*) gene is located directly upstream from the *MSH2* gene and 3' deletions in *EpCAM* result in promoter methylation and inactivation of the *MSH2* gene (Kuiper et al., 2011). If an individual had a tumor with loss of MSH2 and an apparently normal *MSH2* gene, *EpCAM* deletion/duplication testing can sometimes provide the answer. However, the percent of LS caused by *EpCAM* deletions is currently unknown (Vilar et al., 2010). At the present time, *EpCAM* testing is not standard of care (Burt et al., 2010).

Another seemingly rare cause of LS is germline methylation of the *MLH1* promoter causing silencing of the *MLH1* gene (Hitchins et al., 2009). The incidence for germline *MLH1* promoter methylation is unknown, as is the underlying cause for this aberrant methylation. The inheritance pattern is still being debated, but some researches have reported only maternal inheritance with reduced penetrance (Hitchins et al., 2009). In 2011, a group from the University of North South Wales, Australia, demonstrated dominantly inherited *MLH1* germline methylation in a family presumably caused by a variant in the untranslated region of the *MLH1* gene (Hitchins et al., 2011). It is not know however, if this variant exists in others with germline *MLH1* methylation.

In the absence of germline MMR gene mutations, other than promoter methylation of *MLH1* or *MSH2*, no other mechanism of inactivation is known for the MMR genes (Rodriguez-Soler, et al., 2013). Yet, there is no question of the existence

of a group of patients who have the clinical picture of Lynch syndrome who do not have pathogenic mutations in MMR genes.

One explanation is that these patients do indeed have a mutation in one of the MMR genes that was not detected by the current method of Sanger sequencing. One of the reasons we chose to use exome sequencing in this study was to see if next-generation sequencing technology would uncover changes in the known MMR genes that Sanger sequencing missed. While this is a possibility, there is evidence of a distinct clinical difference between sporadic cancer patients, molecularly confirmed LS patients, and LS mutation negative patients. One study showed that relatives of LS patients without germline mutations are less likely to suffer from cancer than relatives of LS patients with germline MMR mutations, but more likely to suffer from cancer than relatives of patients with sporadic cancer. In addition, the age of onset of relatives of LS patients without MMR mutations that do develop cancer is on average younger than relatives of sporadic cancer patients, but older than relatives of patients with LS and MMR mutations (Boland, 2013). This difference suggests that there is indeed some other unknown mechanism causing MMR protein deficiency, not simply an unidentified MMR mutation.

The Role of Exome Sequencing in Lynch Syndrome

All of the intricacies of Lynch syndrome make it a very complex disease to study. Perhaps this is why there remains a portion of patients who have the clinical characteristics of LS, but the underlying mechanism still eludes us. These are the

patients that have a personal history of a Lynch-related tumor displaying the hallmarks of MSI and IHC in which a sporadic cause has been ruled out, often have a striking family history suggestive of an autosomal dominant cancer syndrome, and yet every test in the algorithm of LS genetic testing is still negative. Many researchers are familiar with the existence of this group of patients and acknowledge that there must be an underlying genetic mechanism that is presently unknown (Rodriguez-Soler, et al., 2013). Exome sequencing, a powerful tool of examining nearly all protein coding regions in the genome, was chosen for this research project as the method to explore this possibility.

In addition to having the potential to aid in the discovery of novel causes of LS, exome sequencing has an advantage over traditional diagnostic testing approaches to LS because not only can all known LS genes can be tested at the same time, (instead of in a step-wise fashion), rare genes such as *EpCAM* that are associated with LS can also be tested simultaneously. Whole genome sequencing, while much more comprehensive, is more expensive and creates so much sequence data that it complicates the process of identifying novel candidate genes. For these reasons, exome sequencing if often preferred to whole genome sequencing. To date, whole genome sequencing is only used for research purposes. In contrast, many labs now offer CLIA-approved clinical exome sequencing for diagnostic purposes. If the cost of exome sequencing decreases as it is expected to (Zhuang et al., 2012), it is feasible that for LS, as well as for other highly heterogenous diseases, exome sequencing could become routine clinical care.

With this research study we wanted to use exome sequencing to provide our research subjects with an answer to their long quest for a molecular diagnosis of LS, while also assessing the efficacy of exome sequencing to mine for novel genes in highly heterogeneous conditions. We also wanted to see if this next-generation sequencing technique would uncover mutations in the known MMR genes that the traditional Sanger sequencing missed.

METHODS AND MATERIALS

Illumia Inc., (Hayward, CA) and Ambry Genetics, (Aliso Viejo, CA) agreed to provide support to this research project for a total sample size of 32 human subjects. The UCSF collaborators and I were responsible for selection of the 32 subjects, as well as for the data analysis.

Selection of Research Subjects

Over several years, genetic counselors at the University of California, San Francisco Cancer Risk Program maintained a database of patients who had the clinical picture of Lynch syndrome, but had uninformative genetic test results, (negative results or a variant of unknown significance). At the start of this research project, the database consisted of 146 patients. All had provided prior consent for further contact for research purposes through a long-standing UCSF IRB approved research study, (IRB number 10-02541), and all were over the age of 18. It is from this database that the final 32 subjects originated.

This original pool of 146 was first narrowed down to 68. I excluded patients with a variant of unknown significance (VUS) in the MMR gene corresponding to the MMR protein lost on IHC because a VUS can be reclassified as disease causing, and may actually be the cause of their LS. In this case, exome sequencing would not provide additional information. I also excluded patients who did not have comprehensive genetic testing.

I collected the health records on the remaining 68 prospective subjects from the electronic medical records at the UCSF Cancer Risk Program and as need be, they were supplemented with records from the patients' original paper charts filed and stored at UCSF. I created a database of the prospective subjects that included basic demographic information, personal cancer history, results of any tumor testing, results of any genetic testing, and any family history of cancer. The family pedigrees of each subject were also obtained from the records at UCSF, and any identifying comments about subjects' family members were censured to protect their privacy before they were added into the database. This database of information on the prospective subjects was managed using the REDCap electronic data capture tools hosted at UCSF. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies. The use of REDCap ensured that all protected heath information was secure and password protected throughout the study.

To recruit the strongest candidates for exome sequencing, (i.e. candidates most likely to harbor a novel mutation for LS), I created a tiered system of recruitment. All "Tier 1" prospective subjects were contacted by telephone first. Only after all "Tier 1" candidates were either enrolled in the study or declined to participate in the study were "Tier 2" patients contacted, next "Tier 3" was contacted, and then finally the last tier was contacted, "Tier 4".

See **Table 2** for the inclusion criteria of each "Tier".

Table 2

Tiered System for Selection of Research Subjects

	Criteria		
	Tumor Results	Gene Testing Results	
Tier 1*	Loss of MSH2 and/or MSH6	*MSH2* negative	*EPCAM* deletion negative
Tier 2	Loss of MLH and/or PMS2	*MLH1* negative	*BRAF* negative, methylation studies negative

	Criteria
Tier 3a.	Individuals who would fit Tier 1, but did not have *EPCAM* testing
Tier 3b.	Individuals who would fit Tier 2 criteria, but did not have *BRAF* or methylation studies
Tier 4	Any individual who did not fit in Tier 1-3, but has a unique family or personal history that is indicitive of a autosomal dominant colon cancer predisposition syndrome

**These individuals are most likely to harbor a novel LS gene, as every available clinical testing avenue for LS had been exhausted and tumors with loss of MSH2 are less likely than tumors with loss of MLH1 to be sporadic (Herfarth, et al., 1997).*

Consenting Subjects

I applied for and was granted IRB approval for this research project, (Protocol #1213-09), by the University of California, Stanislaus. I contacted potential subjects using the telephone number listed in the patients' medical record, because all had already consented to be contacted for research purposes through the UCSF IRB. Records of the time and date of all calls were recorded in the subject's REDCap file. I left a voicemail if a subject was not home, and to protect the subject's privacy, after three unreturned voicemail messages the subject was excluded. When I reached a prospective subject by telephone, I identified myself as a genetic counseling student conducting research in association with the UCSF Cancer Risk Program.

All subjects were informed that it was completely voluntary to participate in this research. I explained the specifics of the study to any prospective subject who was interested in hearing more. I explained exome sequencing and the kind of information they may receive from participating in the research. Those who were interested in participating were provided the consent form approved by the UC Stanislaus IRB, either through the mail or email. All final research subjects signed and returned a consent form that was uploaded and stored on their RedCAP file for documentation. All subjects had my personal contact information and were encouraged to contact me if they had any questions or concerns. Subjects were assured that only the primary researchers would have access to their personal health information and their exome sequence data would not be connected to their name, except through codes that linked back to the secure REDCap database.

Subjects who did agree to participate in this study were asked at the time of consent if they would be willing to participate in a second research study being conducted by Kelley Hitch, another genetic counseling student at CSU Stanislaus. Her research involved a phone interview lasting approximately 15 minutes. Subjects were advised that participation in Ms. Hitch's study was also voluntary, and declining it would not affect their ability to participate in the exome study. Subjects could indicate their choice on the consent form. If they consented to participate in Ms. Hitch's study as well, we set a time and date that would be convenient for them, and I passed this information on to Ms. Hitch.

As indicated on the consent form, all subjects were aware that we would disclose to them any information from their personal exome results that were related to their cancer diagnosis or future cancer risks. They understood that any other incidental findings would not be disclosed to them. The subjects were also informed that any results would need to be confirmed by a CLIA approved laboratory before that information could be used in their medical care.

The final group of research subjects consisted of 28 unrelated cancer patients, and their demographic information, tumor test results, genetic test results, and cancer history can be found in **Table 3.** We also included four related individuals, three of which were affected with LS. Their relationship and other information can be found in **Table 4.** The three related individuals who were affected had genetic testing and were found to carry a VUS in the *MSH2* gene, the gene corresponding to their IHC

results. They were included despite this because we felt that having a family in the study would prove to be beneficial.

Table 3

Demographics, Tumor and Genetic Test Results, and Cancer History of Unrelated Research Subjects

Demographic				Tumor Test Results			Gene Test Results***						Cancer History
ID#	Tier	Sex	Race*	MSI	IHC Loss	Methylation Studies	*MLH1*	*MSH2*	*MSH6*	*PMS2*	*BRAF*	*EpCAM*	Location, Age of Onset
3	1	F	C	H	MSH2		-	-				-	colon, 27
5	3b.	F	A	H	PMS2		-	-					bone, 46; endometrial, 50
6	3a.	M	C	H	MSH2, MSH6		-	-	-				colon, 50
8	3b.	F	C	H	MLH1		-	-					rectum, 58
11	3b.	F	AA	H	MLH1, PMS2		-	-	-				colon, 63; kidney clear cell, 65; vulvar BCC, 66
14	3a.	M	C		MSH2, MSH6			-	-				colon, 48
15	3a.	F	C		MSH2		-	-					breast, 28; melanoma 29
22	3a.	M	C		MSH2		-	-					colon, 31; melanoma, 60, 67, 73; sebaceous adenoma, 77
23	1	M	C		MSH2, MSH6		-	-				-	rectum, 43
24	4	F	C	H	Retained		-	-					colon, 25; rectum, 40
29	1	M	C	H	MSH2, MSH6			-	-			-	Wilm's tumor (kidney) 3; colon, 52
31	4	F	A	H	MLH1, PMS2	+(germline)	-	-	-		-		rectum, 37
44	2	M	C	H	MLH1, PMS2	-	-	-			-		colon, 44
45	1	F	C		MSH2, MLH6		-	-	-			-	endometrial, 56
47	1	F	C		MSH2, MLH6		-	-	-			-	colon, 54
48	2	F	C	H	MLH1, PMS2	-	-			VUS	-		colon, 44
50	3b.	M	C	H	PMS2		-			-			colon, 49
51	4	F	C	H	MLH1, PMS2	+(germline)	-	-					endometrial, 42; colon 43
52	1	M	C	H	MSH6			-	-			-	colon, 41
53	1	F	C	H	MSH2, MSH6		-	-	-	-		-	endometrial, 62
54	4	F	C	H	Retained		-	-	-	-		-	colon, 62
55	1	F	C	H	MSH2, MSH6			-	-			-	endometrial, 57
56	1	F	C	H	MSH2, MSH6			-	-			-	endometrial, 46
57	2	M	A		MLH1, PMS2		-	-	VUS				stomach, 84
58	2	M	C	H	MLH1, PMS2	-	-				-		colon, 38
59	4	F	H	H	All four		-	-	-	VUS			endometrial, 53
67	3a.	F	C	H	MSH2, MSH6		VUS	-	-				endometrial, 63
68	1	M	C		MSH2		-	-	-			-	sebaceous carcinoma, 55

Note . Empty cell denotes unknown or not tested

*Abbreviations: C, Caucasion; A, Asian; AA, African American; H, Hispanic

**Gene tests results include gene sequencing and duplication/deletion analysis

Table 4

Demographics, Tumor and Genetic Test Results, and Cancer History of Family 0

	Demographics				Tumor Test Results		Gene Test Results**					Cancer History
Relationship	ID#	Tier	Sex	Race*	MSI	IHC Loss	*MLH1*	*MSH2*	*MSH6*	*PMS2*	*EpCAM*	Location, Age of Onset
Son	38	1	M	C	H	MSH2, MSH6	-	VUS	-		-	colon, 29
Mother	62	1	F	C	H	MSH2, MSH6	-	VUS	-		-	colon, 38, 39; ureter, 71
Daughter	63	1	F	C	H	MSH2, MSH6	-	VUS	-		-	ovary, 41
Father	64	N/A	M	C								unaffected

Note . Empty cell denotes unknown or not tested

*Abbreviations: C, Caucasion

**Gene tests results include gene sequencing and duplication/deletion analysis

Obtaining DNA Samples for Exome Sequencing

A majority of the subjects had a peripheral blood sample stored for research purposes at the UCSF Tissue Core prior to enrolment in this study. Those that did were given a choice to use a portion of their stored sample for the exome sequencing or to have a new sample of blood drawn. All agreed to use their stored sample. An Ambry Genetics blood collection box was mailed to the remaining four subjects who did not have a stored blood sample. This kit included two tubes for blood collection and a prepaid return envelope addressed to the UCSF Cancer Risk Program. The subjects could use any laboratory for the blood draw, and if they incurred a fee they were encouraged to return the receipt with the package for reimbursement. When a blood kit was received at UCSF it was sent to the UCSF Tissue Core for storage.

Once all 32 subjects' blood samples had been collected, they were sent to UCSF Genomics Core for DNA extraction from whole blood. Genomic DNA was extracted using the standard procedures, normalized to 200ng/uL in 200uL solution,

for a total of 4 micrograms of genomic DNA per sample. The remaining DNA from each subject was returned to the UCSF Tissue Core for possible use in future research. The normalized samples were mailed to Ambry Genetics, along with de-identified family pedigrees. Ambry Genetics did not have access to any of the subjects protected health information.

Exome Sequencing

Illumina Inc. provided all of the necessary reagents for the exome sequencing, and Ambry Genetics completed all laboratory processes. The 32 genomic DNA samples were prepped using Illumina's TruSeq DNA Sample Preparation Kits. Targeted exome sequence capture was completed with TruSeq Exome Enrichment Kits, covering over 20,000 genes, (over 200,000 exons), as well as broad coverage of non-coding DNA in exon-flanking promoters and untranslated regions. Illumina provided a TruSeq Cluster Kit with Flow Cell for each sample that allows for improved coverage. The enriched libraries were sequenced using 100bp paired-end, 200 cycle chemistry on the Illumina HiSeq 2000. When this was completed, Ambry Genetics provided me with the raw sequencing data.

RESULTS

The raw exome sequence data from each subject was assembled using the NextGENe Software, (SoftGenetics, LLC, State College, PA), and aligned to the human genome reference sequence hg19 (GRCh37). Variant annotation was also performed using NextGENe Software, based on the dbSNP 135 build.

Preliminary variant filtering was also completed using the NextGENe software based on my preferences. Of the tens of thousands of variants each subject had, it was essential to narrow this amount down to a manageable number. My goal for variant filtering in this study was essentially to identify rare variants in protein coding regions of the genome that were predicted to cause damage at the protein level and could potentially lead to a cancer predisposition.

To focus on variants that would affect the protein and were rare, I first excluded any noncoding and silent variants, as well as any variant with a population frequency greater than 1.0% based on 1000 Genomes Project data. The next variant filtering step aimed to exclude variants that were not predicted to be damaging to the protein. In order to pass this filtering step, a variant had to be predicted damaging or possibly damaging by at least one of the following *in silico* prediction models: Mutation Assessor, PolyPhen-2, SIFT, Fathmm, Mutation Tester, and LTR. The list of remaining variants were then restricted to mutations that occurred only in genes that have been to be causally implicated in oncogenesis by A Census of Human Cancer Genes, Futreal et al., 2004 (**Appendix 1**). This included 809 genes implicated

in cancer by either somatic mutations or germline mutations. Approximately 10% of the genes are implicated in cancer by both somatic and germline mutations (Futreal, et al., 2004).

At this point in my research, each subject had between 21-66 variants. I could no longer use the NextGENe software because the program requires an operating system far beyond what I have access to, and my physical location no longer allowed me to use the computer at the SoftGenetics headquarters. The files were converted to a .VCF (variant calling format) file so that I could access them on my personal computer.

With an average of 35 variants per subject, the data was still too robust for me to analyze without access to software. To further narrow down the list of variants I utilized The Exomiser, a program available through the Welcome Trust Sanger Institute website, (http://www.sanger.ac.uk/resources/databases/exomiser/). The Exomiser annotates and prioritizes variants from exome sequence .VCF files based on user-defined criteria. I set the criteria to remove non-pathogenic variants. The Exomiser relies on predicted pathogenicity data from dbNSFP, a database for nonsynonymous variant functional predictions that compiles prediction scores from four prediction models, (SIFT, Polyphen-2, LRT, and MutationTaster), along with a conservation score. In order to pass my original functional prediction filter, the variant only had to be predicted damaging or possibly damaging on one six models, which included the four used by dbNSFP. Because this dbNSFP score was more stringent that my original filtering, the Exomiser narrowed down the number of

variants to between 0 and 13 for each subject. I drew my final candidate genes from these remaining variants.

The full list of variants identified post-filtering for each subject can be found in **Appendix B,** sorted by Study ID#, and in **Appendix C**, sorted by gene name**.** The lists contain the gene name, genomic location of the variant, the codon with nucleotide change, protein change, rs# if variant is listed in dbSNP, and population frequency (if known).

Variations in seven known cancer susceptibility genes were identified representing the following cancer predisposition syndromes: Lynch Syndrome (*PMS1, MSH3*), Tumor Predisposition syndrome (*BAP1*), Hereditary Breast and Ovarian Cancer syndrome (*BRCA2*), Birt-Hogg-Dube syndrome (*FLCN*), Familial Adenomatous Polyposis Coli syndrome (*APC),* and MutYH-Associated Polyposis syndrome (*MutYH*). The extent to which these variations contributed to the subject's clinical history is unknown, however all of these cancer syndromes are associated with an increased risk of colorectal cancer. None of the variants however, are known deleterious mutations.

Variations in an additional 103 different genes were also identified. While all genes have been casually linked to cancer by A Census of Human Cancer Genes, Futreal et al., 2004 (**Appendix 1**), few had information available regarding the effect of germline mutations on cancer susceptibility. I used any available literature on each gene, as well as information from the public databases such as dbSNP, ClinVar, OMIM etc., to determine the function of the gene in cancer. I took many factors into

consideration in deciding which genes to discuss further, including proportion of subjects with a variation in the gene, gene functions (prioritizing tumor suppressor genes), and evidence of reduced expression and/or somatic mutations in LS-related tumors. *FOXO3, JAG2,* and *MTUS1* emerged as candidate genes, however no definitive link between mutations in these genes and Lynch syndrome can be made.

For the remainder of the text, the participants will be referred to by their Subject ID# that can be found in **Table 3** and **Table 4**.

DISCUSSION

While mining through the exome data, it became apparent that we did not discover pathogenic mutations in any of the known MMR genes that the original Sanger sequencing methods missed. This supports the idea that there is a group of cancer patients who have the clinical picture of Lynch syndrome who do not have pathogenic mutations in MMR genes (Rodriguez-Soler, et al., 2013). The subjects in this research study were carefully chosen as the most likely LS mutation negative patients to have MMR protein deficiency caused by an unknown mechanism, and exome sequencing was selected as the tool to aid us in uncovering a novel gene causing this mechanism.

Three genes with roles in sporadic cancer emerged as possible candidate genes for novel causes of LS; *FOXO3, JAG2,* and *MTUS1.* All of these genes have been recognized as genes that act as tumor suppressors in healthy cells, and have been documented to display abnormal expression levels in some tumors. However, the effect of germline mutations in these genes on cancer susceptibility is unknown.

The possibility that these genes could cause LS relies on two principles of tumor suppressor genes; that some tumor suppressor genes can cause both hereditary cancer and sporadic cancer (Balakrishnan, et al., 2007), and that for some tumor suppressor genes, haploinsufficiency is enough to cause genomic instability that facilitates additional genetic alterations (Salmena & Narod, 2012).

The tumor suppressor genes *TP53* and *PTEN* are examples of the first principle; the same somatic mutations that cause the development of sporadic tumors,

when found in the germline, cause the cancer predisposition syndromes Li-Fraumeni syndrome and Cowden syndrome respectively (Balakrishnan, et al., 2007).

To address the second principle, its important to note that while the two-hit hypothesis is a useful framework for understanding the genetic causes of cancer and some genes do function as classic "two-hit" tumor suppressor genes, we now know that oncogenesis doesn't always follow such a predictable pattern (Berger, et al., 2011). For instance, in individuals heterozygous for a deleterious mutation in the *BRCA1* tumor suppressor, even normal cells show diminished DNA repair capability and increased DNA alterations which suggest that *BRCA1* heterozygosity alone causes genomic instability that increases the mutation rate of other important genes (Salmena & Narod, 2012). This genomic instability then promotes additional genetic changes in these cells early on in oncogenesis, preceding somatic loss of the second functioning allele, i.e. before the "second-hit" (Konishi et al., 2011).

I can't tie any of these genes directly to the MMR system involved in LS, but if heterozygous germline mutations of at least one tumor suppressor (*BRCA1*) is enough to cause genetic alterations in normal cells, perhaps a heterozygote carriers of *FOXO3, JAG2,* and *MTUS1* can cause changes to the expression MMR genes, spurring tumorgenesis. In the following paragraphs I will discuss recent research on FOXO3, JAG2, and MTUS1 that shows how mutations in these genes could possibly lead to a cancer predisposition syndrome. Then I will discuss the known cancer susceptibility genes in the context of the subjects' clinical picture.

Novel Cancer Susceptibility Gene Candidates

FOXO3 (Forkhead box O 3)

Forkhead box O (*FOXO*) transcription factors are regulators of multiple cellular activities, some which are critical aspects of tumorgenesis including cell proliferation, cell cycle arrest, cell differentiation, and cell death (Lou et al., 2013). Seven of the total 32 subjects had one mutation in the *FOXO3* gene (Subjects 24, 47, 51, 52, 54, and 62), while Subject 47 had two mutations in *FOXO3*. All had either the 785G>A mutation or the 527A>G mutation, and Subject 47 had both. Both mutations are predicted to be damaging by dbNSFP. The 785G>A mutation is listed in dbSNP, but without population frequency data or information on clinical significance, while the 527A>G mutation appears to be novel.

There is an abundance of evidence showing that the *FOXO3* gene does play a role in cancer and DNA repair mechanisms. The COMSIC database lists 41 different mutations in FOXO3 that have been found in cancer genomes. Research suggests that disruption of *FOXO* proteins lead to uncontrolled cell proliferation and accumulation of DNA damage and the *FOXO* proteins have been found to help protect cells from accumulating DNA damage by inducing DNA repair mechanisms (Arden, 2006).

The human *FOXO* family members include four related factors; *FOXO1, FOXO3, FOXO4*, and *FOXO6* (Regan-Shaw et al., 2006). A recent paper showed that *FOXO3* protein deficiency results in epithelial proliferation in the colons of mice (Qi, 2013). FOXO3's tumor suppressor ability functions by inducing a program of gene

expression that creates a delay in the cell cycle in response to stress stimuli such as DNA damage. This delay gives DNA repair proteins time to correct DNA errors before cell division (Tran et al., 2002). There is also direct evidence that in human colon cancer cells, inactivation of *FOXO3* leads to a decrease in a cell cycle inhibitor and a loss of cell cycle arrest that leads to poor DNA repair and subsequent colon cancer growth (Dijkers et al., 2000).

Subject 51 had a tumor with loss MLH1 and PMS2. This particular subject is known to have germline methylation of *MLH1*. Subject 54 had a tumor with all four MMR proteins retained. The remaining six subjects with *FOXO3* mutations had tumors showing loss of *MLH2* and/or *MLH6*. None of the subjects with mutations in *FOXO3* displayed classic loss of MLH1 and/or PMS2. An interesting observation is that Subject 31, the only other subject in the study beside Subject 54 who had all four MMR proteins retained, has a mutation in the FOXP1 gene which is a direct transcriptional target of FOXO proteins (Boxtel et al., 2013).

It is unclear from the literature how mutations in *FOXO3* could cause a loss of MSH2 and/or MSH6 MMR proteins in tumors. However, somatic loss of MSH2 in sporadic colon cancer can be caused by acquired deletions or partial deletions in one of four genes that regulate MSH2 protein degradation (Diouf et al., 2011). Using human leukemia cells, Diouf et al. demonstrated that when any of these four genes were suppressed, MSH2 protein degradation was increased, leading to a significant reduction in DNA mismatch repair. They subsequently found somatic deletions in the four genes in adult acute lymphoblastic leukemia and sporadic colorectal cancers.

The research by Diouf et al. demonstrates evidence of at least one alternative pathway to MSH2 protein deficiency that does not have any affect on the *MSH2* gene sequence or *MSH2* gene expression. This suggests to me that there are likely other unknown pathways that could lead to MSH2 deficiency. *FOXO3* could either be involved in regulating *MSH2* gene expression or gene expression of the proteins that regulate *MSH2* protein degradation.

A major setback to *FOXO3* as a candidate gene is that Subject 62, the affected mother in Family O, has a mutation in *FOXO3*, while the two affected children do not, suggesting that *FOXO3* is not segregating with disease.

JAG2 (Jagged 2)

JAG2 came to my attention because it is the only candidate gene that segregates with disease in Family O. The variant shared between affected Mom, Son, and Daughter is 1561C>T, which has a population frequency of .05%. The unaffected father does not have a mutation in *JAG2*. Subject 54 who has all four MMR proteins retained and a tumor with high MSI also has a mutation in *JAG2*, 783C>T.

Jagged-2 (*JAG2*) is one of the five known Notch ligands. The Notch signaling pathway regulates the determination of cell-fate during human development and homeostasis of adult tissues (Zyiad & Iruela-Arispe, 2011), and the pathway is initiated when a Notch ligand such as JAG2 binds to a Notch receptor. A change in the activation rate of the Notch pathway can facilitate malignant transformation and progression of tumors (Ranganathan et al., 2011). The effect of the Notch pathway on

individual cells is dependent on signal dosage and includes increased cell survival or cell death, proliferation or growth arrest, and either differentiation or blockage of differentiation (Roy et al., 2007).

The Notch signaling pathway is unique in that it can act as an oncogene and a tumor suppressor. For example, high levels of NOTCH3, JAG1 and JAG2 protein expression was found in serous and clear cell ovarian carcinomas compared to their benign counterparts, suggesting that here the Notch pathway is likely to play a key role in the oncogenesis of some ovarian cancers by acting as a oncogene (Jung et al., 2010). Likewise, increasing expression levels of Notch ligands such as JAG2 is also correlated with a more aggressive disease course in breast and prostate cancer (Roy et al., 2007). However, in an example of the Notch pathway acting as a tumor suppressor, Notch receptors and ligands (including JAG2) are decreased in endometrial cancer (Jonusiene et al., 2013).

The Notch ligand JAG2 in particular has been found to play an important role in promoting Notch activity; in uveal melanoma a higher amount of JAG2 mRNA was associated with invading cells as compared to non-invading cells. It has been suggested that increased JAG2 expression promotes tumorgenesis and metastasis in lung cancer, breast carcinoma, and multiple myeloma (Asnaghi et al., 2013). There are a total of 52 unique tumor samples with mutations in *JAG2* listed in the COSMIC database. No known phenotype is associated with germline mutations in *JAG2*.

One study found that a truncated JAG1 protein activates the Notch signaling pathway in colorectal cancers leading to more aggressive disease (Lu et al., 2013).

While JAG1 and JAG2 have different biological roles, they do have overlapping NOTCH3 receptor binding specificity (Choi et al., 2009), suggesting truncated *JAG2* could have a similar activating affect on the Notch pathway.

A recent discovery of a colorectal cancer patient with a germline truncating mutation in NOTCH3 and a tumor with biallelic mutations in NOTCH3 suggests that NOTCH3 functions as a cancer predisposition gene in which a "second-hit" in a cell leads to tumorgenesis (Smith et al., 2013). This is evidence that loss of the Notch protein led to colorectal cancer. If loss of the JAG2 protein, a Notch ligand, leads to under expression of the Notch pathway, a similar phenotype could be expected.

MTUS1 (Microtubule-associated tumor suppressor gene)

Seven subjects have mutations in the MTUS1 gene, a known tumor suppressor gene that has been implicated in several cancers, including pancreatic cancer, and head and neck squamous cell carcinomas (Ding, et al., 2011). MTUS1 is characterized as a mitochondrial tumor suppressor gene.

Studies have shown that MTUS1 expression is significantly down-regulated in colon tumors as compared to normal tissues, and reduced MTUS1 expression significantly increases cellular proliferation, suggesting that MUTS1 is involved in the loss of proliferative control in human colon cancer (Zuern et al, 2010).

Only two subjects had the same variant, the rest were unique. All subjects with a mutation *in MTUS1* had tumors with loss of MSH2 and/or MSH6, with the exception of one subject who had all four MMR proteins retained. This gene drew my

attention because of these IHC similarities, and the sheer number of subjects with mutations in *MTUS1*. However, the population frequencies of the variants ranged from 0.05% all the way up to 0.55%. This suggests to me that *MTUS1* may play a role in colon cancer risk, however if it does, it would not be a high penetrance gene as up to 1 in 200 individuals are carriers of the E144G variant.

Known Cancer Susceptibility Genes

MSH3, PMS1 Lynch Syndrome

The *MSH3* (MutS Homolog 3) gene codes for a MMR protein that is involved in the same MMR system associated with Lynch syndrome. The MSH2 protein can form a heterodimer with either MSH6 or MSH3, depending on the type of DNA error that needs to be corrected (Peltomaki, 2003). MSH3 is partially redundant to the function of MSH6, however impaired MSH3 protein activity can lead to a partial defect in the MMR system (Duaturo et al, 2011).

Duaturo et al, screened for mutations in *MSH3* in 79 unrelated LS patients whom where negative for germline mutations in *MLH1*, *MSH2*, and *MSH6*. They found 13 variants, including silent, missense, and intronic variants. One missense mutation in *MSH3*, when present in a carrier of a *MSH2* gene polymorphism, was associated with disease in the family. The authors speculate that variants in *MSH3* gene act as low-risk alleles that contribute to the risk of colon cancer in patients with

Lynch syndrome without germline mutations in *MLH1* or *MSH2.* (Duraturo et al., 2011).

Subject 6 was diagnosed with a MSH2 and MSH6 deficient, MSI-High colon tumor at age 50. He has a missense mutation (2041C>T) in the *MSH3* gene, found in .05% in the population (dbSNP). He does not have any first-degree relatives with a history of LS related cancer. It seems possible that this missense mutation in *MSH3* affects the MMR system and contributed to a risk for colon cancer. However, as there is no evidence in the literature of *MSH3* variants alone causing LS, I would speculate that Subject 6 is a carrier of other LS risk alleles.

PMS1 (Postmeiotic segregation increased 1) is a MMR protein that can complex with MLH1, however *PMS1* seems to play a redundant part to *PMS2*. There is very little information about the role of *PMS1* in LS available in the literature. While mutations in *PMS1* are assumed to cause some disruption to the MMR system, to date no mutations in *PMS1* have been associated with LS or an increased risk for colorectal cancer (Liu et al., 2001).

Subject 11 has a variant in *PMS1* as well as a variant in *FLCN*. The *PMS1* variant Subject 11 carriers is 704G>A, and it has a population frequency of .09% (dbSNP). I can only speculate that this variant in PMS1 could account for the colon cancer that this patient had that did not quite fit in to the BHDS phenotype.

BAP1, Tumor Predisposition Syndrome

BAP1 (*BRCA1* associated protein 1) is a tumor suppressor and somatic mutations in BAP1 in have been documented in multiple human cancers for quite some time. Recently researches identified germline mutations in *BAP1* that are associated with an increased risk for several types of cancer. The association is strong enough that researchers now consider germline *BAP1* mutations the cause of distinct novel cancer predisposition syndrome characterized by a susceptibility for malignant mesothelioma, uveal melanoma, and cutaneous melanoma (Carbone et al., 2013). In addition, susceptibility to renal cell carcinomas was recently added to the *BAP1* syndrome, and research into the role of *BAP1* in other tumors continues (Popova et al., 2013).

BAP1 syndrome, creatively termed Tumor Predisposition Syndrome, is presumably rare and has only recently been recognized, so the full clinical picture is still unknown. Subject 5 was diagnosed at 46 with chondrosarcoma, a type of bone cancer, and at 50 was diagnosed with an endometrial tumor that was PMS2 deficient. This subject has a *BAP1* mutation, 1735G>a (G579R). This variant is presumably novel, as it was not found in dbSNP. Simply because Tumor Predisposition Syndrome is no well understood, this finding would make an interesting case study if we could associate this syndrome with bone and endometrial cancer for the first time.

BRCA2, Hereditary Breast and Ovarian Cancer Syndrome

Deleterious mutations in *BRCA1* and *BRCA2* (Breast cancer gene 1,2) cause Hereditary Breast and Ovarian Cancer (HBOC) syndrome, the most common cancer predisposition syndrome. Woman with HBOC have a high lifetime risk for breast and ovarian cancer. An increased frequency of other cancers has been observed in carriers of *BRCA1/2* genes, including prostate cancer (in men), and pancreatic cancer. Carriers of *BRCA2* mutations also are at increased risk for melanoma (NCCN).

Subjects 55 and 59 have the same variant in the *BRCA2* gene and they also have similar cancer histories. Subject 55 was diagnosed with endometrial cancer showing loss of MSH2 and MSH6 at the age of 57. Subject 59 was diagnosed with endometrial cancer at age 53, however her tumor showed loss of all four MMR proteins. The variant they have in common, 8851G>A, A2951T, is a known variant that is classified as non-pathogenic with a population frequency of approximately 0.7%, or roughly 1 in 140 (dbSNP), while the prevalence of disease causing mutations in *BRCA2* combined with *BRCA1* mutations in the general population is even less, between 0.33% and 0.13% (NCCN). *BRCA2* mutations have been associated with increased risk of endometrial cancer, but some data suggest this is due to the use of chemopreventive drugs and not presence of the mutation (NCCN). Even though it seems unlikely that this *BRCA2* allele is deleterious, the presence of this same variant in two subjects with similar clinical pictures is intriguing.

FLCN, Birt-Hogg-Dube Syndrome

FLCN **(**Folliculin) is the only gene known to cause the autosomal dominant cancer predisposition syndrome Birt-Hogg-Dube (BHDS). The most common clinical manifestations of BHD include cutaneous lesions, pulmonary cysts, and various types of renal tumors (Menko et al, 2009). There is great variation in the clinical symptoms of BHDS both within and between families. Some studies have failed to find an association between mutations in FLCN and colorectal cancer, while others have found an increased risk for colon polyps in individuals with BHDS (Mota-Burgos, et al, 2013). However, somatic mutations in FLNC in sporadic MSI-High colorectal cancers suggest that inactivation of FLCN may contribute to colorectal tumorgenesis. (Nahorski et al., 2010). Most pathogenic mutations in FLNC reported in BHDS are nonsense mutations that result in a truncated protein, however several missense mutations in FLNC have also been reported (Mota-Burgos, et al, 2013).

Subject 11 has a missense mutation (1354G>A) in *FLCN* that has an overall population frequency of .09% or approximately 1 in 1090. However, the variant was not found in any of 122 African American samples (1000G). Subject 11 has a unique cancer history in the context of our other subjects. This individual is a female African American who was diagnosed with colon cancer at age 63, clear cell kidney carcinoma at 65, and vulvar basal cell carcinoma at 66. The colon cancer showed loss of MSH1 and PMS2, and was MSI-High.

The median age of diagnosis for renal tumors in BHDS is 48 (Menko et al, 2009). Renal clear cell carcinoma, while in the minority, is a type of kidney cancer

associated with BHDS. Only six cases of melanoma have been reported in association with BHDS, however a recent case was reported of a male with a heterozygous missense mutation (553T>Y) in *FLCN* who presented with melanoma at age 54. He also had several colon polyps, but the authors were hesitant to associate this last finding with the *FLNC* mutation due to the controversy over the role of colorectal cancer in BHDS (Mota-Burgos, et al, 2013). I could not find an association between BHDS and basal cell carcinoma, but *FLCN* is expressed in skin tissues (Warren et al., 2004).

The presence of Subject 11's *FLCN* variant in dbSNP at a frequency of .09% perhaps rules out her 1354G>A mutation as disease causing. However, because of her unique cancer history, and the fact that she is our only subject with renal clear cell carcinoma, it would be interesting to look further into her medical and family history for other signs of BHDS.

APC, Adenomatous polyposis coli

Germline mutations in the adenomatous polyposis coli gene (APC) cause Familial Adenomatous Polyposis (FAP), an autosomal dominant cancer syndrome characterized by 100s to 1,000s of colon adenomatous polys. There are several subtypes of FAP also caused by mutations in *APC*, including a less severe form called Attenuated FAP, Turcot syndrome which involves colonic polyposis and CNS tumors, and Garner syndrome which involves colon polyposis with osteomas and soft tissue tumors. Approximately 98% of disease-causing mutations in *APC* are nonsense

or frameshift mutations, and the remaining 2% are deletions or duplications and only rarely splice site mutations. Only one missense mutation has been reported pathogenic in the literature (Kerr, et al., 2013).

Subject 59 has a missense variant in the *APC* gene that is not listed in dbSNP (2926A>G). This subject was diagnosed with endometrial cancer at age 53, and her tumor showed loss of all four LS MMR proteins. Her family history is rather unremarkable, except that she has four first and second-degree relatives with stomach cancer. Neither FAP nor its subtypes are associated with endometrial cancer, however gastric polys are common and the risk for gastric cancer is increased, but still less than 1% lifetime risk (Kerr, et al., 2013).

It is unlikely that the 2926A>G APC variant in Subject 59 is disease-causing, as missense mutations are almost unheard of in FAP, and her clinical history is not consistent with FAP or any of the FAP subtypes.

MUTYH: MutYH Associated Polyposis Syndrome

Mutations in the *MUTYH* (mutY homolog) gene cause a cancer predisposition syndrome with a colorectal phenotype similar to FAP and AFAP called MutYH-associated polyposis syndrome (MAP). A major difference between MAP and the other cancer syndromes discussed thus far is that MAP is inherited in an autosomal recessive pattern; individuals with MAP inherit two nonfunctional copies of the *MUTYH* gene. The *MUTYH* protein interacts directly with the MMR heterodimer composed of MSH2 and MSH6 MMR proteins (Sampson, 2005).

Subject 51 was diagnosed with endometrial cancer at age 42 and colon cancer at 43. Her medical records also indicate that she has benign cysts on her leg, back, and face, though I do not have further information regarding their classification. At least one of the subjects tumors showed loss of MLH1 and PMS2, however I do not have information on whether it was the endometrial tumor or colon tumor. After testing negative for germline mutations in *MLH1* and *PMS2*, Subject 51 was found to harbor germline methylation of the *MLH1* promoter leading to silencing of the *MLH1* gene. She does not have any first-degree relatives with any LS related tumors.

Subject 51 has the variant 301C>A in the *MUTYH* gene, which has a population frequency of .05% (dbSNP). The affects of heterozygous *MUTYH* mutations on colon cancer risk is unknown, however a recent case study describes the G382D *MUTYH* mutation causing colorectal cancer in a dominant fashion (Khalaf, 2013). I could not find any evidence of a link between mutations in *MUTYH* and germline methylation of *MLH1*. Subject 51's cancer predisposition is more likely caused by the germline *MLH1* methylation.

Limitations

This research project has serious limitations beginning after the exome sequencing was completed. One main issue was that I lacked continual access to software designed to mine through large amounts of sequencing data. Physical location limited my time at SoftGenetics headquarters to approximately one week. The data set was so large that it took over 48 of continual computing to align the

exome sequences to the reference human genome. The time factor necessitated that I create a plan to filter the data that would quickly narrow down the candidate variants to a number that I could manage on my personal computer.

The use of the 809 genes linked to cancer as a filtering tool created huge limitations in the type of variants uncovered. The list from was from 2004 so it lacked any new cancer genes discovered in the last nine years. Nine years, especially these last nine years, is an extremely long time in terms of genetic research. Later on in my literature review I came across a recent paper that listed 1139 genes known to be involved in colorectal tumorgenesis specifically (Smith, et al., 2013), indicating that there are most definitely potential candidate genes that were overlooked.

I also lacked the ability to do large variant comparisons between groups with similar phenotypes, and between members of Family O. The NextGENe software has a capacity of only three exomes for its variant comparison tool. The ability to find common variants between all research subjects could potentially lead to the discovery of a common novel gene. Furthermore, I lack the materials to conduct any follow-up genetic testing of either tumor material for other family members to attempt to confirm any of the candidate variants that I did identify.

Lastly, a limiting factor was my own inexperience in dealing with large sets of sequence data. Many of the initial decisions I made regarding variant filtering were uninformed and I was unable to go back to the original data to make adjustments.

CONCLUSION

Lynch syndrome is a complicated genetic disorder to study. In addition to being caused by mutations any of in four mismatch repair genes, sporadic cancer can present with similar clinical symptoms and similar tumor pathology. In addition, rare causes of Lynch syndrome such as EpCAM deletions and germline MLH1 promoter methylation complicate the search for a molecular diagnosis. In patients with Lynch syndrome who do not have identifiable MMR gene mutations, genetic counselors and other health care providers find themselves at a loss for the next diagnostic step.

Exome sequencing is emerging as a powerful tool that is able to simultaneously sequence the 1% of the genome that is protein coding. Presumably, novel genetic causes for LS in individuals without germline MMR mutations can be found somewhere in the exome. We used exome sequencing in a cohort of 32 patients to mine for novel causes of LS. We did not find any disease causing mutations in the known MMR genes that the traditional Sanger sequencing missed. While the results do highlight some rare variants that may play a role in cancer predisposition, there is a vast amount of genetic information produced in this study remains to be sorted though. Future studies using this same set of data will hopefully provide our research subjects with the definitive answer they were hoping for.

REFERENCES

ACMG Board of Directors. (2012). Points to consider in the clinical application of genomic sequencing. *Genetics in Medicine,* 14(8), 759-761. doi: 10.1038/gim.2012.74

Aktan-Collan, K., Mecklin, J. P., Jarvinen, H., Nystrom-Lahti, M., Peltomaki, P., Soderling, I., … Kaarlainen, H. (2000). Predictive genetic testing for hereditary non-polyposis colorectal cancer: Uptake and long-term satisfaction. *Internal Journal of Cancer,* 89(1), 44-50.

Arden, K. (2006). Multiple roles of FOXO transcription factors in mammalian cells point to multiple roles in cancer. *Experimental Gerontology,* 41(8), 709-717.

Asnaghi L., Handa J.T., Merbs S.L., Harbour J.W., Eberhart C.G. (2013). A role for Jag2 in promoting uveal melanoma dissemination and growth. *Investigative Opthamology and Visual Science,* 54(1). 295-306.

Balakrishnan, A., Bleeker, F., Lamba, S., Rodolfo, M., Daniotti, M., Scarpa, A., A., Leenstra, S., Zanon., C., Bardelli, A. (2007). Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic carcinoma. *Cancer Research,* 67. 3545-3550.

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., & Shendure, J. (2011). Exome sequencing as a tool for mendelian disease gene discovery. *Nature Reviews Genetics*, 12, 745-754. doi: 10.1038/nrg3031

Bansidhar, B. J., & Silinsky, J. (2012). History and Pathogenesis of LS. *Clinical Colon Rectal Surgery*, 25, 63-66.

Berger, A., Knudson, A., Pandolfi, P. (2011) A continuum model for tumor suppression. *Nature*, 476(7359). 163-169.

Boland, C. R. (2013) The mystery of mismatch repair deficiency; Lynch or Lynch-like? *Gastroenterology*, 144(5). 868-870. Editorial.

Burt, R. W., Barthel, J. S., Cannon, J., Dunn, K. B., David, D. S., Early, D. S., Weinberg, D. (2010). *National Comprehensive Cancer Network (NCCN) clinical practice guidelines in oncology [Colorectal cancer screening]*. Version 2.2011, NCCN.org

Chapelle, A. D. (2005). The Incidence of Lynch Syndrome. *Familial Cancer*, 4(3), 233-237.

Choi, K., Ahn, Y., Gibbons, D., Tran, H., Creighton, C., Girard, L., Minna, J., Qin, X., Kurie, J. (2009) Distinct biological roles for the Notch ligands Jagged-1 and Jagged-2. *The Journal of Biological Chemistry*, 284(26).

Choi, M., Scholl, U., Weizhen, J., Lui, T., Zumbo, P., Nayir, A., Bakkaloglu, A., et al. (2009). Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 106(45), 19096-19101. doi:10.1073/pnas.0910672106

Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S,…Futreal PA (2002). Mutations of the BRAF gene in human cancer. *Nature* 417(6892). 949–54.

De la Chapelle, A., & Hampel, H. (2010). Clinical relevance of microsatellite instability in colorectal cancer. *Journal of Clinical Oncology*, 28(20). doi:10.1200/JCO.2009.27.0652

Dijkers, P., Medema, R., Pals, C., Banerji, L., Thomas, N., Lam, E., ... Coffer, P.

(2000). Forkhead transcription factor FKHR-L1 modulates cytokine-dependent transcriptional regulation of p27(KIP1). *Molecular Cell Biology*, 20(24), 9138-9148.

Dimmock, D. P. (2011). Making a definitive diagnosis: Successful clinical application of WES in a child with intractable inflammatory bowel disease. *Genetics in Medicine*, 13(3), 255-262.

Diouf, B., Chen, Q., Krynetskaia, N.F., Yang, W., Cheok, M., Pei, D., Fan, Y., Cheng, C.,...Evans, W.E. (2011) Somatic deletions of genes regulating MSH2 protein stability cause DNA mismatch repair deficiency and drug resistance in human leukemia cells. *Nature Medicine,* 17(10). 1298-1303

Duraturo, F., Liccardo, R., Cavallo, A., De Rosa, m., Izzo, P. (2011) Association of low-risk MSH3 and MSH2 cariant alleles with Lynch syndrome: a probability of synergistic effects. *International Journal of Cancer,* 129(7). 1643-1650.

Foster, M. W., Mulvihill, J. J., & Sharp, R. R. (2009). Evaluating the utility of personal genomic information. *Genetics in Medicine*, 11(8), 570-574. doi: 10.1097/GIM.0b013e3181a2743e

Futreal, A. P., Coinn, L., Marshall, M., Down, T., Hubbard, T.,Wooster, R., Rahman,

N., and Stratton, M. (2004). A census of Human cancer genes. *Nature* Reviews, Cancer, 4(3), 177-183.

Haimovich, A. D. (2011). Methods, challenges, and promise of next-generation sequencing in cancer biology. *Yale Journal of Biology and Medicine*, (84), 439-446.

Hendricks, M. C., Jong, A., Morreau, J., Tops, C., Vasen,, H., Wijnen, J.Breuning, M., Brocker-Vriends, A.(2006) Diagnostic approach and management of Lynch syndrome (hereditary nonpolyposis colorectal carcinoma): A guide for clinicians. *A Cancer Journal for Clinicians,* 56(4), 213-225.

Herfarth, K.K., Kodner, I.J., Whelan, A.J., Ivanovich, J.L., Bracamontes, J.R., Wells, S.A. Jr., Goodfellow, J.P. (1997) Mutations in MLH1 are more frequent than MSH2 in sporadic colorectal cancers with microsatellite instability. *Genes, Chromosomes, and Cancer*, 18(1). 42-49.

Hitchins, M. P., & Ward, R. L. (2009). Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. *Journal of Medical Genetics*, 46, 793-802. doi: 10.1136/jmg.2009.068122

Hitchins, M.P., Rapkins, R.W., Kwok, C.T., Srivastava, S., Wong, J.J., Khachigian,

L.M., Polly, P., Goldblatt, J., Ward, R.L. (2001) Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. (2011). *Cancer Cell,* 20(2), 200-213. doi: 10.1016/j.ccr.2011.07.003

Jamal, S. M., Yu, J. H., Chong, J. X., Dent, K. M., Conta, J. H., Tabor, H. K., & Bamshad, M. J. (2013). Practices and policies of clinical exome sequencing providers: Analysis and implications. *American Journal of Medical Genetics,*161(5), 935-950. doi: 10.1002/ajmg.a.35942

Jensen, L. H., Lindebjerg, J., Kolvraa, S., & Cruger, D. G. (2009). Molecular screening for Lynch syndrome: From bench to bedside. *American Society of Clinical Oncology, 27(34).* doi: 10.1200/JCO.2009.24.6744. Letter to the editor

Jonusiene, V., Sasnauskiene, A., Kanopiene, D., Sasnauskiene, S., Kazbariene, B., Didziapetriene, J. (2013) Down-regulated expression of Notch signaling molecules in human endometrial cancer. *Medical Oncology,* 30(1). 483.

Jung, G., Kwon, Y., Song J., Back, M., Lee, S., Lee, C., Hwang, Y., An, H. (2010). Prognostic significance of Notch 3 gene expression in ovarian serous carcinoma. *Cancer Science.* 101(9). 1977-1983.

Khalaf, R., Jones, C., Strutt, W., Williamson, P. (2013) Colorectal cancer in a monoallelic MYH mutation carrier. *Journal of Gastrointestinal Surgery.*

Kibushi, H., Mohseni, M., Tamaki, A., Garay, J., Croessmann, S., Karnan, S., Ota, A., Wong, H., Konishi, Y., Karakas, B., Tahir, K……Park, B. H. (2011) Mutation of a single allele of the cancer susceptibility gene BRCA1 leads to genomic instability in human breast epithelial cells. *Proceedings of the National Academy of Sciences of the United States of America*, 108(43). 17773-17778.

Ku C. S., Cooper D. N., Iacopetta B., & Roukos D. H. (2012). Integrating next-generation sequencing into the diagnostic testing of inherited cancer predisposition. *Clinical Genetics*, 83(1), 2-6.

Ku, C., Cooper, D., Polychronakos, C., Naidoo, N., Wu, M., & Soong, R. (2012). Exome sequencing: dual role as a discovery and diagnostic tool. *Annals of Neurology*, 71, 5-14. doi:10.1002/ana.22647

Kuiper, R. P., Vissers L., Venkatachalam, R., Bodmer, D., Hoenselaar, E., Goossens, M., Haufe, A., Kamping, E., Niessen, R., Hogervorst, F., Gille, J., Redeker, B., Tops, C., van Gijn, N., van den Ouweland, A. et al., (2011). Recurrence and variability of germline EPCAM deletions in LS. *Human*

Mutation, Variation, Informatics, and Disease, 32(4), 407-414. doi:10.1002/humu.21446

Lindor, N. M., Petersen, G. M., Hadley, D. W., Kinney, A. Y., Miesfeldt S., Lu K. H., Lynch, P., Burke, W., & Press, N. (2006). Recommendations for the care of individulas with an inherited predisposition to Lynch syndrome. *Journal of the American Medical Association*, 296(12), 1507-1517.

Liu, T., Yan, H., Kuismanen, S., Percesepe, A., Bisgaard, M., Pedroni, M., Banatti, P, Zinzler, K., Vogelstein, B., Leon, M., Lindblom, A. (2001) The role of hPMS1 and hPMS2 in predisposing to colorectal cancer. *The Journal of Cancer Research*, 61.

Lou, H., Yang, Y., Wu, P., Jiang, Q., & Xu, C. (2013). PTEN-regulated AKT/FoxO3a/Bim signaling contributes to reactive oxygen species-mediated apoptosis in selenite-treated colorectal cancer cells. *Cell Death and Disease,* 4(E481). doi: 10.1038/cddis.2013.3.

Lu J., Ye X., Fan F., Xia L., Bhattacharya R., Bellister S., Tozzi F., Sceusi E., Zhou Y., Tachibana I., Maru D.M., Hawke D.H., Rak J., Mani S.A., Zweidler-McKay P., Ellis L.M. (2013). Endothelial cells promote the colorectal cancer

stem cell phenotype through a soluble form of Jagged-1. *Cancer Cell*, 23(2). 171-175.

Medema,R.H.,Kops,G.J.,Bos,J.L.,andBurgering,B.M.(2000)AFX-like Forkhead transcription factors mediate cell-cycle regulation by Ras and PKB through p27kip1. *Nature* 404. 782–787.

Menko, F., van Steensel, M., Giraud, S., Frils-Hansen, L., Richard, S., Ungari, S., Nordenskjold, M., Hansen T., Solly, J., Maher, E. (2009) Birt-Hogg-Dube syndrome: diagnosis and management. *The Lancet Oncology*, 10(12). 1199-1206.

Mota-Burgos, A., Herrera, A., Marquez, F., Mendiola, M., Herrera-Ceballos, E. (2013) Birt-Hogg-Dube syndrome in a patient with melanoma and a novel mutation in the FCLN gene. *International Journal of Dermatology*, 52(3). 323-326.

Mvundura, M., Grosse, S. D., Hampel, H., & Palomaki, G. E. (2010). The cost-effectiveness of genetic testing strategies for LS among newly diagnosed patients with colorectal cancer. *Genetics in Medicine,* 12(2), 93-104. doi: 10.1097/GIM.0b013e3181cd666c

Nahorski, M., Martin, L., Gille, J., McKay, K., Rehal, P., Ploeger, H., van Steensel, M., Tomlinson, I., Latif, F., Menko, F., Maher, E. (2010) Investigation of Birt-Hogg-Dube tumor suppressor gene (FLCN) in familial and sporadic colorectal cancer. *Journal of Medical Genetics*, 47(6). 385-390.

Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., … Bamshad, M.J. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics,* 42(1), 30-35. doi: 10.1038/ng.499

Peltomaki, P. Role of DNA mismatch repair defects in the pathogenesis of human cancer. (2003) *Journal of Clinical Oncology,* 21(6). 1174-1179.

Qi, W., Fitchev, P., Cornwell, M., Greenberg, J., Cabe, M., Weber, C., ... Savkovic, S. (2013). FOXO3 growth inhibition of colonic cells is dependent on Intraepithelial lipid droplet density. *The Journal of Biological Chemistry*, 288(23), 16274-16281.

Ranganathan, P., Weaver, K., Capobianoco, A. (2011) Notch signaling in solid tumors: a little bit of everything but not all the time. *Nature Reviews: Cancer*, 11. 338-351.

Reagan-Shaw, S., & Ahmad, N. (2007). The role of Forkhead-box Class O (FoxO)

transcription factors in cancer: A target for the management of cancer. *Toxicology and Applied Pharmacology*, 224(3), 360-368.

Robinson, K. L., Liu, T., Vandrovocova, J., Halvarsson, B., Clendenning, M., Frebourg, T., … Kinzler, K. W. (2007). Lynch syndrome (hereditary nonpolyposis colorectal cancer) diagnostics. *Journal of the National Cancer Institute*, 99(4), 291-299. doi: 10.1093/jnci/djk051

Rodriguez-Soler, M., Perez-Carbonell, L., Guarinos, C., Zapater, P., Castillejo, A., Barbera, V….Jover, R. Risk of cancer in cases of suspected Lynch syndrome without germline mutation. (2013) *Gastroenterology*, 144. 926-932.

Roy, M., Pear, W., Aster, J. (2007) The multifaceted role of Notch in cancer. *Current Opinion in Genetics and Development*, 17. 52-59.

Salmena, L. & Narod, S. (2012) BRCA1 haploinsufficiency: consequences for breast cancer. *Women's Health,* 8(2). 127-129.

Sampson, J., Jones, S., Dolwani, S, Cheadle, J. (2005) MutYH (MYH) and colorectal cancer. *Biochemical Society Transactions*, 33. 679-683.

Schneider, K. A. (2012). Counseling about cancer, strategies for genetic counseling. Hoboken, NJ: Wiley-Blackwell.

Senior, P., Zhang, B., Chan, S. (2010) Loss of cell-surface receptor EphB2 is important for growth, migrations, and invasiveness of a colon cancer cell line. *International Journal of Colorectal Disease* 25(6). 687-694.

Sherr, C. (2004) Principals of tumor suppression. *Cell,* 116. 235-246.

Singleton, A. B. (2011). Exome sequencing: a transformative technology. *Lancelet Neurology*, (10), 942-946

Smith, C., Naven, M., Harris, R., Colley, J., West, H., Li, N., Liu, Y., Adams, R.,…Cheadle, J. (2013). Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. *Human Mutation*, 34(7). 1026-1034.

Tran, H., Brunet, A., Grenier, J., Fornance, A., Jr., DiStefanco, P., Chiang, L., & Greenberg, M. (2002). DNA repair pathway stimulated by the forkhead transcription factor FOXO3a through the Gadd45 protein. *Science,* 296(5567), 530-534.

Vilar, E., & Gruber, S. B. (2010). Microsatellite instability in colorectal cancer- the stable evidence. *Nature Reviews, Clinical Oncology*. doi: 10.1038/nrclinonc.2009.237

Warren, M., Torres-Cabala, C., Turner, M., Merino M., Matrosova, V., Nickerson, M., Ma, W., Linehan, W., Zbar, B., Schmidt, L. (2004) Expression of Birt-Hogg-Dube gene mRNA in normal and neoplastic human tissues. *Modern Pathology*, 17(8). 998-1011.

Worthey, E. A., Mayer, A. N., Syverson, G. D., Helbling, D., Bonacci, B. B., Decker, B., Dimmock, D. P. (2011). Making a definitive diagnosis: Successful clinical application of WES in a child with intractable inflammatory bowel disease. *Genetics in Medicine*, 13(3), 255-262.

Yang, A. (2000). Structure and function of mismatch repair proteins. *Mutation Research*, 469(3-4), 245-56.

Zhang X, Li J. Era of universal testing of microsatellite insta- bility in colorectal cancer. *World J Gastrointest Oncol* 2013; 5(2): 12-19 DOI: http://dx.doi.org/10.4251/ wjgo.v5.i2.12

Zhuang, Z., Gusev, A., Cho J., Pe'er, I. (2012) Detecting identity by descent and

homozygosity mapping in whole-exome sequencing data. *PLos ONE* 7(10). doi: 10.1371/journal.pone.0047618

Ziyad, S. & Iruela-Arispe, M. L. (2011). Molecular mechanisms of tumor angiogenesis. *Genes & Cancer*, 2(12). 1085-1096.

Zogopoulos, G., Jorgensen, C., Montpetit, A., Lepage, A., Ferretti, V., Chad, L., Selvarajah, S., Zanke, B., J. Hudson, Pawson, T, Gallinger, S. (2008) Germline EPHB2 receptor variants in familial colorectal cancer. *PLoS ONE*, 3(8).

Zuern, C., Heimrisk, J., Kaufmann, R., Richter, KK., Settmacher, U., Wanner, C., Galle, J., Seibold, S. (2010). Down-regulation of MTUS1 in human colon tumors. *Oncology Reports*, 32(1). 183-189.

APPENDIX

APPENDIX A

GENES CAUSALLY IMPLICATED IN ONCOGENESIS

From a Census of Human Cancer Genes (Futreal et al., 2004)

Ensembl ID	Gene Name	Ensembl ID	Gene Name
ensg00000002822	MD1L1	ensg00000135318	5NTD
ensg00000002834	LASP1	ensg00000135363	RBTN2
ensg00000003400	CASPA	ensg00000135446	CDK4
ensg00000004534	RBM6	ensg00000135541	AHI1
ensg00000004838	ZMY10	ensg00000135605	TEC
ensg00000004948	CALCR	ensg00000135679	MDM2
ensg00000005073	HXA11	ensg00000135828	RN5A
ensg00000005339	CBP	ensg00000135903	PAX3
ensg00000005893	LAMP2	ensg00000136167	PLSL
ensg00000005961	ITA2B	ensg00000136231	IF2B3
ensg00000006468	ETV1	ensg00000136244	IL6
ensg00000006704	GT2D1	ensg00000136492	FANCJ
ensg00000006744	RNZ2	ensg00000136717	BIN1
ensg00000007237	GAS7	ensg00000136754	ABI1
ensg00000007350	TKTL1	ensg00000136848	DAB2P
ensg00000007372	PAX6	ensg00000136936	XPA
ensg00000008226	DLEC1	ensg00000136997	MYC
ensg00000009709	PAX7	ensg00000137070	I11RA
ensg00000010704	HFE	ensg00000137074	APTX
ensg00000011052	NDKA	ensg00000137193	PIM1
ensg00000012048	BRCA1	ensg00000137265	IRF4
ensg00000012061	ERCC1	ensg00000137309	HMGA1
ensg00000012171	SEM3B	ensg00000137364	TPMT
ensg00000012232	EXTL3	ensg00000137497	NUMA1
ensg00000015285	WASP	ensg00000137713	2AAB
ensg00000019549	SNAI2	ensg00000137727	RHG20
ensg00000020922	MRE11	ensg00000137745	MMP13
ensg00000023287	RBCC1	ensg00000137812	CASC5
ensg00000023445	BIRC3	ensg00000137868	STRA6
ensg00000025293	PHF20	ensg00000137948	BRDT
ensg00000026103	TNR6	ensg00000138039	LSHR

ensg00000027075	KPCL	ensg00000138293	NCOA4
ensg00000029725	RABE1	ensg00000138294	MSMB
ensg00000032514	ERCC6	ensg00000138311	TALAN
ensg00000034152	MP2K3	ensg00000138336	CXXC6
ensg00000037280	VGFR3	ensg00000138363	PUR9
ensg00000038427	CSPG2	ensg00000138376	BARD1
ensg00000039068	CADH1	ensg00000138463	DIRC2
ensg00000040633	PHF23	ensg00000138615	CILP1
ensg00000043093	DCNL1	ensg00000138674	SC31A
ensg00000044115	CTNA1	ensg00000138698	GDS1
ensg00000044524	EPHA3	ensg00000138758	SEPT11
ensg00000046604	DSG2	ensg00000139083	ETV6
ensg00000047410	TPR	ensg00000139618	BRCA2
ensg00000047932	GOPC	ensg00000139679	P2RY5
ensg00000047936	ROS	ensg00000139687	RB
ensg00000048052	HDAC9	ensg00000139910	NOVA1
ensg00000048392	RIR2B	ensg00000140262	HTF4
ensg00000048462	TNR17	ensg00000140396	NCOA2
ensg00000049167	ERCC8	ensg00000140416	TPM1
ensg00000049319	S5A2	ensg00000140464	PML
ensg00000051180	RAD51	ensg00000140538	NTRK3
ensg00000054118	TR150	ensg00000140937	CAD11
ensg00000060718	COBA1	ensg00000141342	PKHM1
ensg00000061337	LZTS1	ensg00000141367	CLH1
ensg00000062038	CADH3	ensg00000141376	BCAS3
ensg00000064012	CASP8	ensg00000141380	SSXT
ensg00000064933	PMS1	ensg00000141510	P53
ensg00000065320	NET1	ensg00000141524	TMC6
ensg00000065361	ERBB3	ensg00000141646	SMAD4
ensg00000065559	MP2K4	ensg00000141736	ERBB2
ensg00000066455	GOGA5	ensg00000141867	BRD4
ensg00000066468	FGFR2	ensg00000141985	SH3G1
ensg00000067066	SP100	ensg00000142405	NAL12
ensg00000067082	KLF6	ensg00000142599	RERE
ensg00000067182	TNR1A	ensg00000142611	PRD16
ensg00000067208	EVI5	ensg00000142867	BCL10
ensg00000067798	NAV3	ensg00000143252	C560
ensg00000067955	PEBB	ensg00000143294	PRCC
ensg00000068078	FGFR3	ensg00000143297	FCRL5
ensg00000068323	TFE3	ensg00000143322	ABL2

ensg00000068976	PYGM	ensg00000143416	SBP1
ensg00000069399	BCL3	ensg00000143437	ARNT
ensg00000069974	RB27A	ensg00000143473	KCNH1
ensg00000070371	CLH2	ensg00000143514	ASPP2
ensg00000070404	FSTL3	ensg00000143549	TPM3
ensg00000071243	ING3	ensg00000143924	EMAL4
ensg00000071564	TFE2	ensg00000144218	AFF3
ensg00000071967	CYBR1	ensg00000144381	CH60
ensg00000072274	TFR1	ensg00000144476	CXCR7
ensg00000072364	AFF4	ensg00000144554	FACD2
ensg00000072609	CHFR	ensg00000145012	LPP
ensg00000072694	FCG2B	ensg00000145022	TCTA
ensg00000073009	NEMO	ensg00000145216	FIP1
ensg00000073282	P63	ensg00000145623	OSMR
ensg00000073670	ADA11	ensg00000145715	RASA1
ensg00000073756	PGH2	ensg00000145819	RHG26
ensg00000073792	IF2B2	ensg00000146648	EGFR
ensg00000073921	PICAL	ensg00000146676	PURB
ensg00000074047	GLI2	ensg00000147065	MOES
ensg00000074800	ENOA	ensg00000147133	TAF1
ensg00000075043	KCNQ2	ensg00000147140	NONO
ensg00000075292	ZN638	ensg00000147257	GPC3
ensg00000075539	FRYL	ensg00000147324	MFHA1
ensg00000075618	FSCN1	ensg00000147548	NSD3
ensg00000075624	ACTB	ensg00000147654	RCAS1
ensg00000075856	SART3	ensg00000147883	CDN2B
ensg00000075891	PAX2	ensg00000147889	CD2A1
ensg00000076242	MLH1	ensg00000148303	RL7A
ensg00000077150	NFKB2	ensg00000148400	NOTC1
ensg00000077498	TYRO	ensg00000148737	TF7L2
ensg00000077782	FGFR1	ensg00000149177	PTPRJ
ensg00000077942	FBLN1	ensg00000149311	ATM
ensg00000078399	HXA9	ensg00000149418	ST14
ensg00000078403	AF10	ensg00000149948	HMGA2
ensg00000078674	PCM1	ensg00000150275	PCD15
ensg00000078699	MTG8R	ensg00000150527	CTGE5
ensg00000078900	P73	ensg00000150593	PDCD4
ensg00000079102	MTG8	ensg00000150907	FOXO1
ensg00000079263	LY10	ensg00000151348	EXT2
ensg00000079385	CEAM1	ensg00000151640	DPYL4

ensg00000079432	CIC	ensg00000151702	FLI1
ensg00000079999	KEAP1	ensg00000152213	ARL11
ensg00000080644	ACHA3	ensg00000153208	MERTK
ensg00000080824	HS90A;HS902	ensg00000153487	ING1
ensg00000081913	PHLPP	ensg00000153814	JAZF1
ensg00000082805	RB6I2	ensg00000153944	MSI2H
ensg00000083093	PALB2	ensg00000154277	UCHL1
ensg00000083168	MYST3	ensg00000154767	XPC
ensg00000083642	PDS5B	ensg00000154803	FLCN
ensg00000083799	CYLD	ensg00000155380	MOT1
ensg00000084676	NCOA1	ensg00000155465	YLAT1
ensg00000085276	EVI1	ensg00000156006	ARY2
ensg00000085832	EP15	ensg00000156298	TSN7
ensg00000085999	RAD54	ensg00000156650	MYST4
ensg00000086758	HUWE1	ensg00000156970	BUB1B
ensg00000087088	BAX	ensg00000156976	IF4A2
ensg00000087245	MMP2	ensg00000157168	NRG1
ensg00000087460	ALEX	ensg00000157404	KIT
ensg00000087494	PTHR	ensg00000157554	ERG
ensg00000087586	STK6	ensg00000157764	BRAF1
ensg00000087916	S6A14	ensg00000157950	SSX2
ensg00000088808	ASPP1	ensg00000158169	FANCC
ensg00000089280	FUS	ensg00000158636	EMSY
ensg00000091138	S26A3	ensg00000158715	S45A3
ensg00000091483	FUMH	ensg00000158747	NBL1
ensg00000091592	NALP1	ensg00000158813	EDA
ensg00000095002	MSH2	ensg00000159110	INAR2
ensg00000095585	BLNK	ensg00000159113	I10R2
ensg00000096384	HS90B	ensg00000159216	RUNX1
ensg00000096968	JAK2	ensg00000159921	GLCNE
ensg00000097007	ABL1	ensg00000160182	TFF1
ensg00000099869	IG2AS	ensg00000160183	TMPS3
ensg00000099956	SNF5	ensg00000160213	CYTB
ensg00000100095	SE6L1	ensg00000160224	AIRE
ensg00000100105	PATZ1	ensg00000160613	PCSK7
ensg00000100280	AP1B1	ensg00000160801	PTHR1
ensg00000100311	PDGFB	ensg00000160886	LY6K
ensg00000100345	MYH9	ensg00000160957	RECQ4
ensg00000100393	EP300	ensg00000161011	SQSTM
ensg00000100503	NIN	ensg00000162337	LRP5

ensg00000100504	PYGL	ensg00000162367	TAL1
ensg00000100526	CDKN3	ensg00000162374	ELAV4
ensg00000100721	TCL1A	ensg00000162594	IL23R
ensg00000100814	CIP1	ensg00000162676	GFI1
ensg00000100815	TRIPB	ensg00000162735	PEX19
ensg00000100985	MMP9	ensg00000162775	RBM15
ensg00000101017	TNR5	ensg00000162924	REL
ensg00000101191	DIDO1	ensg00000163161	ERCC3
ensg00000101311	FERM1	ensg00000163297	ANTR2
ensg00000101384	JAG1	ensg00000163513	TGFR2
ensg00000101977	MCF2	ensg00000163518	FCRL4
ensg00000102034	ELF4	ensg00000163568	AIM2
ensg00000102145	GATA1	ensg00000163599	CTLA4
ensg00000102245	CD40L	ensg00000163655	GUAA
ensg00000102466	FGF14	ensg00000163902	RPN1
ensg00000102575	PPA5	ensg00000163930	BAP1
ensg00000102678	FGF9	ensg00000164050	PLXB1
ensg00000102854	MSLN	ensg00000164266	ISK1
ensg00000103126	AXN1	ensg00000164362	TERT
ensg00000103197	TSC2	ensg00000164398	ACSL6
ensg00000103266	STUB1	ensg00000164438	TLX3
ensg00000103313	MEFV	ensg00000164611	PTTG1
ensg00000103522	IL21R	ensg00000164690	SHH
ensg00000103876	FAAA	ensg00000164692	CO1A2
ensg00000104044	P	ensg00000164919	COX6C
ensg00000104213	PGFRL	ensg00000164951	PDP1
ensg00000104219	ZDHC2	ensg00000164985	PSIP1
ensg00000104320	NBN	ensg00000165025	KSYK
ensg00000104368	TPA	ensg00000165240	ATP7A
ensg00000104419	NDRG1	ensg00000165280	FANCG
ensg00000104447	TRPS1	ensg00000165288	BRWD3
ensg00000104881	IASPP	ensg00000165392	WRN
ensg00000104884	ERCC2	ensg00000165409	TSHR
ensg00000104899	MIS	ensg00000165471	MBL2
ensg00000104903	LYL1	ensg00000165556	CDX2
ensg00000104967	NOVA2	ensg00000165671	NSD1
ensg00000105146	AURKC	ensg00000165699	TSC1
ensg00000105205	LPPL	ensg00000165731	RET
ensg00000105221	AKT2	ensg00000165795	NDRG2
ensg00000105372	RS19	ensg00000166140	ZFY19

ensg00000105619	TFPT	ensg00000166407	RBTN1
ensg00000105639	INSL3	ensg00000166548	KITM
ensg00000105641	SC5A5	ensg00000166825	AMPN
ensg00000105656	ELL	ensg00000166949	SMAD3
ensg00000105662	CRTC1	ensg00000167034	NKX31
ensg00000105810	CDK6	ensg00000167085	PHB
ensg00000105976	MET	ensg00000167323	STIM1
ensg00000106031	HXA13	ensg00000167460	TPM4
ensg00000106327	TFR2	ensg00000167601	UFO
ensg00000106348	IMDH1	ensg00000167772	ANGL4
ensg00000106366	PAI1	ensg00000167861	CQ028
ensg00000106483	SFRP4	ensg00000167895	TMC8
ensg00000106536	PO6F2	ensg00000167972	ABCA3
ensg00000106635	BCL7B	ensg00000168036	CTNB1
ensg00000106688	EAA3	ensg00000168092	PA1B2
ensg00000107779	BMR1A	ensg00000168267	PTF1A
ensg00000107789	MINP1	ensg00000168283	BMI1
ensg00000107807	TLX1	ensg00000168421	RHOH
ensg00000107882	SUFU	ensg00000168477	TENX
ensg00000108091	CCDC6	ensg00000168610	STAT3
ensg00000108231	LGI1	ensg00000168646	AXN2
ensg00000108292	AF17	ensg00000168692	TSPY1
ensg00000108395	TRI37	ensg00000169031	CO4A3
ensg00000108753	HNF1B	ensg00000169032	MP2K1
ensg00000108821	CO1A1	ensg00000169083	ANDR
ensg00000108924	HLF	ensg00000169087	HBAP1
ensg00000108946	KAP0	ensg00000169174	PCSK9
ensg00000109047	RECO	ensg00000169184	MN1
ensg00000109132	PHX2B	ensg00000169218	RSPO1
ensg00000109220	CHIC2	ensg00000169427	KCNK9
ensg00000109339	MK10	ensg00000169432	SCN9A
ensg00000109471	IL2	ensg00000169679	BUB1
ensg00000109670	FBXW7	ensg00000169684	ACHA5
ensg00000109685	NSD2	ensg00000169696	ASPC1
ensg00000109854	HTAI2	ensg00000169714	CNBP
ensg00000109906	ZBT16	ensg00000169855	ROBO1
ensg00000110092	CCND1	ensg00000169925	BRD3
ensg00000110367	DDX6	ensg00000170266	BGAL
ensg00000110395	CBL	ensg00000170734	POLH
ensg00000110619	SYCC	ensg00000170791	CHCH7

ensg00000110628	S22AI	ensg00000170881	RN139
ensg00000110711	AIP	ensg00000170961	HAS2
ensg00000110713	NUP98	ensg00000171094	ALK
ensg00000110777	OBF1	ensg00000171155	C1GLC
ensg00000110987	BCL7A	ensg00000171310	CHSTB
ensg00000111057	K1C18	ensg00000171320	ESCO2
ensg00000111087	GLI1	ensg00000171444	CRCM
ensg00000111276	CDN1B	ensg00000171476	HOP
ensg00000111537	IFNG	ensg00000171680	PKHG5
ensg00000111696	NT5D3	ensg00000171723	GEPH
ensg00000111790	FGOP2	ensg00000171735	CMTA1
ensg00000111859	CASL	ensg00000171777	GRP4
ensg00000112039	FANCE	ensg00000171791	BCL2
ensg00000112081	SFRS3	ensg00000171843	AF9
ensg00000112448	TRI27	ensg00000171862	PTEN
ensg00000112531	QKI	ensg00000172175	MALT1
ensg00000112561	TFEB	ensg00000172409	CLP1
ensg00000112576	CCND3	ensg00000172493	AFF1
ensg00000112761	WISP3	ensg00000172660	RBP56
ensg00000113083	LYOX	ensg00000172867	K22E
ensg00000113263	ITK	ensg00000173267	SYUG
ensg00000113302	IL12B	ensg00000173432	SAA
ensg00000113318	MSH3	ensg00000174115	TBC3A
ensg00000113594	LIFR	ensg00000174325	DIRC1
ensg00000113721	PGFRB	ensg00000174775	RASH
ensg00000113916	BCL6	ensg00000174808	BTC
ensg00000114026	OGG1	ensg00000174842	GLMN
ensg00000114062	UBE3A	ensg00000175054	ATR
ensg00000114270	CO7A1	ensg00000175197	DDIT3
ensg00000114354	TFG	ensg00000175387	SMAD2
ensg00000114378	HYAL1	ensg00000175595	XPF
ensg00000114737	CISH	ensg00000175832	ETV4
ensg00000114861	FOXP1	ensg00000177000	MTHR
ensg00000114999	TTL	ensg00000177030	DEAF1
ensg00000115170	ACVR1	ensg00000177374	HIC1
ensg00000115339	GALT3	ensg00000177575	C163A
ensg00000115486	VKGC	ensg00000177646	ACAD9
ensg00000115504	EHBP1	ensg00000177666	PLPL2
ensg00000115850	LPH	ensg00000178053	MLF1
ensg00000115904	SOS1	ensg00000178104	MYOME

ensg00000116014	KISSR	ensg00000178105	DDX10
ensg00000116062	MSH6	ensg00000178235	SLIK1
ensg00000116128	BCL9	ensg00000178522	AMBN
ensg00000116132	PRRX1	ensg00000178573	MAF
ensg00000116251	RL22	ensg00000178691	SUZ12
ensg00000116560	SFPQ	ensg00000178999	AURKB
ensg00000116984	METH	ensg00000179094	PER1
ensg00000116990	MYCL1	ensg00000179295	PTN11
ensg00000117000	RLF	ensg00000179583	C2TA
ensg00000117118	DHSB	ensg00000180843	ALO17
ensg00000117298	ECE1	ensg00000181163	NPM
ensg00000117335	MCP	ensg00000181690	PLAG1
ensg00000117400	TPOR	ensg00000182158	CR3L2
ensg00000117425	PTC2	ensg00000182185	RA51B
ensg00000117560	TNFL6	ensg00000182197	EXT1
ensg00000117586	TNFL4	ensg00000182636	NECD
ensg00000117632	STMN1	ensg00000182712	MTCPA
ensg00000117984	CATD	ensg00000182866	LCK
ensg00000118046	STK11	ensg00000182871	COIA1
ensg00000118058	HRX	ensg00000182944	EWS
ensg00000118260	CREB1	ensg00000182985	CADM1
ensg00000118689	FOXO3	ensg00000182986	ZN320
ensg00000118939	UCHL3	ensg00000183072	NKX25
ensg00000118971	CCND2	ensg00000183117	CSMD1
ensg00000118972	FGF23	ensg00000183161	FANCF
ensg00000119121	TRPM6	ensg00000183214	MICA
ensg00000119139	ZO2	ensg00000183628	DGCR6
ensg00000119335	SET	ensg00000183722	LHFP
ensg00000119397	CP110	ensg00000183765	CHK2
ensg00000119508	NR4A3	ensg00000184012	TMPS2
ensg00000119535	CSF3R	ensg00000184258	CDR1
ensg00000119537	KDSR	ensg00000184292	TACD2
ensg00000119684	MLH3	ensg00000184384	MAML2
ensg00000119866	BC11A	ensg00000184402	S18L1
ensg00000119899	S17A5	ensg00000184481	FOXO4
ensg00000119950	MXI1	ensg00000184489	TP4A3
ensg00000120008	BRWD2	ensg00000184507	NUT
ensg00000120659	TNF11	ensg00000184640	SEPT9
ensg00000120889	TR10B	ensg00000184675	F123B
ensg00000120942	UBIA1	ensg00000184702	SEPT5

ensg00000121454	LHX4	ensg00000184916	JAG2
ensg00000121741	ZMYM2	ensg00000184937	WT1
ensg00000121879	PK3CA	ensg00000185068	TF2H5
ensg00000122025	FLT3	ensg00000185275	CD24
ensg00000122194	PLMN	ensg00000185345	PRKN2
ensg00000122507	PTHB1	ensg00000185515	BRCC3
ensg00000122512	PMS2	ensg00000185630	PBX1
ensg00000122566	ROA2	ensg00000185811	IKZF1
ensg00000122779	TIF1A	ensg00000185862	EVI2B
ensg00000123080	CDN2C	ensg00000185920	PTC1
ensg00000123219	CENPK	ensg00000186051	TAL2
ensg00000123268	ATF1	ensg00000186340	TSP2
ensg00000123364	HXC13	ensg00000186575	MERL
ensg00000123388	HXC11	ensg00000186660	ZFP91
ensg00000123473	STIL	ensg00000186716	BCR
ensg00000124243	BCAS4	ensg00000186831	K1C17
ensg00000124529	H4	ensg00000186832	K1C16
ensg00000124713	GNMT	ensg00000187140	FOXD3
ensg00000124795	DEK	ensg00000187239	FNBP1
ensg00000124813	RUNX2	ensg00000187266	EPOR
ensg00000125347	IRF1	ensg00000187323	DCC
ensg00000125354	SEPT6	ensg00000187398	LUZP2
ensg00000125378	BMP4	ensg00000187621	TNG2
ensg00000125454	DNC	ensg00000187735	TCEA1
ensg00000125618	PAX8	ensg00000187736	NHEJ1
ensg00000126215	XRCC3	ensg00000187741	FANCA
ensg00000126233	SLUR1	ensg00000187754	SSX2
ensg00000126524	SBDS	ensg00000187908	DMBT1
ensg00000126583	KPCG	ensg00000188153	CO4A5
ensg00000126746	ZN384	ensg00000188641	DPYD
ensg00000126752	SSX1	ensg00000188987	H4
ensg00000126777	KTN1	ensg00000189067	LITAF
ensg00000126778	SIX1	ensg00000189143	CLD4
ensg00000126860	EVI2A	ensg00000189283	FHIT
ensg00000126883	NU214	ensg00000196092	PAX5
ensg00000126934	MP2K2	ensg00000196176	H4
ensg00000127083	OMD	ensg00000196526	AFAP1
ensg00000127152	BC11B	ensg00000196531	NACP1;NACA
ensg00000127914	AKAP9	ensg00000196549	NEP
ensg00000127946	HIP1	ensg00000196588	MKL1

ensg00000127947	PTN12	ensg00000196712	NF1
ensg00000128487	SPEC1	ensg00000196914	ARHGC
ensg00000128512	DOCK4	ensg00000197061	H4
ensg00000128591	FLNC	ensg00000197275	RA54B
ensg00000128602	SMO	ensg00000197299	BLM
ensg00000128656	CHIN	ensg00000197323	TRI33
ensg00000128713	HXD11	ensg00000197565	CO4A6
ensg00000128714	HXD13	ensg00000197579	TOPRS
ensg00000129204	UBP6	ensg00000197594	ENPP1
ensg00000129422	MTUS1	ensg00000197746	SAP
ensg00000129757	CDN1C	ensg00000197880	MDS2
ensg00000129993	MTG16	ensg00000198400	NTRK1
ensg00000130288	NDUAD	ensg00000198467	TPM2
ensg00000130368	MAS	ensg00000198518	H4
ensg00000130382	ENL	ensg00000198553	KCNRG
ensg00000130396	AFAD	ensg00000198561	CTND1
ensg00000130402	ACTN4	ensg00000198712	COX2
ensg00000130675	MNX1	ensg00000198795	ZN521
ensg00000130826	DKC1	ensg00000198807	PAX9
ensg00000130844	ZN331	ensg00000198900	TOP1
ensg00000131016	AKA12	ensg00000198946	SSX4
ensg00000131759	RARA	ensg00000204103	MAFB
ensg00000131899	L2GL1	ensg00000204370	DHSD
ensg00000132155	RAF1	ensg00000204496	TNFB
ensg00000132170	PPARG	ensg00000204531	PO5L4;PO5F1
ensg00000132434	LANC2	ensg00000204645	SSX4
ensg00000132781	MUTYH	ensg00000204764	RBP17
ensg00000133020	MYH8	ensg00000204977	TRI13
ensg00000133027	PEMT	ensg00000205056	CLLU1
ensg00000133105	RXFP2	ensg00000205094	DUX4
ensg00000133116	KLOT	ensg00000205095	DUX4
ensg00000133216	EPHB2	ensg00000205336	GPR56
ensg00000133392	MYH11	ensg00000205413	SAMD9
ensg00000133401	PDZD2	ensg00000205927	OLIG2
ensg00000133454	MY18B	ensg00000206115	MDS1
ensg00000133639	BTG1	ensg00000206308	2DRA
ensg00000133703	RASK	ensg00000206349	PO5L4;PO5F1
ensg00000133818	RRAS2	ensg00000206454	PO5L4;PO5F1
ensg00000133895	MEN1	ensg00000211610	KV313
ensg00000134086	VHL	ensg00000211896	IGHG1

ensg00000134323	MYCN	ensg00000212638	DUX4
ensg00000134371	CDC73	ensg00000212640	DUX4
ensg00000134508	CABL1	ensg00000212710	CTGE1
ensg00000134533	RERG	ensg00000213066	FR1OP
ensg00000134574	DDB2	ensg00000213088	DUFFY
ensg00000134757	DSG3	ensg00000213190	AF1Q
ensg00000134824	FADS2	ensg00000213231	TCL1B
ensg00000134853	PGFRA	ensg00000213281	RASN
ensg00000134899	ERCC5	ensg00000213672	SPN90
ensg00000134954	ETS1	ensg00000213689	TREX1
ensg00000134982	APC	ensg00000213705	GNAS3
ensg00000135100	HNF1A	ensg00000213859	KCD11
		ensg00000214699	NAT6
		ensg00000214827	MTCPB
		ensg00000219870	BRK1

APPENDIX B

GENETIC VARIANTS BY SUBJECT ID

Subject ID	Gene	Location of Variant	Variant	dbSNP rs#	%ESP
3	AFF1	chr4:88047328	c.1544C>T p.P515L	rs144038967	0.32%
3	CISH	chr3:50645081	c.734G>A p.R245Q	rs138916184	0.18%
3	DCC	chr18:50848468	c.1070A>G p.N357S	rs35691189	0.18%
3	ERCC6	chr10:50681043	c.221C>T p.T74M	rs142580756	0.05%
5	BAP1	chr3:52437309	c.1735G>A p.G579R		
5	BCR	chr22:23656786	c.2378C>G p.A793G	rs56265970	
5	BIN1	chr2:127809910	c.938C>T p.P313L	rs141119288	0.41%
5	ERCC5, BIVM-ERCC5	chr13:103525650	c.2921A>T p.D974V		
5	FCRL4	chr1:157545384	c.1478A>G p.Y493C	rs3811028	0.82%
5	NUMA1	chr11:71726658	c.580C>T p.R194W		
5	PLAG1	chr8:57079258	c.801A>C p.E267D		
5	TRPS1	chr8:116631850	c.436G>T p.D146Y		
6	CIC	chr19:42791024	c.169G>A p.E57K		
6	ETV4	chr17:41606065	c.446C>T p.A149V	rs117364926	0.09%
6	MSH3	chr5:80063896	c.2041C>T p.P681S	rs115198722	0.05%
6	PAX5	chr9:37015054	c.26G>A p.R9Q		
6	PCSK7	chr11:117097992	c.650T>C p.M217T	rs200328848	
6	SMAD3	chr15:67482825	c.644T>G p.V215G		
6	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417739	
8	ETV1	chr7:13940369	c.897G>A p.M299I		
8	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
8	TPM3	chr1:154145607	c.67A>C p.K23Q		
11	AFAP1	chr4:7787983	c.1468G>A p.G490S	rs141071297	0.09%
11	AKAP9	chr7:91659223	c.3038C>T p.S1013L	rs143565222	0.09%
11	AKAP9	chr7:91726522	c.937C>T p.R313C	rs146495719	0.05%
11	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
11	FGFR1	chr8:38275420	c.1247G>A p.R416H		
11	FLNC	chr7:128478800	c.1354G>A p.V452M	rs192163925	0.09%
11	GLI2	chr2:121712961	c.223G>A p.A75T	rs111840592	0.69%
11	ITK	chr5:156659403	c.767C>T p.T256I	rs115751349	0.09%
11	PLAG1	chr8:57079258	c.801A>C p.E267D		
11	PMS1	chr2:190742119	c.704G>A p.R235H	rs147566508	0.09%
11	PSIP1	chr9:15474104	c.734A>G p.K245R	rs143334212	0.09%
14	AFF3	chr2:100218034	c.1234G>A p.G412S		

14	BCR	chr22:23654017	c.2083G>A p.D695N		
14	CIC	chr19:42799341	c.4825T>G p.*1609G		
14	DCC	chr18:50592531	c.221A>G p.K74R	rs144623089	0.14%
14	DSG2	chr18:29101156	c.473T>G p.V158G	rs191143292	0.23%
14	MET	chr7:116415124	c.1928C>T p.P643L		
14	PLAG1	chr8:57079258	c.801A>C p.E267D		
14	TREX1	chr3:4850839	c.341C>A p.P114Q	rs72556554	0.06%
15	BCR	chr22:23654017	c.2083G>A p.D695N		
15	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
15	PER1	chr17:8051031	c.1301G>A p.R434H	rs137923123	0.06%
15	PYGL	chr14:51382637	c.563C>T p.P188L	rs143759519	0.14%
15	SMAD3	chr15:67482825	c.644T>G p.V215G		
15	TMC6	chr17:76115102	c.574C>G p.L192V	rs149053739	
15	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417739	
22	AIRE	chr21:45710706	c.17G>A p.S6N		
22	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
22	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
22	MN1	chr22:28196476	c.56G>A p.G19D	rs200030766	0.09%
22	MTUS1	chr8:17611593	c.1724A>G p.H575R	rs209569	0.41%
22	MTUS1	chr8:17513484	c.431A>G p.E144G	rs61733708	0.55%
22	PER1	chr17:8047029	c.2627C>A p.P876H		
23	ABCA3	chr16:2367764	c.875A>T p.E292V	rs149989682	0.14%
23	ACTN4	chr19:39218649	c.1744G>A p.V582M	rs141727248	0.18%
23	APTX	chr9:32987594	c.167C>A p.S56Y	rs34778324	0.60%
23	FANCA	chr16:89865598	c.773C>A p.S258Y	rs185984960	0.05%
23	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
24	BCR	chr22:23654017	c.2083G>A p.D695N		
24	BCR	chr22:23655130	c.2146A>T p.T716S		
24	BCR	chr22:23656786	c.2378C>G p.A793G		
24	FANCC	chr9:97912307	c.584A>T p.D195V	rs1800365	0.46%
24	FOXO3	chr6:108882938	c.527A>G p.K176R		
24	INSRR	chr1:156823922	c.785G>A p.S262N	rs199833934	
24	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
24	MTUS1	chr8:17510766	c.748G>C p.E250Q	rs61733705	0.50%
29	ACSL6	chr5:13130371	c.373C>G p.H125D		
29	AHI1	chr6:135639715	c.1718C>T p.S573F	rs117447608	0.23%
29	ELL	chr19:18561595	c.758C>T p.P253	rs202017480	0.18%
29	ERBB2	chr17:37879585	c.1132A>G p.I378V	rs1801201	0.27%
29	FGFR3	chr4:1806629	c.1351C>T p.P451S	rs61735104	0.73%
29	FUS	chr16:31195601	c.404A>G p.Y135C		

29	FUS	chr16:31195609	c.412C>G p.Q138E		
29	GLI2	chr2:121712955	c.217G>A p.G73R		
29	KEAP1	chr19:10602452	c.280G>A p.A94T		
29	MTUS1	chr8:17611593	c.1724A>G p.H575R	rs209569	0.41%
29	MYH9	chr22:36682873	c.4952T>C p.M1651T	rs142094977	0.09%
29	NIN	chr14:51224303	c.3445G>T p.D1149Y		
29	OSMR	chr5:38933482	c.488C>G p.P163R	rs34080825	0.41%
29	TMC6	chr17:76117124	c.422C>T p.P141	rs75400929	0.41%
31	BARD1	chr2:215674240	c.54C>G p.N18K		
31	BCR	chr22:23656786	c.2378C>G p.A793G	rs56265970	unknown
31	CCND3	chr6:41905041	c.263C>T p.S88		
31	ERBB3	chr12:56493772	c.448G>A p.A150T		
31	ERCC6	chr10:50686474	c.322A>C p.I108L		
31	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
31	PAX8	chr2:113999656	c.530G>A p.G177D		
31	PCSK7	chr11:117076843	c.2228C>A p.T743N	rs148634347	0.02%
31	TCL1A	chr14:96180363	c.41A>C p.H14P		
31	TMC6	chr17:76113988	c.653C>T p.P218	rs148239603	0.05%
38	CSMD1	chr8:2824257	c.6925G>A p.G2309S	rs117633452	0.82%
38	DIDO1	chr20:61511906	c.5402G>T p.G1801V		
38	FANCF	chr11:22646639	c.718T>A p.W240R		
38	JAG2	chr14:105609400	c.1561C>T p.R521W	rs34728766	0.50%
38	MEFV	chr16:3293403	c.341A>G p.K114R	rs104895094	0.25%
38	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417740	
44	ABCA3	chr16:2374455	c.397G>A p.V133M	rs201400258	0.05%
44	HIP1	chr7:75368208	c.31G>T p.V11		
44	HIP1	chr7:75368209	c.30G>C p.Q10H		
44	HIP1	chr7:75368210	c.29A>C p.Q10P		
44	HIP1	chr7:75368211	c.28C>T p.Q10*		
44	INSRR	chr1:156815768	c.1954T>G p.Y652D		
44	INSRR	chr1:156815771	c.1951C>G p.L651V		
44	ROBO1	chr3:78766988	c.586G>A p.A196T		
44	TMC6	chr17:76113988	c.653C>T p.P218	rs148239603	0.05%
44	TMC6	chr17:76120670	c.145G>T p.A49S		
45	AIP	chr8:67257557	c.517G>A p.E173K	rs138902236	0.05%
45	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
45	EGFR	chr7:55238182	c.2063A>C p.H688P		
45	FSCN1	chr7:5633310	c.743T>G p.V248G		
45	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
45	MTUS1	chr8:17513386	c.529C>G p.Q177E	rs150374461	0.23%

45	PDZD2	chr5:32098669	:c.8147C>A p.A2716D	rs142643164	0.23%
45	SMAD3	chr15:67482825	c.644T>G p.V215G		
45	TMC6	chr17:76113988	c.653C>T p.P218	rs148239603	0.05%
47	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
47	DMBT1	chr10:124342121	c.628G>A p.E210K		
47	FANCA	chr16:89836245	c.2504A>C p.K835T		
47	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
47	FOXO3	chr6:108882938	c.527A>G p.K176R		
47	HNF1B	chr17:36070591	c.436A>C p.T146P		
47	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
47	MN1	chr22:28196290	c.242G>C p.G81A	rs201422000	0.15%
47	MTUS1	chr8:17611381	c.1936G>C p.E646Q	rs61733691	0.27%
47	MYH8	chr17:10317540	c.977T>C p.I326T	rs34124921	
47	NAT6	chr3:50334633	c.262A>C p.T88P		
47	PCSK9	chr1:55512275	c.479G>A p.R160Q		
47	SMAD3	chr15:67482825	c.644T>G p.V215G		
48	EVI5	chr1:93089803	c.1709G>A p.R570Q	rs201719783	
48	ING1	chr13:111371966	c.320C>T p.P107	rs140897205	0.05%
48	PHF20	chr20:34535529	c.3019G>A p.A1007T	rs148822034	0.14%
50	BCR	chr22:23656786	c.2378C>G p.A793G	rs56265970	unknown
50	ELL	chr19:18557592	c.1099C>G p.P367A	rs151228060	0.09%
50	ERCC2	chr19:45855903	c.596G>A p.R199Q		
50	FANCC	chr9:98011497	c.77C>T p.S26F	rs1800361	0.32%
51	DOCK4	chr7:111462470	c.1201C>A p.R401S	rs186031092	0.09%
51	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
51	MUTYH	chr1:45797157	c.301C>A p.L101M	rs144079536	0.05%
51	PCM1	chr8:17824658	c.3391A>C p.N1131H	rs35041534	0.05%
52	BCR	chr22:23655107	c.2123T>C p.M708T		
52	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
52	JAK2	chr9:5126715	c.3323A>G p.N1108S	rs142269166	0.09%
52	LZTS1	chr8:20107433	c.1414G>A p.V472M	rs139782965	0.05%
52	MTUS1	chr8:17541943	c.167A>C p.K56T	rs61748836	0.32
52	NBN	chr8:90983460	c.397C>T p.R133W	rs34767364	0.23%
52	PYGL	chr14:51390736	c.20A>G p.Y7C	rs34096980	0.18%
53	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
53	MYC	chr8:128750680	c.172A>C p.T58P		
53	PAX5	chr9:36966688	c.314C>T p.S105	rs137870876	0.09%
53	TMC6	chr17:76113988	c.653C>T p.P218L	rs148239603	0.05%
53	TSC2	chr16:2124321	c.1321A>C p.T441P	rs45517238	0.08%
53	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417739	

54	FOXP1	chr3:71247489	c.44C>T p.A15V		
54	JAG2	chr14:105612960	c.793C>G p.R265G		
54	TMC6	chr17:76113988	c.653C>T p.P218L	rs148239603	0.05%
55	AURKB	chr17:8110585	c.184G>A p.V62M	rs150216235	0.09%
55	BRCA2	chr13:32953550	c.8851G>A p.A2951T	rs11571769	0.69%
55	EDA	chrX:68836358	c.206G>T p.R69L	rs132630309	0.12%
55	MYC	chr8:128750680	c.172A>C p.T58P		
55	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417739	
56	HIP1	chr7:75172179	c.2728G>A p.E910K	rs140538287	0.05%
56	PER1	chr17:8049320	c.2126G>C p.S709T		
57	ABL1	chr9:133730200	c.266G>A p.R89Q		
57	AFF4	chr5:132270228	c.529C>T p.R177C		
57	BCR	chr22:23654017	c.2083G>A p.D695N		
57	BCR	chr22:23656786	c.2378C>G p.A793G	rs56265970	
57	CASC5	chr15:40916632	c.3720T>A p.D1240E	rs141726041	0.92%
57	FANCC	chr9:97887391	c.973G>A p.A325T	rs201407189	0.18%
57	FCRL4	chr1:157545384	c.1478A>G p.Y493C	rs3811028	0.82%
57	PAX8	chr2:113992993	c.986C>T p.S329L		
58	EPHB2	chr1:23233346	c.1858G>A p.D620N	rs28936395	0.23%
58	LZTS1	chr8:20110406	c.1036C>G p.R346G	rs148775156	0.18%
58	MXI1	chr10:111967693	c.127G>A p.A43T	rs141964073	0.09%
58	NFKB2	chr10:104155730	c.14A>G p.Y5C	rs200361192	0.05%
58	NTRK1	chr1:156845431	c.571G>A p.E191K	rs144901788	0.10%
58	TCL1B	chr14:96152818	c.14C>T p.A5V		
59	APC	chr5:112176317	c.2926A>G p.R976G		
59	BRCA2	chr13:32953550	c.8851G>A p.A2951T	rs11571769	0.69%
59	LIFR	chr5:38493789	c.1984T>C p.C662R		
59	PDZD2	chr5:32074181	c.2447C>T p.P816L	rs139754344	0.60%
59	PMS2	chr7:6045573	c.113C>T p.A38V	rs148270248	0.01%
59	ST14	chr11:130058112	c.185T>C p.L62P		
62	AKAP9	chr7:91630913	c.557C>T p.A186V	rs149162273	0.02%
62	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
62	DMBT1	chr10:124342121	c.628G>A p.E210K		
62	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
62	JAG2	chr14:105609400	c.1561C>T p.R521W	rs34728766	0.50%
62	MSH3	chr5:79966040	c.704A>G p.Q235R		
63	DIDO1	chr20:61511906	c.5402G>T p.G1801V		
63	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
63	DMBT1	chr10:124342121	c.628G>A p.E210K		
63	FANCF	chr11:22646639	c.718T>A p.W240R		

63	HNF1B	chr17:36070591	c.436A>C p.T146P		
63	JAG2	chr14:105609400	c.1561C>T p.R521W	rs34728766	0.50%
63	JAK2	chr9:5072561	c.1711G>A p.G571S	rs139504737	
63	MSH2	chr2:47641544	c.650T>C p.L217P	rs63750640	
63	MYC	chr8:128750680	c.172A>C p.T58P		
63	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417741	
64	BCR	chr22:23654017	c.2083G>A p.D695N		
64	CSMD1	chr8:2824257	c.6925G>A p.G2309S	rs117633452	0.82%
64	DIDO1	chr20:61511906	c.5402G>T p.G1801V		
64	FANCF	chr11:22646639	c.718T>A p.W240R		
64	HNF1B	chr17:36070591	c.436A>C p.T146P		
64	JAK2	chr9:5072561	c.1711G>A p.G571S	rs139504737	
64	MEFV	chr16:3293403	c.341A>G p.K114R	rs104895094	0.25%
64	PDGFB	chr22:39631786	c.112G>A p.G38R	rs148252800	0.09%
64	TMC8	chr17:76129619	c.664C>T p.R222W	rs117156381	0.18%
67	BCR	chr22:23654017	c.2083G>A p.D695N		
67	RERE	chr1:8419873	c.1906_1907AG		
67	RERE	chr1:8419874	c.1906A>C p.K636Q		
67	SCN91	chr2:167089942	c.3799C>G p.L1267V	rs180922748	0.14%

APPENDIX C

GENETIC VARIANTS BY GENE NAME

Subject ID	Gene	Location of Variant	Variant	dbSNP rs#	%ESP
23	ABCA3	chr16:2367764	c.875A>T p.E292V	rs149989682	0.14%
44	ABCA3	chr16:2374455	c.397G>A p.V133M	rs201400258	0.05%
57	ABL1	chr9:133730200	c.266G>A p.R89Q		
29	ACSL6	chr5:13130371	c.373C>G p.H125D		
23	ACTN4	chr19:39218649	c.1744G>A p.V582M	rs141727248	0.18%
11	AFAP1	chr4:7787983	c.1468G>A p.G490S	rs141071297	0.09%
3	AFF1	chr4:88047328	c.1544C>T p.P515L	rs144038967	0.32%
14	AFF3	chr2:100218034	c.1234G>A p.G412S		
57	AFF4	chr5:132270228	c.529C>T p.R177C		
29	AHI1	chr6:135639715	c.1718C>T p.S573F	rs117447608	0.23%
45	AIP	chr8:67257557	c.517G>A p.E173K	rs138902236	0.05%
22	AIRE	chr21:45710706	c.17G>A p.S6N		
11	AKAP9	chr7:91659223	c.3038C>T p.S1013L	rs143565222	0.09%
11	AKAP9	chr7:91726522	c.937C>T p.R313C	rs146495719	0.05%
62	AKAP9	chr7:91630913	c.557C>T p.A186V	rs149162273	0.02%
59	APC	chr5:112176317	c.2926A>G p.R976G		
23	APTX	chr9:32987594	c.167C>A p.S56Y	rs34778324	0.60%
55	AURKB	chr17:8110585	c.184G>A p.V62M	rs150216235	0.09%
5	BAP1	chr3:52437309	c.1735G>A p.G579R		
31	BARD1	chr2:215674240	c.54C>G p.N18K		
5	BCR	chr22:23656786	c.2378C>G p.A793G	rs56265970	
14	BCR	chr22:23654017	c.2083G>A p.D695N		
15	BCR	chr22:23654017	c.2083G>A p.D695N		
24	BCR	chr22:23654017	c.2083G>A p.D695N		
24	BCR	chr22:23655130	c.2146A>T p.T716S		
24	BCR	chr22:23656786	c.2378C>G p.A793G		
31	BCR	chr22:23656786	c.2378C>G p.A793G	rs56265970	unknown
50	BCR	chr22:23656786	c.2378C>G p.A793G	rs56265970	unknown
52	BCR	chr22:23655107	c.2123T>C p.M708T		
57	BCR	chr22:23654017	c.2083G>A p.D695N		
57	BCR	chr22:23656786	c.2378C>G p.A793G	rs56265970	
64	BCR	chr22:23654017	c.2083G>A p.D695N		
67	BCR	chr22:23654017	c.2083G>A p.D695N		
5	BIN1	chr2:127809910	c.938C>T p.P313L	rs141119288	0.41%
55	BRCA2	chr13:32953550	c.8851G>A p.A2951T	rs11571769	0.69%

59	BRCA2	chr13:32953550	c.8851G>A p.A2951T	rs11571769	0.69%
57	CASC5	chr15:40916632	c.3720T>A p.D1240E	rs141726041	0.92%
31	CCND3	chr6:41905041	c.263C>T p.S88		
6	CIC	chr19:42791024	c.169G>A p.E57K		
14	CIC	chr19:42799341	c.4825T>G p.*1609G		
3	CISH	chr3:50645081	c.734G>A p.R245Q	rs138916184	0.18%
38	CSMD1	chr8:2824257	c.6925G>A p.G2309S	rs117633452	0.82%
64	CSMD1	chr8:2824257	c.6925G>A p.G2309S	rs117633452	0.82%
3	DCC	chr18:50848468	c.1070A>G p.N357S	rs35691189	0.18%
14	DCC	chr18:50592531	c.221A>G p.K74R	rs144623089	0.14%
38	DIDO1	chr20:61511906	c.5402G>T p.G1801V		
63	DIDO1	chr20:61511906	c.5402G>T p.G1801V		
64	DIDO1	chr20:61511906	c.5402G>T p.G1801V		
11	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
45	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
47	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
47	DMBT1	chr10:124342121	c.628G>A p.E210K		
62	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
62	DMBT1	chr10:124342121	c.628G>A p.E210K		
63	DMBT1	chr10:124380869	c.136C>A p.H46N	rs201533269	
63	DMBT1	chr10:124342121	c.628G>A p.E210K		
51	DOCK4	chr7:111462470	c.1201C>A p.R401S	rs186031092	0.09%
14	DSG2	chr18:29101156	c.473T>G p.V158G	rs191143292	0.23%
55	EDA	chrX:68836358	c.206G>T p.R69L	rs132630309	0.12%
45	EGFR	chr7:55238182	c.2063A>C p.H688P		
29	ELL	chr19:18561595	c.758C>T p.P253	rs202017480	0.18%
50	ELL	chr19:18557592	c.1099C>G p.P367A	rs151228060	0.09%
58	EPHB2	chr1:23233346	c.1858G>A p.D620N	rs28936395	0.23%
29	ERBB2	chr17:37879585	c.1132A>G p.I378V	rs1801201	0.27%
31	ERBB3	chr12:56493772	c.448G>A p.A150T		
50	ERCC2	chr19:45855903	c.596G>A p.R199Q		
5	ERCC5, BIVM-ERCC5	chr13:103525650	c.2921A>T p.D974V		
3	ERCC6	chr10:50681043	c.221C>T p.T74M	rs142580756	0.05%
31	ERCC6	chr10:50686474	c.322A>C p.I108L		
8	ETV1	chr7:13940369	c.897G>A p.M299I		
6	ETV4	chr17:41606065	c.446C>T p.A149V	rs117364926	0.09%
48	EVI5	chr1:93089803	c.1709G>A p.R570Q	rs201719783	
23	FANCA	chr16:89865598	c.773C>A p.S258Y	rs185984960	0.05%
47	FANCA	chr16:89836245	c.2504A>C p.K835T		

24	FANCC	chr9:97912307	c.584A>T p.D195V	rs1800365	0.46%
50	FANCC	chr9:98011497	c.77C>T p.S26F	rs1800361	0.32%
57	FANCC	chr9:97887391	c.973G>A p.A325T	rs201407189	0.18%
38	FANCF	chr11:22646639	c.718T>A p.W240R		
63	FANCF	chr11:22646639	c.718T>A p.W240R		
64	FANCF	chr11:22646639	c.718T>A p.W240R		
5	FCRL4	chr1:157545384	c.1478A>G p.Y493C	rs3811028	0.82%
57	FCRL4	chr1:157545384	c.1478A>G p.Y493C	rs3811028	0.82%
11	FGFR1	chr8:38275420	c.1247G>A p.R416H		
29	FGFR3	chr4:1806629	c.1351C>T p.P451S	rs61735104	0.73%
11	FLNC	chr7:128478800	c.1354G>A p.V452M	rs192163925	0.09%
15	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
22	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
24	FOXO3	chr6:108882938	c.527A>G p.K176R		
47	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
47	FOXO3	chr6:108882938	c.527A>G p.K176R		
51	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
52	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
53	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
62	FOXO3	chr6:108985481	c.785G>A p.S262N	rs199833934	
54	FOXP1	chr3:71247489	c.44C>T p.A15V		
45	FSCN1	chr7:5633310	c.743T>G p.V248G		
29	FUS	chr16:31195601	c.404A>G p.Y135C		
29	FUS	chr16:31195609	c.412C>G p.Q138E		
11	GLI2	chr2:121712961	c.223G>A p.A75T	rs111840592	0.69%
29	GLI2	chr2:121712955	c.217G>A p.G73R		
44	HIP1	chr7:75368208	c.31G>T p.V11		
44	HIP1	chr7:75368209	c.30G>C p.Q10H		
44	HIP1	chr7:75368210	c.29A>C p.Q10P		
44	HIP1	chr7:75368211	c.28C>T p.Q10*		
56	HIP1	chr7:75172179	c.2728G>A p.E910K	rs140538287	0.05%
47	HNF1B	chr17:36070591	c.436A>C p.T146P		
63	HNF1B	chr17:36070591	c.436A>C p.T146P		
64	HNF1B	chr17:36070591	c.436A>C p.T146P		
48	ING1	chr13:111371966	c.320C>T p.P107	rs140897205	0.05%
24	INSRR	chr1:156823922	c.785G>A p.S262N	rs199833934	
44	INSRR	chr1:156815768	c.1954T>G p.Y652D		
44	INSRR	chr1:156815771	c.1951C>G p.L651V		
11	ITK	chr5:156659403	c.767C>T p.T256I	rs115751349	0.09%
38	JAG2	chr14:105609400	c.1561C>T p.R521W	rs34728766	0.50%

54	JAG2	chr14:105612960	c.793C>G p.R265G		
62	JAG2	chr14:105609400	c.1561C>T p.R521W	rs34728766	0.50%
63	JAG2	chr14:105609400	c.1561C>T p.R521W	rs34728766	0.50%
52	JAK2	chr9:5126715	c.3323A>G p.N1108S	rs142269166	0.09%
63	JAK2	chr9:5072561	c.1711G>A p.G571S	rs139504737	
64	JAK2	chr9:5072561	c.1711G>A p.G571S	rs139504737	
29	KEAP1	chr19:10602452	c.280G>A p.A94T		
8	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
22	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
23	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
24	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
31	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
45	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
47	LHX4	chr1:180240971	c.608T>G p.V203G	rs201275928	
59	LIFR	chr5:38493789	c.1984T>C p.C662R		
52	LZTS1	chr8:20107433	c.1414G>A p.V472M	rs139782965	0.05%
58	LZTS1	chr8:20110406	c.1036C>G p.R346G	rs148775156	0.18%
38	MEFV	chr16:3293403	c.341A>G p.K114R	rs104895094	0.25%
64	MEFV	chr16:3293403	c.341A>G p.K114R	rs104895094	0.25%
14	MET	chr7:116415124	c.1928C>T p.P643L		
22	MN1	chr22:28196476	c.56G>A p.G19D	rs200030766	0.09%
47	MN1	chr22:28196290	c.242G>C p.G81A	rs201422000	0.15%
63	MSH2	chr2:47641544	c.650T>C p.L217P	rs63750640	
6	MSH3	chr5:80063896	c.2041C>T p.P681S	rs115198722	0.05%
62	MSH3	chr5:79966040	c.704A>G p.Q235R		
22	MTUS1	chr8:17611593	c.1724A>G p.H575R	rs209569	0.41%
22	MTUS1	chr8:17513484	c.431A>G p.E144G	rs61733708	0.55%
24	MTUS1	chr8:17510766	c.748G>C p.E250Q	rs61733705	0.50%
29	MTUS1	chr8:17611593	c.1724A>G p.H575R	rs209569	0.41%
45	MTUS1	chr8:17513386	c.529C>G p.Q177E	rs150374461	0.23%
47	MTUS1	chr8:17611381	c.1936G>C p.E646Q	rs61733691	0.27%
52	MTUS1	chr8:17541943	c.167A>C p.K56T	rs61748836	0.32
51	MUTYH	chr1:45797157	c.301C>A p.L101M	rs144079536	0.05%
58	MXI1	chr10:111967693	c.127G>A p.A43T	rs141964073	0.09%
53	MYC	chr8:128750680	c.172A>C p.T58P		
55	MYC	chr8:128750680	c.172A>C p.T58P		
63	MYC	chr8:128750680	c.172A>C p.T58P		
47	MYH8	chr17:10317540	c.977T>C p.I326T	rs34124921	
29	MYH9	chr22:36682873	c.4952T>C p.M1651T	rs142094977	0.09%
47	NAT6	chr3:50334633	c.262A>C p.T88P		

52	NBN	chr8:90983460	c.397C>T p.R133W	rs34767364	0.23%
58	NFKB2	chr10:104155730	c.14A>G p.Y5C	rs200361192	0.05%
29	NIN	chr14:51224303	c.3445G>T p.D1149Y		
58	NTRK1	chr1:156845431	c.571G>A p.E191K	rs144901788	0.10%
5	NUMA1	chr11:71726658	c.580C>T p.R194W		
29	OSMR	chr5:38933482	c.488C>G p.P163R	rs34080825	0.41%
6	PAX5	chr9:37015054	c.26G>A p.R9Q		
53	PAX5	chr9:36966688	c.314C>T p.S105	rs137870876	0.09%
31	PAX8	chr2:113999656	c.530G>A p.G177D		
57	PAX8	chr2:113992993	c.986C>T p.S329L		
51	PCM1	chr8:17824658	c.3391A>C p.N1131H	rs35041534	0.05%
6	PCSK7	chr11:117097992	c.650T>C p.M217T	rs200328848	
31	PCSK7	chr11:117076843	c.2228C>A p.T743N	rs148634347	0.02%
47	PCSK9	chr1:55512275	c.479G>A p.R160Q		
64	PDGFB	chr22:39631786	c.112G>A p.G38R	rs148252800	0.09%
45	PDZD2	chr5:32098669	:c.8147C>A p.A2716D	rs142643164	0.23%
59	PDZD2	chr5:32074181	c.2447C>T p.P816L	rs139754344	0.60%
15	PER1	chr17:8051031	c.1301G>A p.R434H	rs137923123	0.06%
22	PER1	chr17:8047029	c.2627C>A p.P876H		
56	PER1	chr17:8049320	c.2126G>C p.S709T		
48	PHF20	chr20:34535529	c.3019G>A p.A1007T	rs148822034	0.14%
5	PLAG1	chr8:57079258	c.801A>C p.E267D		
11	PLAG1	chr8:57079258	c.801A>C p.E267D		
14	PLAG1	chr8:57079258	c.801A>C p.E267D		
11	PMS1	chr2:190742119	c.704G>A p.R235H	rs147566508	0.09%
59	PMS2	chr7:6045573	c.113C>T p.A38V	rs148270248	0.01%
11	PSIP1	chr9:15474104	c.734A>G p.K245R	rs143334212	0.09%
15	PYGL	chr14:51382637	c.563C>T p.P188L	rs143759519	0.14%
52	PYGL	chr14:51390736	c.20A>G p.Y7C	rs34096980	0.18%
67	RERE	chr1:8419873	c.1906_1907AG		
67	RERE	chr1:8419874	c.1906A>C p.K636Q		
44	ROBO1	chr3:78766988	c.586G>A p.A196T		
67	SCN91	chr2:167089942	c.3799C>G p.L1267V	rs180922748	0.14%
6	SMAD3	chr15:67482825	c.644T>G p.V215G		
15	SMAD3	chr15:67482825	c.644T>G p.V215G		
45	SMAD3	chr15:67482825	c.644T>G p.V215G		
47	SMAD3	chr15:67482825	c.644T>G p.V215G		
59	ST14	chr11:130058112	c.185T>C p.L62P		
31	TCL1A	chr14:96180363	c.41A>C p.H14P		
58	TCL1B	chr14:96152818	c.14C>T p.A5V		

15	TMC6	chr17:76115102	c.574C>G p.L192V	rs149053739	
29	TMC6	chr17:76117124	c.422C>T p.P141	rs75400929	0.41%
31	TMC6	chr17:76113988	c.653C>T p.P218	rs148239603	0.05%
44	TMC6	chr17:76113988	c.653C>T p.P218	rs148239603	0.05%
44	TMC6	chr17:76120670	c.145G>T p.A49S		
45	TMC6	chr17:76113988	c.653C>T p.P218	rs148239603	0.05%
53	TMC6	chr17:76113988	c.653C>T p.P218L	rs148239603	0.05%
54	TMC6	chr17:76113988	c.653C>T p.P218L	rs148239603	0.05%
64	TMC8	chr17:76129619	c.664C>T p.R222W	rs117156381	0.18%
8	TPM3	chr1:154145607	c.67A>C p.K23Q		
14	TREX1	chr3:4850839	c.341C>A p.P114Q	rs72556554	0.06%
5	TRPS1	chr8:116631850	c.436G>T p.D146Y		
53	TSC2	chr16:2124321	c.1321A>C p.T441P	rs45517238	0.08%
6	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417739	
15	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417739	
38	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417740	
53	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417739	
55	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417739	
63	TSHR	chr14:81609723	c.1321A>C p.T441P	rs201417741	