Boston University

OpenBU **http://open.bu.edu**

Boston University Theses & Dissertations Boston University Theses & Dissertations

2016

Integrated genomic and transcriptomic analyses of radiation-induced malignancies

https://hdl.handle.net/2144/16992

"Downloaded from OpenBU. Boston University's institutional repository."

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

INTEGRATED GENOMIC AND TRANSCRIPTOMIC ANALYSES OF RADIATION-INDUCED MALIGNANCIES

by

YONG EUN LEE

B.S., University of California, Berkeley, 2012

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2016

© 2016 by
YONG EUN LEE
All rights reserved

Approved by

First Reader _____

R. J. Rushmore III, Ph.D.
Assistant Professor of Anatomy & Neurobiology

Second Reader _____

Jean L. Nakamura, M.D.
Associate Professor of Radiation Oncology
University of California, San Francisco

DEDICATION

I would like to dedicate this to the Nakamura lab members.

ACKNOWLEDGMENTS

I would like to thank Dr. Jean Nakamura and the wonderful members of the Nakamura Lab for their guidance.

I would like to also thank Drs. R. Jarret Rushmore and Gwynneth Offner for their guidance during my time in the MAMS program.

INTEGRATED GENOMIC AND TRANSCRIPTOMIC ANALYSES OF RADIATION-INDUCED MALIGNANCIES

YONG EUN LEE

ABSTRACT

Cancer is a genetic disease caused by an unregulated expansion of a clone of cells (Sompayrac, 2004). The genetic abnormalities in cancer are the consequences of defective DNA replication, repair, maintenance, and modification, genetic background, and exposure to mutagens (Alexandrov et al., 2013).

Ionizing radiation (IR), a mutagen exposed to cancer patients during clinical radiotherapy (RT), can cause DNA damage, genomic instability, and mutagenesis (Sherborne et al., 2015). While RT has been effective in treating cancer, it increases the risk of second malignant neoplasm (SMN), a severe delayed complication associated with mainly pediatric cancer survivors many decades after the treatment of their first cancer (Robison & Hudson, 2014). As the mortality of patients with childhood cancer has been decreasing, cases of radiation-induced cancers has been increasing (Robison & Hudson, 2014). The considerable contribution by RT to SMN risk illustrate the need to characterize the genetic mechanism directly responsible for radiation-induced malignancies.

To better our understanding of the mutational landscape of SMNs, our specific aims are to identify potential driver mutations implicated in radiation-induced malignancies through genome and transcriptome analysis and to assess

whether genetic background, specifically germline polymorphisms and mutations in tumor suppressor gene *TP53*, has an impact on the formation of secondary malignancies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CCSS.....Childhood Cancer Survivor Study

cnLOH .....Loss of Heterogeneity

HER2.....Human Epidemeral Growth Factor Receptor-2

IR..... Ionizing Radiation

PCR..... Polymerase Chain Reaction

RT.....Radiation Therapy

SMN.....Second Malignant Neoplasm

SNV .....Single Nucleotide Variant

UV..... Ultraviolet

WES .....Whole Exome Sequencing

INTRODUCTION

Cancer is a major public health problem in United States. According to DeSantis *et al.*, about 589,430 Americans died in 2015 from cancer, listing as the second leading cause of death following heart disease. The most common cancers are lung and bronchus, prostate, and colorectum amongst men, and lung and bronchus, breast, and colorectum in women (DeSantis et al., 2014). While men are more likely to be diagnosed with cancer than women in their lifetime, cancer risk is higher for women than men under the age of 50 (Siegel, Miller, & Jemal, 2015). Incidence of childhood and adolescent cancer makes 1% of cancer diagnoses; leukemia and cancers of the central nervous system accounts for 50% of these cases (Robison & Hudson, 2014). Because cancer is very diverse with over 100 identified types with different causes and symptoms (Almeida & Barry, 2011), the disease is difficult to cure. A tumor can form in any tissue type; while many commonalities exist, each cancer is unique (Almeida & Barry, 2011).

In the most basic sense, cancer can be defined as a disease in which a group of unregulated cells grow abnormally, ignoring all the rules of cell division (Hejmadi, 2009). The normal cells transform into cancer cells that disrupt the molecular networks that control cell proliferation, differentiation, and death(Hejmadi, 2009). When this proliferation persists, the tumor grows and metastasizes, which makes up 90% of cancer-related deaths (Hejmadi, 2009).

Cancer pathogenesis is caused by somatic mutations that allow a clone of cells to grow in an unregulated fashion (Stratton, 2011). Somatic mutations include base substitutions, insertions and deletions (indels), and structural variations such as copy-number variations, translocations, inversions, and copy-neutral regions of loss of heterogeneity (cnLOH) (Stratton, 2011). While somatic mutations exist in all cells, a subset called driver mutations contribute to the cancer phenotype by altering tumor suppressor genes and oncogenes (also known as recessive and dominant cancer genes, respectively) that normally regulate cell growth, proliferation, differentiation, and death (Stratton, 2011). An individual's likelihood of acquiring driver mutations depends on exogenous and endogenous mutagenic exposures and predisposed germline mutations (Alexandrov et al., 2013). For example, exposure to cigarette smoke and ultraviolet radiation are major risk factors of lung cancer and skin cancer, respectively (International Agency for Research on Cancer Working Group on artificial ultraviolet (UV) light and skin cancer, 2007; Pirozynski, 2006), and overexpression of human epidermal growth factor receptor-2 (HER2), an oncogene, is associated with poor prognosis in breast cancer (Yarden, 2001). Both lifestyle and inherent factors could have an impact on cancer susceptibility.

In order to better understand the biological processes of cancer, it is important to identify driver mutations that contribute to carcinogenesis (Mwenifumbo & Marra, 2013). Discovering common genetic aberrations could lead to revolutionizing the understanding of cancer biology and improving

targeted therapy strategies. For example, identifying *BRAF* V600E, an activating mutation prevalent in 50-70% of all melanomas, has shed light on its underlying molecular mechanisms in the constitutive activation of MAPK signaling and has allowed for the development of *BRAF* pharmaceutical inhibitors (Fang, Hutchinson, Deng, & Green, 2016; Whipple & Brinckerhoff, 2014). Identification of other recurrent genetic alterations, such as in-frame deletion of EGFR exon 19 E746-A750 in non-small cell lung cancer (Cooper, Lam, O'Toole, & Minna, 2013) and chromosomal gain at 3q in cervical squamous cell carcinoma (Thomas et al., 2014), have also reaped benefits in gaining insight into their genetic mechanisms and improving the efficacy of cancer treatment (Pfeifer & Besaratinia, 2009).

While somatic mutations of each type of cancer are highly heterogeneous despite the same tissue of origin, next-generation sequencing has given the means to discover somatic mutations that influence pathways and deregulate cell and tissue homeostasis (Mwenifumbo & Marra, 2013). Expanding the catalogue of cancer-related DNA sequence changes will lead to deciphering the mutational landscape of different types of cancer (Pfeifer & Besaratinia, 2009). As more significant driver mutations and cancer genes get catalogued and identified, characterization of cancer subtypes and personalization of cancer therapy will greatly advance in the next five to ten years (Mwenifumbo & Marra, 2013).

Assessing the Risk of Second Malignant Neoplasms

Although the incidence of childhood cancer is extremely low, the prevalence of pediatric cancer survivors has increased over the past 30 years (Robison & Hudson, 2014). While the increase of survivors reflects the improved efficacy of childhood cancer therapy, survivors are still at high risk of second malignant neoplasms (SMNs) due to the harsh toxic treatments used to treat their first cancer (Robison & Hudson, 2014). Radiation therapy (RT), which uses focal, fractionated ionizing radiation (IR), a known mutagen, is often associated with SMNs (Sherborne et al., 2015). However, both the genetic consequences of IR exposure from RT and the impact of genetic background on SMN risk are poorly characterized.

In order to identify significant genetic events related to treatment-induced SMNs, we identified single nucleotide variants (SNVs) from SMNs of two pediatric cancer survivors by performing whole exome sequencing (WES) and comparing the differences between their sarcoma DNA to their corresponding germline control DNA. From a comprehensive list of somatic variants found in patient tumor samples via WES, we were able to further analyze and identify potentially deleterious driver mutations by using predictor programs available online and categorizing each potential driver mutation based on relative risk.

Furthermore, to characterize the influence of genetic background on IR-induced tumorigenesis, we established a validation cohort of childhood cancer survivors who were diagnosed with subsequent treatment-induced cancers.

Because germline mutations in tumor suppressor gene *TP53* are known to be indicative of genetic susceptibility to SMNs (Malkin et al., 1992), Sanger sequencing was performed on the germline of these patients to evaluate whether there is a higher prevalence in inherent polymorphisms of *TP53* among radiation-induced SMN patients. Our validation cohort indicated that there is an increase in the recurrence of germline *TP53* polymorphisms in both the proline-rich domain and DNA binding domain of the gene compared to the population. These observations suggest screening for germline mutations in *TP53* in pediatric cancer survivors could facilitate in identifying their genetic susceptibility to developing SMNs and possibly modifying therapies for individuals who harbor this alteration, especially to those who were unaware of their predisposition.

METHODS

Whole Exome Sequencing

Whole exome sequencing was performed using the NimbleGen Human exome v3.0 kit. Captured material was indexed and sequenced on the Illumina GAII and HiSeq2000 platform at the Institute for Human Genetics at UCSF. Successfully sequenced reads were then mapped to the human reference genome (GRCh37) using GATK best practices. MuTect was used for somatic mutation detection. Annovar was used for variant annotation and to filter known human mutations to produce a cleaner list of variants. Mutation Assessor, SIFT, and PROVEAN were used to assign functional predictions. Validation of a subset of somatic variants by Sanger sequencing confirmed 92% of SNVs (23 of 25 tested).

Germline TP53 analysis of pediatric cancer survivors with subsequent SMNs

To determine if there is a higher rate of polymorphisms of *TP53* in pediatric cancer survivors with subsequent SMNs, we selected a validation cohort of 41 patients that are registered in the Childhood Cancer Survivor Study (CCSS), all of whom are pediatric cancer survivors with radiation-induced second cancers. The validation cohort consisted of patients who were initially treated of cancers such as Hodgkin lymphoma, sarcomas, and CNS tumors and developed SMNs following radiation therapy such as infiltrating ductal carcinoma of the breast and sarcomas. Furthermore, none of these survivors are known to be suffering from any known a hereditary tumor predisposition syndrome, such as Li-Fraumeni, Neurofibromatosis I, Tuberous Sclerosis, Neurofibromatosis 2, and Ataxia Telangiectasia).

Polymerase chain reaction (PCR) was performed on the validation cohort to prepare for Sanger sequencing. Sequencing primers, as shown in Table 1, were used for PCR to screen 41 patients for variants in exons 2-11 of *TP53*. All of these primers were designed using *Primer3* and checked using *Blat* and *In Silico PCR* from the UCSC Genome Bioinformatics site.

Table 1. List of *TP53* sequencing primers. Primers listed were used to perform PCR for Sanger sequencing on exons 2-11 of *TP53*. F (Forward); R(Reverse). All of the primers used were individually designed using *Primer 3* and validated using *blat* and *In-Silico PCR.*

Exon	Primer sequences 5'-3'
2	F: CAGCCATTCTTTTCCTGCTC
	R: AGCCCAACCCTTGTCCTTAC
3 & 4	F: TGAGTGGATCCATTGGAAGG
	R: GCCAAAGGGTGAAGAGGAAT
	F: CCCCTCTGAGTCAGGAAACA
	R: GCCAAAGGGTGAAGAGGAAT
5	F: CTCTCTAGCTCGCTAGTGGGT
	R: CGAAAAGTGTTTCTGTCATCCAAAT
6	F: GCCATGGCCATCTACAAGCA
	R: TGGGGTTATAGGGAGGTCAAA
5 & 6	F: GCCGTCTTCCAGTTGCTTTA
	R: CTTAACCCCTCCTCCCAGAG
7	F: ACAGGTCTCCCCAAGG
	R: AAACTGAGTGGGAGCAGTAAGGAGA
	F: CCTGCTTGCCACAGGTCT
	R: TGATGAGAGGTGGATGGGTAG
	F: TAACCCCATGAGATGTGCAA
	R: GGGATGTGATGAGAGGTGGA
8	F: GGACAAGGGTGGTTGGGAGTAGA
	R: CCCAATTGCAGGTAAAACAGTCAAG
9	F: GCGCACAGAGGAAGAGAATC
	R: TGTCTTTGAGGCATCACTGC
10	F: CTGGGCAACAGAGTGAGACC
	R: CACTGAGGCAAGAATGTGGTT
	F: CAGTTTCTACTAAATGCATGTTGCT
	R: ATACACTGAGGCAAGAATGTGGTTA
11	F: GATTTGAATTCCCGTTGTCC
	R: CCACAACAAAACACCAGTGC

RESULTS

Identification and Classification of Potential Driver Mutations

We identified potential driver mutations from non-silent variants found by WES in both Patient 1 and Patient 2 tumor samples by matching the gene and location of variant to the COSMIC cancer gene census (August 2015). While we obtained one tumor sample from Patient 1, we obtained two samples from Patient 2—samples A and B—from two distinct locations of the same tumor. SNVs that have been previously annotated in COSMIC were categorized based on criteria established by Murugaesu *et al.* (2015). After the SNV was validated to be a tumor suppressor gene via COSMIC, we proceeded to use three computational approaches—Sift, Polyphen, and MutationTaster—to predict the possible damaging effects of the mutation. A variant would be classed as Category 1, or high confidence driver mutation, if two of the three programs predicted the specific variant to be deleterious. If it did not match the criteria, then the proximity of the SNV to other mutation entries in COSMIC was considered. If there were more than 3 other entries listed on the COSMIC database within 15 bps of the variant, then it would be classed as Category 2, or putative driver mutation. If it failed to be classed in any of criteria described, then it would be classified as Category 3, or low confidence driver mutation. If the gene was determined to be a dominant oncogene according to COSMIC, then it would automatically be Category 1 if there were more than 3 entries of the same mutation in COSMIC. If there were less than three identical matches of the

variant, then it would be classed as Category 2 or 3 based on the same criteria as tumor suppressor genes. If the variant was not entered in the COSMIC gene census, then it would be classed a Category 2 if two out of the three predictor programs indicated the SNV to be harmful, otherwise Category 3. All other mutations were classified as Category 4, or mutations of unknown significance. Deletions and insertions were not categorized based on the criteria described above. However, indels were validated on COSMIC and checked for deleterious effects on MutationTaster. In addition, the number of entries on COSMIC within 15 base pairs of each indel were counted.

Table 2 displays high category (Category 1-3) SNVs and indels with either frameshift mutations or have more than three entries in COSMIC gene census within 15 base pairs. We found two high confidence driver mutations, one putative driver mutation, and one low confidence driver mutation in Patient 1. Patient 1 also had two indels with potential driver mutation characteristics. In Patient 2, there are two high confidence driver mutations, one putative driver mutation, and one low confidence driver mutation in both tumor samples A and B, one low confidence driver mutation exclusively in tumor sample A, and one high confidence and one low confidence driver mutation exclusively in tumor sample B. There are also two frameshift deletions exclusively in tumor A of Patient 2 and two frameshift deletions exclusively in tumor B of Patient 2.

Table 2: Overview of potential driver mutations found in Patient 1 and Patient 2 with SNM. 15bp (number of mutations found in COSMIC within 15 base pairs of variant). N/A (program does not recognize the variant or program does not provide information on variant). Driver mutations listed on the table have all been previously identified in COSMIC cancer gene census (September 2015). Driver mutations have been categorized, according to criteria established by Murugaesu *et al*. Patient 1 has two high confidence driver mutations (Category 1), Patient 2 has three high confidence driver mutations from both sample A and B. Through this classification process, we can better understand which gene mutations are responsible for tumorigenesis of SMNs.

Patient 1 only						
Gene	**Nucleotide**	**Category**	**SIFT**	**Polyphen**	**Mutation Taster**	**15bp**
CSMD1	G7802A	3	tolerated	probably damaging	poly-morphism	0
GRIK3	G2518A	2	N/A	Benign	disease causing	4
LILRB1	1417_1419 del	-	N/A	N/A	poly-morphism	5
MTRR	G364A	1	tolerated	probably damaging	disease causing	3
NCOA2	G3523T	1	N/A	probably damaging	disease causing	1
SMAD7	627_628ins C	-	N/A	N/A	disease causing	6
Patient 2 both tumor samples A and B						
Gene	**Nucleotide**	**Category**	**SIFT**	**Polyphen**	**Mutation Taster**	**15bp**
COL4A3	G3476A	2	N/A	benign	poly-morphism	23
MAATS1	C1813T	1	damaging	N/A	disease causing	0
MOV10L1	C656T	1	N/A	probably damaging	disease causing	10
NTSR2	C868T	2	tolerated	probably damaging	N/A	5
Patient 2 tumor A only						
Gene	**Nucleotide**	**Category**	**SIFT**	**Polyphen**	**Mutation Taster**	**15bp**
ETV3L	C407T	2	N/A	N/A	poly-morphism	4
FBN3	2259delC	-	N/A	N/A	disease causing	4

RQCD1	350delT	-	N/A	N/A	disease causing	3
Patient 2 tumor B only						
Gene	**Nucleotide**	**Category**	**SIFT**	**Polyphen**	**Mutation Taster**	**15bp**
CYHR1	430delG	-	N/A	N/A	disease causing	4
NALCN	G2081A	1	tolerated	probably damaging	disease causing	11
TCF20	737_739del	-	N/A	N/A	disease causing	3

From the potential driver mutations found in SMN sarcomas from Patient 1 and Patient 2 via WES, we hypothesize that the top two driver mutation candidates are *SMAD7* and *NCOA2*. In addition to the high confidence driver mutation (Category 1) of *NCOA2*, a gene that encodes a transcriptional coactivator for nuclear receptors known to be mutated in sarcomas, Patient 1 also harbor a frameshift mutation, a known deleterious genetic event, in *SMAD7*, a gene that encodes a protein mediator in the TGF-B tumor suppressor pathway known to be abnormally expressed in various cancers.

Germline TP53 Validation Cohort

An accumulation of genetic mutations is known to induce cancer pathogenesis (Almeida & Barry, 2011). Most of these alterations are located within the following three categories of genes: proto-oncogenes, tumor suppressor genes, or DNA repair genes (Cooper et al., 2013). Proto-oncogenes control cell division and growth while tumor suppressor genes suppress cell growth and division (Pfeifer & Besaratinia, 2009). Mutations within these three categories of genes have a direct correlation with oncogenesis (Iengar, 2012).

The tumor suppressor protein p53, encoded by gene *TP53*, is in the center of regulation of cell cycle checkpoints, apoptosis, DNA repair, cellular metabolism, and senescence (Gibbons, Byers, & Kurie, 2014). When genotoxic stress is induced on the cell, stabilization of p53 promotes the transcription of genes that plays a role in cell-cycle arrest, apoptosis, and DNA repair (Liu, Song, & Xu, 2010). Mutations of p53 is known to have a driving role in tumor progression by deregulating these important mechanisms of cancer suppression (Sigal & Rotter, 2000). 50% of all human cancers are known to have p53 mutations, with 74% of these mutations being missense mutations (Meek, 2015).

P53 mutations are mainly missense mutations that not only inactivate the normal protein function, but also gain altered functions that further promote oncogenesis (Liu et al., 2010, p. 53). Common missense mutations occur in the DNA binding domain (exons 5-8) that changes the conformation of the protein, interfering with the binding domain and its ability to regulate transcription of

downstream genes (Sigal & Rotter, 2000). Frequent somatic mutations within this hotspot include R175H, commonly detected in breast cancer (Byler et al., 2014), and R249S in liver cancer (Goldstein et al., 2011). Germline p53 gene mutations are mainly associated with Li-Fraumeni syndrome, a familial autosomal dominant predisposition to a wide array of cancers, mainly early onset of sarcomas, gliomas, and breast carcinomas, with 57% probability of developing a second cancer in the next 30 years (Malkin et al., 1992). While families with Li-Fraumeni Syndrome (LFS) are closely monitored and studied, those who lack family histories of LFS cancers but harbor de novo *tp53* mutations are poorly characterized (Kilpivaara & Aaltonen, 2013). However, there has been emerging evidence that germline *TP53* mutations not related to LFS also have oncogenic roles in different cancer types (Kamihara, Rana, & Garber, 2014).

To identify germline mutations of *TP53* in non-LFS patients with second malignant neoplasms (SMN), we sequenced the germline DNA of 41 patients from a validation cohort comprised of pediatric cancer survivors with subsequent neoplasms registered in CCSS. Sanger sequencing was performed on exons 2-11 of *TP53*, and we observed a synonymous mutation in exon 6 codon 213 (A639G in the gene, R213R in the protein) in 4 out of 37 patients (11%). This single nucleotide polymorphism rs1800372 has a minor allele frequency of 0.5% in the general population (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1800372). This significant increase in frequency of this SNP in the validation cohort compared to

the general population leads us to believe the mutation may have a deleterious effect on the DNA binding domain and on the genetic susceptibility to SMN. A second germline variant identified in this validation cohort was G215C in exon 4 (R72P in the protein), which was present in 6 patients and mutually exclusive with the germline A639G variant, resulting in a total of ten out of 38 evaluable patients (26%) harboring a germline *TP53* variant. While the R72P polymorphism is common in the general population (Vietri et al., 2007), the A639G *TP53* variant is uncommon in the general population but significantly enriched in our cohort. Shown in Figure 1 are sequencing chromatograms from four pediatric cancer survivors who harbor germline variant A639G and six individuals demonstrating variant G215C. Both sets of variants are compared to the normal control DNA on the first row as reference. These chromatograms demonstrate clear heterozygous mutations harbored by patients in the validation cohort.

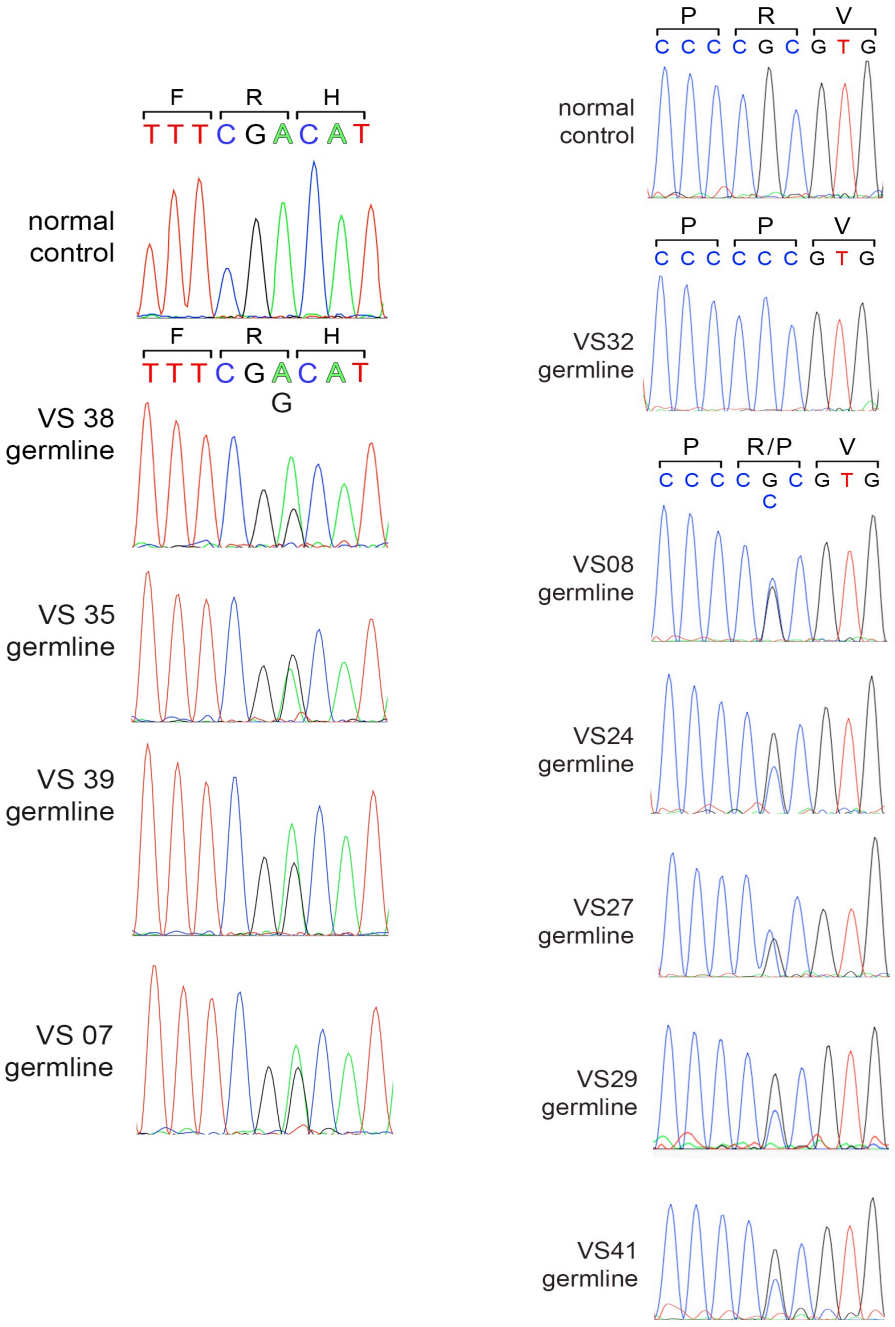

Figure 1. Representative SMN samples showing *TP53* mutations found in the validation cohort of pediatric cancer survivors with SMNs. Synonymous mutation A639G was identified in four patients—VS 38, VS 35, VS 39, VS07; Missense mutation R72P was harbored in six patients mutually exclusive from patients who harbor A639G—VS32, VS08, VS24, VS27, VS29, VS41. These

SMN patient samples were obtained from the validation cohort of pediatric patients registered in CCSS.

DISCUSSION

Identification of Deleterious Driver Mutations

Cancer development involves stages of initiation, promotion, and progression (Weinberg, 2013). These stages are characterized by aberrant genetic events that accumulate over time. To identify possible significant genetic events related to radiation-induced malignancies, WES was performed on two pediatric cancer survivors with SMNs, Patient 1 and Patient 2. Through WES, we identified a frameshift deletion in codon 209 of gene SMAD7 in Patient 1, which has been found to be implicated in pancreatic, gastric, skin, breast, liver, and prostate cancer (Stolfi, Marafini, De Simone, Pallone, & Monteleone, 2013, p. 7). As shown in Figure 2, codon 209 is part of the N-terminal MAD homology 1 domain(MH1), which is related to protein regulation and stability. SMAD7 is part of the Smad protein family that is activated via phosphorylation by TGF-B Type I receptor (Yan & Chen, 2011). Transforming Growth Factor (TGF-B) is a tumor suppressor cytokine that is known to inhibit proliferation and promote apoptosis in normal cells (Luo, Li, Lv, & Huang, 2014). On the other hand, TGF-B signaling is also known to drive late cancer cells to promote epithelial-mesenchymal transition that leads to invasion and metastasis (Stolfi et al., 2013). Studying the dual contradicting roles of TGF-B has paved the way of understanding its downstream mechanisms and discovering intracellular signal proteins, among which Smad7 is a negative regulator of the TGF-B cascade signaling pathway (Stolfi et al., 2013). Smad proteins are grouped into three classes: Receptor

activated Smads, inhibitory Smads, and common-mediator Smads (Luo et al., 2014).

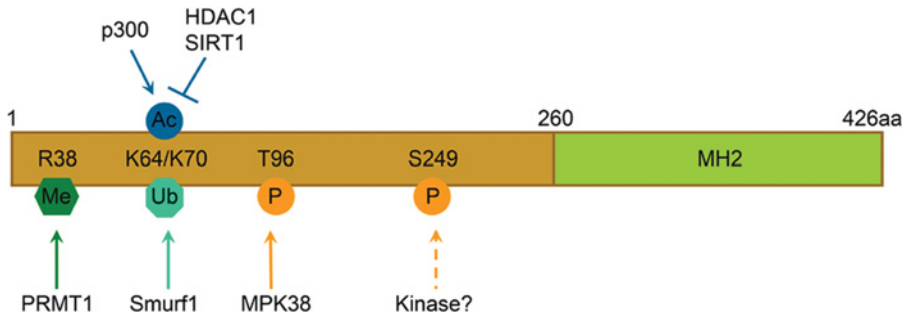

Figure 2. Structure of Smad7. We detected a deletion of two nucleotides in codon 209 in the SMN of Patient 1 located at the N-terminal of the protein. The N-terminus interacts with the MH2 domain while inhibiting TGF-B Type I receptor. Post-translational modifications occur in the N-terminus of the protein, regulating protein stability and functions (Yan & Chen, 2011).

As seen in Figure 3, TGF-B is a cytokine that signals through its interactions with its membrane-bound receptors, TGF-B Type I (TBRI) and TGF-B Type II (TBRII) (Stolfi et al., 2013). These receptors interact with regulatory Smad (R-Smad) proteins and Co-smad proteins to promote cell-cycle arrest, differentiation, or apoptosis (Stolfi et al., 2013). Smad7, an I-smad, suppresses TGF-B signaling in both the cytoplasm and the nucleus via various mechanisms. In the cytoplasm, Smad7 blocks R-Smad phosphorylation, promotes degradation of TGF-B RI, and suppresses TGF-B RI activation by dephosphorylation (Stolfi et al., 2013). In the nucleus, Smad7 inhibits the formation of R-Smad/Smad4 complexes and disrupts the R-Smad/Smad4 complexes from binding to DNA (Stolfi et al., 2013).

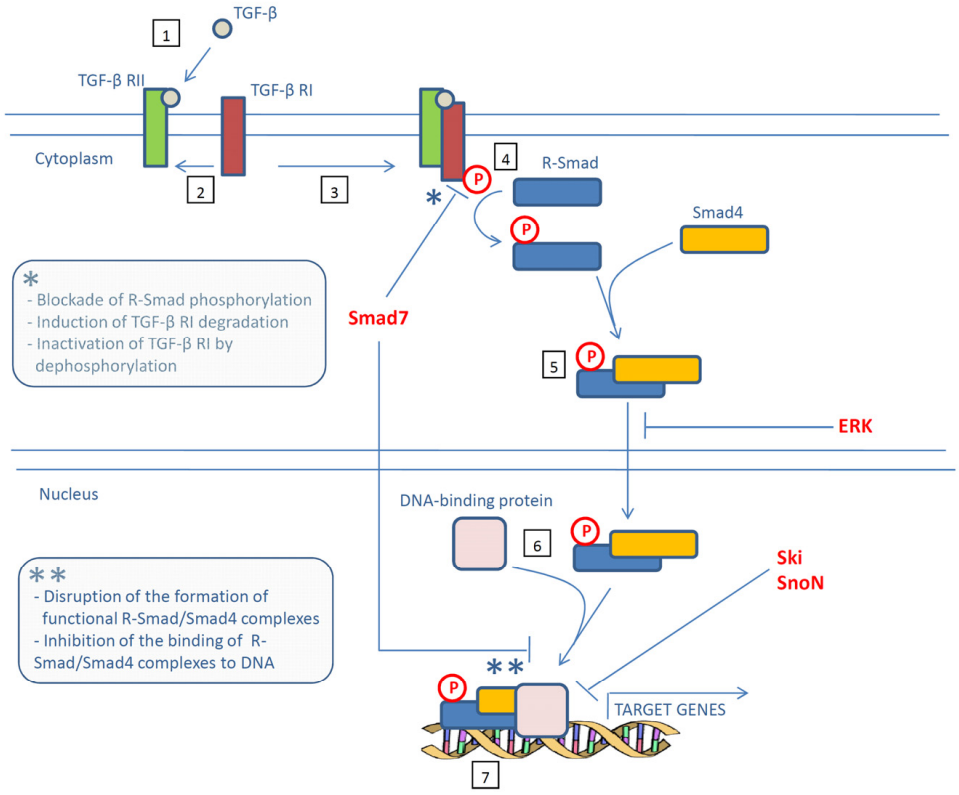

Figure 3. Schematic of the transforming growth factor (TFG-B) signaling pathway. When TGF-B binds to its type II receptor, TGF-B type I forms a receptor complex with TGF-B RI. When the receptor complex is activated, TGF-B RI phosphorylates an R-smad, which allows the R-smad to bind to Smad4. This complex moves into the nucleus to bind to the DNA and promotes the transcription of target genes. Smad7 antagonizes the signaling pathway in both the cytoplasm and the nucleus, as indicated by * and ** (Stolfi et al., 2013).

In addition to acting as an inhibitor in the TGF-B signaling pathway, Smad7 also promotes cancer progression (Luo et al., 2014). On one hand, there is an increase in the expression of SMAD7 in a variety of cancers, such as breast, colorectal, and prostate cancers (Luo et al., 2014). Furthermore, an over-expression of SMAD7 is correlated to poor prognosis in colon adenocarcinoma (Luo et al., 2014). Upregulation of SMAD7 in these cancers indicate that SMAD7

inhibited the tumor suppressor role of TGF-B signaling (Luo et al., 2014). However, other tumors also indicate a deletion of the gene can promote invasion and metastasis, shown in liver, breast, and skin cancer (Yan & Chen, 2011). While Smad7's pro- and anti-tumor roles are contradictory, this is not surprising given the dual roles of TGF-B signaling pathway in different stages of cancer (Stolfi et al., 2013).

Given the complex role of Smad7 in TGF-B signaling pathway, any genetic aberrations on this gene could have a deleterious effect on cancer cells, promoting malignancies and cancer progression. Furthermore, due to the damaging consequences of frameshift mutations (Iengar, 2012), the Smad7 protein in Patient 1 is most likely abnormal, potentially leading to secondary tumor formation post-RT.

Analysis of *TP53* Polymorphisms found in Second Malignant Neoplasms

Cancer is a disease that results from an accumulation of cells that are unable to maintain homeostasis due to irregular genetic and epigenetic events (Cooper et al., 2013). Sequencing has paved the way to identifying and understanding the various genetic alterations that occur in cancer genomes (Stratton, 2011). Through genomic sequencing, tumor suppressor gene *TP53* has been confirmed to be the most mutated in human cancers, mutated in over 50% of human cancers (Bieging & Attardi, 2012). The majority of these mutations has been classified as missense mutations, mutations that change a single amino acid that result in the expression of a mutated protein that potentially leads to deleterious effects (Goldstein et al., 2011).

While the role of *TP53* has been widely studied in various cancers, it is still poorly characterized in SMN patients (Sherborne et al., 2015). To identify possible recurrent *TP53* polymorphisms and mutations in our validation cohort of SMN patients, Sanger sequencing was performed on exons 2-11 of the tumor suppressor gene. This study identified the recurrent SNP rs1800372 in *TP53*, as shown in Figure 4. Although this SNP results in a synonymous mutation and does not change the sequence of the encoded protein, it has been correlated with poor prognosis in primary breast cancer (Berns et al., 1998). Furthermore, synonymous mutations, or “silent” mutations, have also been found to be driver mutations in human cancers, contributing to the tumor phenotype (Supek, Miñana, Valcárcel, Gabaldón, & Lehner, 2014). Silent mutations are known to

alter translational accuracy by affecting mechanisms related to mRNA conformation, mRNA translation, and pre-mRNA splicing (Supek et al., 2014).

A second germline variant identified in this validation cohort is G215C in exon 4, or R72P in the protein, in 6 additional patients. While this polymorphism is quite common in the general population, this polymorphism has a profound change on the primary structure of the p53 protein. As indicated in Figure 4, the variant is located on the proline-rich domain, a region required to induce apoptosis (Pim & Banks, 2004). It has been previously shown that the Arg72 variant demonstrated higher efficiency at inducing apoptosis than the Pro72 variant, indicating higher risk for neoplastic formation for those who harbor the variant (Vietri et al., 2007).

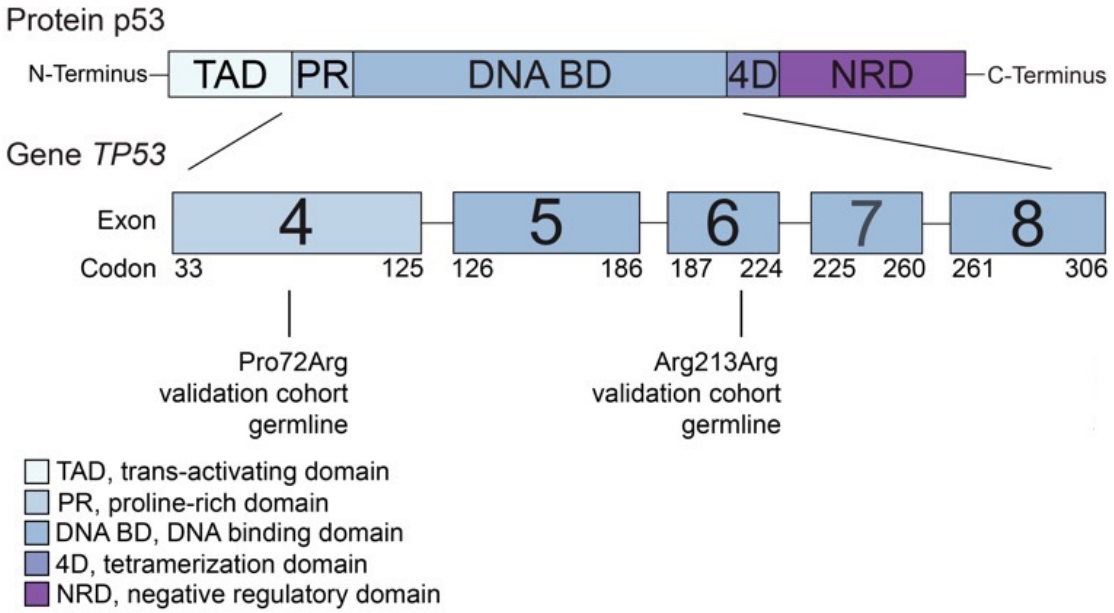

Figure 4. Localization of *TP53* mutations in relation to the mature functional region of p53 protein. Note the locations of germline variants found in 37 pediatric cancer patients with radiation-induced SMNs. The first recurrent variant found from the validation cohort is R213R, a synonymous mutation located on exon 6, which is part of DNA binding domain of protein p53, also known as the "hotspot" of p53 mutations implicated in cancer. P72R variant was also identified in the validation cohort, which is located on exon 4 that is part of the proline-rich domain on the functional protein p53. Although a common polymorphism in the general population, it is implicated in poor apoptosis induction, promoting cell proliferation and suppressing anti-tumor properties.

Importance of Cancer Genome Sequencing

Next-generation sequencing, or second-generation sequencing, has revolutionized the understanding of the molecular biology of cancer (Pfeifer & Besaratinia, 2009). These large scale sequencing studies, such as WES, have provided the means to identify driver mutations, to discover mutational signatures, and to pave the way of personalized medicine (Stratton, 2011). As exhaustive sequencing of cancer genomes is under way, the result of these analyses will not only change our understanding, but it will also lead to new target therapies to treat and to hopefully prevent the disease altogether (Mwenifumbo & Marra, 2013). This new comprehensive data will help scientists and clinicians predict cancer behavior, tumor progression and responses to therapies, such as radiation therapy (Kilpivaara & Aaltonen, 2013).

REFERENCES

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., … Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. http://doi.org/10.1038/nature12477

Almeida, C. A., & Barry, S. A. (2011). *Cancer: Basic Science and Clinical Aspects*. John Wiley & Sons.

Berns, E. M., van Staveren, I. L., Look, M. P., Smid, M., Klijn, J. G., & Foekens, J. A. (1998). Mutations in residues of TP53 that directly contact DNA predict poor outcome in human primary breast cancer. *British Journal of Cancer*, *77*(7), 1130–1136.

Bieging, K. T., & Attardi, L. D. (2012). Deconstructing p53 transcriptional networks in tumor suppression. *Trends in Cell Biology*, *22*(2), 97–106. http://doi.org/10.1016/j.tcb.2011.10.006

Byler, S., Goldgar, S., Heerboth, S., Leary, M., Housman, G., Moulton, K., & Sarkar, S. (2014). Genetic and Epigenetic Aspects of Breast Cancer Progression and Therapy. *Anticancer Research*, *34*(3), 1071–1077.

Cooper, W. A., Lam, D. C. L., O'Toole, S. A., & Minna, J. D. (2013). Molecular biology of lung cancer. *Journal of Thoracic Disease*, *5*(Suppl 5), S479–S490. http://doi.org/10.3978/j.issn.2072-1439.2013.08.03

DeSantis, C. E., Lin, C. C., Mariotto, A. B., Siegel, R. L., Stein, K. D., Kramer, J. L., … Jemal, A. (2014). Cancer treatment and survivorship statistics, 2014. *CA: A Cancer Journal for Clinicians*, *64*(4), 252–271. http://doi.org/10.3322/caac.21235

Fang, M., Hutchinson, L., Deng, A., & Green, M. R. (2016). Common BRAF(V600E)-directed pathway mediates widespread epigenetic silencing in colorectal cancer and melanoma. *Proceedings of the National Academy of Sciences of the United States of America*. http://doi.org/10.1073/pnas.1525619113

Gibbons, D. L., Byers, L. A., & Kurie, J. M. (2014). Smoking, p53 Mutation, and Lung Cancer. *Molecular Cancer Research : MCR*, *12*(1), 3–13. http://doi.org/10.1158/1541-7786.MCR-13-0539

Goldstein, I., Marcel, V., Olivier, M., Oren, M., Rotter, V., & Hainaut, P. (2011). Understanding wild-type and mutant p53 activities in human cancer: new

landmarks on the way to targeted therapies. *Cancer Gene Therapy*, *18*(1), 2–11. http://doi.org/10.1038/cgt.2010.63

Hejmadi, M. (2009). *Introduction to Cancer Biology*. Bookboon.

Iengar, P. (2012). An analysis of substitution, deletion and insertion mutations in cancer genes. *Nucleic Acids Research*, *40*(14), 6401–6413. http://doi.org/10.1093/nar/gks290

International Agency for Research on Cancer Working Group on artificial ultraviolet (UV) light and skin cancer. (2007). The association of use of sunbeds with cutaneous malignant melanoma and other skin cancers: A systematic review. *International Journal of Cancer. Journal International Du Cancer*, *120*(5), 1116–1122. http://doi.org/10.1002/ijc.22453

Kamihara, J., Rana, H. Q., & Garber, J. E. (2014). Germline TP53 Mutations and the Changing Landscape of Li–Fraumeni Syndrome. *Human Mutation*, *35*(6), 654–662. http://doi.org/10.1002/humu.22559

Kilpivaara, O., & Aaltonen, L. A. (2013). Diagnostic Cancer Genome Sequencing and the Contribution of Germline Variants. *Science*, *339*(6127), 1559–1562. http://doi.org/10.1126/science.1233899

Liu, D., Song, H., & Xu, Y. (2010). A common Gain of function of p53 cancer mutants in inducing genetic instability. *Oncogene*, *29*(7), 949–956. http://doi.org/10.1038/onc.2009.376

Luo, L., Li, N., Lv, N., & Huang, D. (2014). SMAD7: a timer of tumor progression targeting TGF-β signaling. *Tumor Biology*, *35*(9), 8379–8385. http://doi.org/10.1007/s13277-014-2203-7

Malkin, D., Jolly, K. W., Barbier, N., Look, A. T., Friend, S. H., Gebhardt, M. C., … Strong, L. C. (1992). Germline Mutations of the p53 Tumor-Suppressor Gene in Children and Young Adults with Second Malignant Neoplasms. *New England Journal of Medicine*, *326*(20), 1309–1315. http://doi.org/10.1056/NEJM199205143262002

Meek, D. W. (2015). Regulation of the p53 response and its relationship to cancer. *Biochemical Journal*, *469*(3), 325–346. http://doi.org/10.1042/BJ20150517

Mwenifumbo, J. C., & Marra, M. A. (2013). Cancer genome-sequencing study design. *Nature Reviews Genetics*, *14*(5), 321–332. http://doi.org/10.1038/nrg3445

Pfeifer, G. P., & Besaratinia, A. (2009). Mutational spectra of human cancer. *Human Genetics*, *125*(5-6), 493–506. http://doi.org/10.1007/s00439-009-0657-2

Pim, D., & Banks, L. (2004). p53 polymorphic variants at codon 72 exert different effects on cell cycle progression. *International Journal of Cancer*, *108*(2), 196–199. http://doi.org/10.1002/ijc.11548

Pirozynski, M. (2006). 100 years of lung cancer. *Respiratory Medicine*, *100*(12), 2073–2084. http://doi.org/10.1016/j.rmed.2006.09.002

Reference SNP (refSNP) Cluster Report: rs1800372 ** With Benign allele **. (n.d.). Retrieved April 1, 2016, from http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=1800372

Robison, L. L., & Hudson, M. M. (2014). Survivors of childhood and adolescent cancer: life-long risks and responsibilities. *Nature Reviews Cancer*, *14*(1), 61–70. http://doi.org/10.1038/nrc3634

Sherborne, A. L., Davidson, P. R., Yu, K., Nakamura, A. O., Rashid, M., & Nakamura, J. L. (2015). Mutational Analysis of Ionizing Radiation Induced Neoplasms. *Cell Reports*, *12*(11), 1915–1926. http://doi.org/10.1016/j.celrep.2015.08.015

Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA: A Cancer Journal for Clinicians*, *65*(1), 5–29. http://doi.org/10.3322/caac.21254

Sigal, A., & Rotter, V. (2000). Oncogenic Mutations of the p53 Tumor Suppressor: The Demons of the Guardian of the Genome. *Cancer Research*, *60*(24), 6788–6793.

Sompayrac, L. (2004). *How Cancer Works*. Jones & Bartlett Learning.

Stolfi, C., Marafini, I., De Simone, V., Pallone, F., & Monteleone, G. (2013). The Dual Role of Smad7 in the Control of Cancer Growth and Metastasis. *International Journal of Molecular Sciences*, *14*(12), 23774–23790. http://doi.org/10.3390/ijms141223774

Stratton, M. R. (2011). Exploring the Genomes of Cancer Cells: Progress and Promise. *Science*, *331*(6024), 1553–1558. http://doi.org/10.1126/science.1204040

Supek, F., Miñana, B., Valcárcel, J., Gabaldón, T., & Lehner, B. (2014). Synonymous Mutations Frequently Act as Driver Mutations in Human Cancers. *Cell*, *156*(6), 1324–1335. http://doi.org/10.1016/j.cell.2014.01.051

Thomas, L. K., Bermejo, J. L., Vinokurova, S., Jensen, K., Bierkens, M., Steenbergen, R., … Reuschenbach, M. (2014). Chromosomal gains and losses in human papillomavirus-associated neoplasia of the lower genital tract – A systematic review and meta-analysis. *European Journal of Cancer*, *50*(1), 85–98. http://doi.org/10.1016/j.ejca.2013.08.022

Vietri, M. T., Riegler, G., Ursillo, A., Caserta, L., Cioffi, M., & Molinari, A. M. (2007). p53 codon 72 polymorphism in patients affected with ulcerative colitis. *Journal of Gastroenterology*, *42*(6), 456–460. http://doi.org/10.1007/s00535-007-2026-z

Weinberg, R. (2013). *The Biology of Cancer, Second Edition*. Garland Science.

Whipple, C. A., & Brinckerhoff, C. E. (2014). BRAFV600E melanoma cells secrete factors that activate stromal fibroblasts and enhance tumourigenicity. *British Journal of Cancer*, *111*(8), 1625–1633. http://doi.org/10.1038/bjc.2014.452

Yan, X., & Chen, Y.-G. (2011). Smad7: not only a regulator, but also a cross-talk mediator of TGF-β signalling. *Biochemical Journal*, *434*(1), 1–10. http://doi.org/10.1042/BJ20101827

Yarden, Y. (2001). Biology of HER2 and its importance in breast cancer. *Oncology*, *61 Suppl 2*, 1–13. http://doi.org/55396

CURRICULUM VITAE

YONG EUN LEE
Address: 25 Saint Mary's Court Apt 2, Brookline, MA 02446
Phone: (510) 672-3124 • E-mail: lylee@bu.edu • Year of Birth: 1987

EDUCATION

Boston University School of Medicine, Boston, MA
Master of Science, Medical Sciences (Expected May 2016)

University of California, Berkeley, Berkeley, CA
Bachelor of Science, Nutritional Science, December 2012

RESEARCH EXPERIENCE

University of California, San Francisco, Department of Radiation Oncology
Master's Research Assistant, July 2015-May 2016
Principle Investigator: Dr. Jean Nakamura, MD

University of California, Berkeley, Department of Nutritional Sciences & Toxicology
Undergraduate Research Assistant, November 2006-
May 2008 Principle Investigator: Dr. Hei Suk Sul, PhD

- Assisted in study of the role of DNA-dependent protein kinase in feeding/insulin-dependent activation of lipogenic genes, linking DNA-PK to insulin-signaling pathway

PROFESSIONAL EXPERIENCE

CEP America, Doctors Medical Center, San Pablo, CA
Scribe, January 2014-Present

- Document patient histories, physical exams, and lab results while directly observing each case the doctor sees
- Handle confidential information
- Gain direct knowledge of how an ER is run in a low income environment
- Interact with the doctors, nurses, and hospital staff to create a positive work environment

Aspire Education Project, Oakland, CA
Instructor, December 2013-Present

- Instruct high school students on high-level math and science subjects
- Establish positive synergistic relationships with students and parents through

effective communication and constant availability

Serenity MedSpa, San Francisco, CA
Back Office Medical Assistant, October 2012-August 2013

- Prepared patients and arranged rooms for doctor visits while independently managing back office to ensure smooth workflow throughout entirety of workday
- Served as a primary liaison between doctors and patients
- **Extreme Learning, Santa Fe Elementary School**, Oakland, CA
  Academic Coach, January-March 2010
- Organized and directed an after-school tutoring program of twenty elementary school students to improve their fundamental reading, writing, and math skills
- Mentored students who often struggled in school and came from underserved communities
- Conducted meetings with parents to formulate productive curricula tailored to each student

---

VOLUNTEER WORK

Prevention International: No Cervical Cancer, Leon Nicaragua
Support Volunteer, June 2013

- Demonstrated meticulous organizational skills to accumulate patient information for efficient data entry by evaluating medical charts in Spanish and assessing exam forms from clinicians
- Heavily relied on by doctors and nurse practitioners to manage and to delegate tasks including: interviewing and referring patients, cleaning and inventorying medical equipment, and dispensing medications
- Trekked to health centers in neighboring rural areas to participate in the Community Outreach and Education project to raise awareness of cervical cancer prevention programs

Staff Volunteer, June 2013-present

- Coordinated successful fundraising events, including an end-of-the-year phonathon with committee members

Suitcase Clinic, Berkeley, CA
Continuity of Care Advocate, May 2007-April 2008

- Counseled uninsured, underserved, and homeless clients in obtaining free medical, legal, food and housing services in addition to those available at the weekly drop-in center
- Offered and executed referrals to the local primary care clinic by scheduling and accompanying clients to their appointments to ascertain proper care and to establish positive relationships for follow-up

- Compiled an exhaustive handbook of free health and social services available in the San Francisco Bay Area to serve as a beneficial resource for the clients
- Maintained a synergistic relationship between primary care clinic and Suitcase Clinic *Caseworker*, August 2006-May 2007
- Accompanied the client throughout duration of visit to the drop-in center with active listening skills and empathetic support
- Completed mandatory service-training course of various topics related to homelessness and social justice to become an effective caseworker

---

TEACHING & TUTORING EXPERIENCE

Self-Employed, Albany, CA
Private Tutor, December 2009-June2012
- Tutored middle school and high school students in math and science courses, and in their preparation for the SATI/II exam

A+ Tutors, Richmond, CA
Private Tutor, August 2009-August 2011
- Created lesson plans and helped with homework in math and science courses

---

HONORS/AWARDS

University of California, Berkeley, Incentive Awards Program, Berkeley, CA
Incentive Awards Program Scholar, 2005-2009
- Received full academic scholarship