8-2021

Identification and characterization of de novo germline TP53 mutation carriers in families with Li-Fraumeni Syndrome

Carlos C. Vera Recio

Identification and characterization of de novo germline *TP53* mutation carriers in families with Li-Fraumeni Syndrome

by

Carlos C Vera Recio, BS

APPROVED:

____

Wenyi Wang, Ph.D., Advisory Professor

____

Guillermina Lozano, Ph.D.

____

Nicholas Navin, Ph.D.

____

Ryan Sun, Ph.D.

____

Elmer Bernstam, Ph.D.

APPROVED:

____

Dean, The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Science

Identification and characterization of de novo germline *TP53* mutation carriers in families with Li-Fraumeni Syndrome

A

Thesis

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

by

Carlos C Vera Recio, BS

Houston, Texas

August 2021

Dedication

To my parents Awilda Recio Camacho, Carlos Felipe Vera Muñoz. To my sister, Carla Crystal Vera Recio. To my grandparents, Jose Guadalupe, Carlos Manuel, Isabel and Livia. To all my cousins, to the ones I share blood with, and the ones I chose along the way. Finally, to my lovely fiancée, Nichollette F. Gonzalez Massa.

Acknowledgements

First and foremost, I would like to acknowledge my mentor Dr. Wenyi Wang. She gave me the opportunity to be a part of her lab, found several incredibly interesting research projects that were a good fit for me and has been invested in every step of my progress as a researcher. Wenyi cares about her students and is dedicated to their success, and my case was not the exception. I would not have come as far as I did without her support.

Secondly, I would like to thank my advisory committee members. Dr. Lozano, Dr. Bernstam, Dr. Navin, Dr. Sun, Dr. Subudhi and Dr. Morris. All of you have provided me with invaluable lessons and insight that have guided me and my project. Thank you for your help, in spite of your busy schedule. I want to give a very special thanks to Dr. Lozano, who was also my secondary mentor as an NLM trainee, and also gave me the opportunity to work with her mice model data.

Thirdly, I would like to thank all my brilliant and friendly lab members, who made Dr. Wang's lab a great place to be a graduate student. Thanks to Peng Yang and Shaolong Cao, whom besides my lab members, I'm happy to be able to call my friends, and who've provided commentary and help whenever it was asked. A very special thanks to Elissa Dodd-Eaton, who has been an invaluable presence in the lab since day one, and who I've worked closely on many projects, especially this one.

Fourthly, I would like to thank graduate students from other labs who've become my friends during my time in Houston, and who've helped me even when they had absolutely no ties to my work. Ramiz Iqbal from Dr. Chen’s lab, Chachad Dhruv from Dr. Lozano’s lab, and Dr. Naveen Ramesh formerly from Dr. Navin’s lab. You have been priceless housemates, friends, editors and support.

Last but definitely not least, I would like to thank my family. My parents, who have provided me with inestimable amounts of support every day of their life. My sister, whose self-proclaimed “annoying phone calls” made me happier in many occasions. Finally, to Nichollette who will never understand how much more wonderful she has made this experience for me by simply being there and supporting me in whichever way she could.

Identification and characterization of de novo germline *TP53* Mutation carriers in families with Li-Fraumeni Syndrome

Carlos C Vera Recio, M.S.

Advisory Professor: Wenyi Wang, Ph.D

Li-Fraumeni syndrome (LFS) is an inherited cancer syndrome caused by a deleterious mutation in *TP53*. An estimated 48% of LFS patients present due to a de novo mutation (DNM) in *TP53*. The knowledge of DNM status, DNM or familial mutation (FM), of an LFS patient requires genetic testing of both parents which is often inaccessible, making de novo LFS patients difficult to study. Famdenovo.*TP53* is a Mendelian Risk prediction model used to predict DNM status of *TP53* mutation carriers based on the cancer-family history and several input genetic parameters, including disease-gene penetrance. The good predictive performance of Famdenovo.*TP53* was demonstrated using data collected from four historical US cohorts. We hypothesize that by incorporating penetrance estimates that are specific for different types of cancers diagnosed in family members, we can develop a model with further improved calibration, accuracy and prediction. We present Famdenovo.CS, which uses cancer-specific penetrance estimates that were derived previously using a Bayesian semi-parametric competing risk model, to calculate the DNM probability. We validate Famdenovo.CS on 206 LFS families with known DNM status, from five different US cohorts. We used the concordance index (AUC), observed:expected ratios (OE) and Brier score (BS) to measure our model’s discrimination, calibration and accuracy, respectively. We use our model to analyze 101 families recently collected from the Clinical Cancer Genetic program at MD Anderson Cancer Center (CCG-MDA). We estimate the proportion of probands that present a DNM and compare DNM to FM carriers in several areas: cancer types diagnosed, age at diagnosis and mutations in *TP53*. Famdenovo.CS showed similar performance to Famdenovo.*TP53* in terms of discrimination with AUC of 0.95 and 0.77 in validation sets A and B respectively; while improving on the model accuracy and calibration, demonstrated by a significant decrease in the BS (-0.091, 95%. CI [-0.19, -0.024]) and improved OE (1.17, 95% CI [0.90, 1.46]). Of the 101 probands in the CCG-MDA cohort, we predict 39 to be DNMs and 62 to be FMs. The cancer types and ages of diagnosis observed in FMs and DNMs are similarly distributed. DNMs in *TP53* are a prevalent cause of LFS and we did not find differences in the clinical characteristics of DNM and FM carriers. Our model allows for a systematic identification and characterization of *TP53* DNM carriers.

Table of Contents

List of Illustrations

List of Tables

Chapter 1 – Background

The importance of de novo mutations

De novo mutations are defined as germline mutations that appear for the first time in an individual but while absent from his/her parents. An average rate of 1.2 x 10^{-8} de novo mutations are expected per nucleotide each generation (Jónsson et al., 2018; Ohno, 2019). Thanks to the increased availability of next generation sequencing (NGS) and study designs where whole-exome or whole-genome sequencing (WES, WGS) is generated for familial-trios, de novo mutations have been increasingly identified as causal of sporadic genetic diseases (Jin et al., 2018) including a plethora of neurological disorders (Turner and Eichler, 2019) such as Autism, CHARGE syndrome, Amyotrophic Lateral Sclerosis, Alzheimer's and Parkinson's disease (Brandler and Sebat, 2015; Jin et al., 2018; Nicolas and Veltman, 2019; Ronemus et al., 2014). De novo mutations have also been reported to cause Li-Fraumeni syndrome (LFS) (Gonzalez et al., 2009; Renaux-Petel et al., 2018), an inherited cancer syndrome cause by a pathogenic mutation in the tumor suppressor gene, *TP53* (Li and Fraumeni Jr, 1969; Malkin, 2011; Malkin et al., 1990).

The ongoing research efforts on de novo mutations have reported an increased genome-wide de novo mutation rate associated to older parental ages (Cioppi et al., 2019; Goldmann et al., 2019; Jónsson et al., 2017), recombination rate (Francioli et al., 2015), GC content (Michaelson et al., 2012), DNA hypersensitivity (Michaelson et al., 2012); and associated de novo mutations in cancer genes to older maternal (tumor suppressor) and paternal (oncogenes) ages (Acuna-Hidalgo et al., 2016). An increased number of genome-wide de novo mutations has also been associated with worse disease presentation in neurodegenerative disorders, particularly Autism (Michaelson et al., 2012). In Autism, de novo mutations are now being studied in an effort to personalize treatment (Brandler and Sebat, 2015).

Although most of research on de novo mutations have been focused on identifying the risk factors that increase the genome-wide accumulation of said mutations and the cumulative effects of the inflated count of de novo mutations, very little research has focused on understanding what are the mechanisms causal or associated to the accumulation of highly-pathogenic de novo mutations, such as mutations in *TP53* that cause LFS. Genomic characterization of patients with LFS that present due to de novo mutations (DNMs) might provide insight on the acquisition of deleterious germline mutations in *TP53*, which in turn could aid in developing strategies for early identification of DNMs or perhaps novel prenatal testing or implantation genetic testing (Gao et al., 2020; Schneider et al., 2019).

However, when a patient is diagnosed with LFS their DNM status, whether the patient is a DNM or if he carries a familial mutation (FM), is unknown. Confirmation of the DNM requires genetic testing of both parents which is often unavailable (Gao et al., 2020; Gonzalez et al., 2009; Renaux-Petel et al., 2018). Because of the critical role of the *TP53* gene, and the complex nature of LFS, systematically identifying DNMs in families with LFS without additional genetic testing requires sophisticated statistical methods (Gao et al., 2020). To better illustrate the clinical and statistical challenges that must be overcome, over the next sections, we will give a comprehensive overview of *TP53*, the common pathogenic mutations that affect this gene, LFS and introduce the statistical modeling approaches that will be used to address this need.

The *TP53* gene and its hotspot mutations

The *TP53* is a tumor suppressor gene located in chromosome 17 of the human genome, on the band 17p13.1 and spans the nucleotide positions 7,661,779 to 7,687,538 (Yates et al., 2020). This gene codes for at least 15 isoforms of the p53 protein (Vieler and Sanyal, 2018), which serve multiple molecular functions that are essential to response to stress including: regulation of apoptosis, cell cycle regulation, DNA repair, cell senescence and metabolism (Aubrey et al., 2016). Because most of these functions directly control molecular features that are essential to the hallmarks of cancer (Hanahan and Weinberg, 2011), and because *TP53* is the most commonly mutated gene in cancer (Aubrey et al., 2016; Barbosa et al., 2019; Levine, 2020; Olivier et al., 2010), it is commonly referred to as “The Guardian of the Genome” (Aubrey et al., 2016).

Although mutations in *TP53* are the most common genomic alteration in cancer, a subset of mutations are much more frequent and therefore termed hotspot mutations (Walerych et al., 2012). It is estimated that up to 90% of mutations in this gene in the context of cancer, occur on the DNA binding domain as missense mutations in one of 190 codons (Baugh et al., 2018). The seven most common of these mutations make up to 28% of the mutations in cancer, and cause the amino acid changes *TP53*-p.R175H, *TP53*-p.R248Q, *TP53*-p.R273H, *TP53*-p.R248W, *TP53*-p.R273C, *TP53*-p.R282W and *TP53*-p.G245S (Baugh et al., 2018). Both the *TP53*-p.R248 and the *TP53*-p.R273 residues make direct contact with the genomic DNA, hence, hotspot mutations in these two residues incapacitate the p53 protein from attaching to the DNA, causing loss of function and earning them the alias “contact mutants” (Walerych et al., 2012). The other mutations do not affect residues in direct contact with the DNA, however, due affect the structure of the functional p53 protein and are thus dubbed “structural mutants” (Baugh et al., 2018; Walerych et al., 2012).

Although these hotspot mutations result in the classic loss of function that cascades into tumorigenesis, gain of function (GoF) properties of mutations in *TP53* have been reported and implicated in tumorigenesis (Oren and Rotter, 2010; Stein et al., 2019). Although these GoF properties are less understood, some studies indicate that they might be a result of that mutant-p53 binding to other proteins (Kim and Lozano, 2018). Interestingly, some studies argue that some mutations might have stronger or more intense GoF properties than others. For example, *TP53*-p.R248Q was demonstrated to increase mitotic activity in cells with fragmented DNA in due to increased AKT signaling in mice; while also leading to worse survival in humans than other hotspot mutations in *TP53* (Hanel et al., 2013).

In the context of LFS, it is of clinical and research interest to understand if these possible GoF mutations translate to a different clinical presentation or different penetrance in the age of onset of disease. Interestingly enough, on previous work studying DNMs and FMs in LFS, the hotspot mutation causing *TP53*-p.R248Q was identified in both DNMs and FMs, while its counterpart, *TP53*-p.R248W was only identified in DNMs, even though both were in similar frequency (Gao et al., 2020). The natural follow up question is: Is the *TP53*-p.R248W mutation absent in DNMs or not ascertained? If its absent, then studies that sequence family trios where the offspring has a DNM in *TP53* are indicated to understand would why some mutations in *TP53* are selectively allowed to be established in the germline as DNMs. If the mutation has occurred in DNMs but has not been ascertained into clinical cohorts, then there might be differences in the clinical characteristics of DNMs that depend on the deleterious mutation in TP53, and we might need to further study these carriers to improve on the current ascertainment strategies. Whichever the case, the critical step is identifying the DNMs in families with LFS, that, as discussed above, requires statistical modeling.

Overview of Li-Fraumeni Syndrome

Li-Fraumeni syndrome (LFS) is a rare, autosomal dominant cancer predisposition syndrome first described by Dr. Frederick Li and Joseph Fraumeni (Li and Fraumeni Jr, 1969) after reviewing 648 childhood rhabdomyosarcoma cases. LFS is caused by a deleterious germline mutation in the *TP53* tumor suppressor gene, and is characterized by early onset of a wide range of cancer types and multiple events of primary cancers throughout one's lifetime (Malkin et al., 1990) (Malkin, 2011). The lifetime risk of cancer for males and females with LFS is over 70% and 90%, respectively, with five cancer types accounting for the majority of cases: osteosarcomas, soft-tissue sarcomas, central nervous system tumors, breast tumors and adrenocortical carcinomas (Schneider et al., 2019). LFS patients are also at increased risk of other cancers, including (but not limited to) leukemias, lymphomas, lung, skin, prostate, ovary, gastrointestinal, thyroid (Schneider et al., 2019). Although LFS is classically described as a familial syndrome, recent estimates have demonstrated that the de novo mutation carrier (DNM) rate in LFS is up to 48% (Gao et al., 2020; Gonzalez et al., 2009). However, confirming the DNM status, DNM or familial mutation carrier (FM), of an LFS patient requires genetic testing for the same deleterious *TP53* mutation in both parents which is not always possible, (parents deny testing, already dead, FH for insurance), making de novo LFS patients a population that has been challenging to study. There is a critical need for methods that can be used to predict DNM status using the patient and family history collected in a genetic counseling session. In the following sections, we will dive into the clinical diagnosis, testing and management of patients with LFS, in order to further understand the complexity of the syndrome, and why sophisticated statistical techniques are required for appropriate modeling.

Diagnosis and testing of Li-Fraumeni

A patient is diagnosed with LFS if they meet all three requirements of the classic LFS criteria (Mai et al., 2012), or if they test positive for a deleterious germline mutation in *TP53* (Schneider et al., 2019). The classic LFS criteria is:

1. Sarcoma diagnosed before the age of 45
2. A first-degree relative with a cancer diagnosed before the age of 45
3. A first or second-degree relative with any cancer diagnosed before the age of 45 or sarcoma at any age

A proband that does not meet the classic LFS criteria, should still be suspected for LFS if they meet one of the three following requirements:

1. The proband meets the Chompret Criteria 2015 (Bougeard et al., 2015)(Valdez et al., 2017). A proband is said to meet the Chompret Criteria 2015 if he meets any of the four following criteria:
    a. Criteria 1- The proband has had both of the following:
        i. A tumor in the LFS spectrum before the age of 46
        ii. One first or second degree relative with either of the two:
            1. A tumor in the LFS spectrum before the age of 56
            2. Multiple primary cancers
    b. Criteria 2 – The proband has had multiple primary cancers in the LFS spectrum of which at least 1 was diagnosed before age 46. Multiple breast tumors do not count as multiple primaries.

    c. Criteria 3 – The proband has been diagnosed with adrenocortical carcinoma, choroid plexus tumor or embryonal anaplastic rhabdomyosarcoma.
    d. Criteria 4 – the proband is a female who was diagnosed with breast cancer before age 31

2. The proband presents with pediatric hypodiploid acute lymphoblastic leukemia (age ≤ 21) (Holmfeldt et al., 2013).
3. The proband has a somatic tissue (such as a tumor) with a deleterious *TP53* mutation with variant allele fraction (VAF) close to 0.5 or higher.

The probands who meet any of the three prior criteria, should be suspected for LFS and tested for a germline mutation in *TP53* through single-gene testing or multigene panel-testing. In single-gene testing, the DNA sequence of the *TP53* gene is verified for missense mutations, nonsense mutations, splice variants or small insertion/deletions (indels) events (Schneider et al., 2019). If analyzing the DNA sequence of the *TP53* reports a wild-type sequence (negative for variants), then suspicion must arise for large genomic events such as large deletions or duplications, and genetic testing to rule out those must follow (Schneider et al., 2019). If testing for LFS is done using multi-gene panels, the aforementioned panel must include testing for large deletion/duplication events in *TP53*, in addition from the standard gene sequence analysis (Schneider et al., 2019).

Somatic *TP53* variants and/or clonal hematopoiesis of indeterminate potential (CHIP) are common confounders of genetic testing for germline *TP53* variants (Weitzel et al., 2018). A test is most likely a false positive in cases the patient does not meet the classic LFS criteria, the genetic testing was performed using a multigene panel, and the variant allele frequency of the identified mutation was below 0.2 (Weitzel et al., 2018).

Common findings in the patient’s disease history that should raise suspicion for a somatic variant or CHIP also include exposure to carcinogens, such as cigarette smoking or chemotherapy, and leukemia or other current malignancy (Weitzel et al., 2018). In the cases were a somatic *TP53* variant or CHIP is suspected, Weitzel recommends several additional genetic tests to confirm the germline status of the mutation including: testing of additional tissues, for example skin fibroblasts (eyebrow plucks) or saliva; genetic testing of additional family members including the proband’s offspring or additional affected family members.

Management and surveillance of patients with LFS

Due to the high lifetime risk of cancer, patients with LFS need to constantly undergo comprehensive evaluations for cancer (Schneider et al., 2019). Surveillance protocol includes a complete physical exam and whole body MRI every six months (adults) or every four months (pediatric patients) (Kratz et al., 2017; Villani et al., 2016). Whole body MRI has shown significant improvement in outcomes (Anupindi et al., 2015; Villani et al., 2016), although at the expense of increased false positive findings (Ballinger et al., 2017). Constant whole-body MRI has been shown to increase the feeling of control and hope in some patients, but can increase stress and burden in others (McBride et al., 2017). Targeted, cancer-specific screening guidelines are also suggested including: abdominal and pelvic ultrasound for early identification of adrenal corticoid carcinoma and sarcomas; clinical breast exam, mammogram and breast MRI for early identification of breast cancer; upper and lower endoscopy to identify gastrointestinal cancers; dermatologic exam to identify melanoma; neurological exam and brain MRI to identify central nervous system tumors (Kratz et al., 2017; Macfarland et al., 2019; Villani et al., 2016)

Avoidance of all known carcinogens (sun exposure, tobacco smoking) is recommended (Schneider et al., 2019). Treatment of malignancies follows standard protocol for the specific malignancy, however, radiation therapy is avoided if possible (Schneider et al., 2019). Patients with breast cancer are encouraged to undergo bilateral mastectomy instead of lumpectomy to reduce risk of recurrence, contralateral breast malignancy, and reduced exposure to radiation therapy (Schon and Tischkowitz, 2018). LFS patients that have not had breast cancer can also consider bilateral mastectomy as a preventive measure to reduce risk of breast cancer (Schon and Tischkowitz, 2018).

Estimated contribution of de novo mutations to Li-Fraumeni Syndrome

Although classically described as an inherited syndrome with strong family history as a requirement for the diagnosis, it is now well understood that de novo mutations (DNMs) are a frequent cause of LFS (Bougeard et al., 2015; Gao et al., 2020; Gonzalez et al., 2009; Renaux-Petel et al., 2018). The initial reports on the contribution of DNMs in LFS opted to estimate a lower bound for the DNM rate. Gonzalez et al., 2009 reported 75 patients ascertained for early-onset breast cancer with a mutation in *TP53* of which 15 lacked family history of cancer and were suspected to be DNMs. Although 15 were identified as highly likely DNM, genetic testing was only available in 5 cases, all of which were confirmed DNMs, and reporting an DNM rate of 5% - 20% (Gonzalez et al., 2009). Similarly, Reneaux-Petel et al., 2018, reported 40 DNMs among 336 unrelated *TP53* mutation carriers, for a DNM rate of at least 14%. Amore recent DNM rate estimate by Gao et al., 2020, controlled for bias in ascertainment criteria on four historical US cohorts by focusing on patients with early-onset breast cancer and patients ascertained due to multiple primary cancers at similar ages, and reported an ascertainment corrected DNM rate estimates of up to 48% (Gao et al., 2020).

Although DNMs are a well understood cause of LFS, the DNM status of an LFS patient is often unknown, as it requires genetic testing of both parents which is often unavailable and inaccessible (Gao et al., 2020; Gonzalez et al., 2009; Renaux-Petel et al., 2018). Conclusions and analysis regarding DNMs have been historically extrapolated from the very small sample size available of confirmed cases (Gonzalez et al., 2009; Renaux-Petel et al., 2018). The difficulties of drawing conclusions from this limited sample size is further complicated by the complexity of LFS and the distinct ascertainment practices of different cohorts, making DNM carriers of *TP53* mutations in LFS an understudied population. Due to the understudied nature of carriers of DNMs in *TP53*, it is not known whether LFS patients in this population have the same disease presentation as that classically described for FM carriers, or whether current ascertainment practices are sufficiently encompassing for all patients who present due to DNMs in *TP53*. Because of the need for early, ongoing surveillance and genetic counseling in these families, it is also of clinical interest to identify DNMs who might otherwise go unnoticed for several generations (Gao et al., 2020).

Mendelian risk prediction modeling of inherited cancer syndromes

Because our goal is to estimate the probability of a genotype, the genotype in this case being DNM in *TP53*, using the family and disease history collected in a routine genetic counseling session, using a Mendelian risk prediction modeling approach is the natural choice (Chen et al., 2004). Mendelian models have been widely and successfully used in the past to accurately predict the genotype of a counselee in several inherited cancer syndromes including: breast and ovarian (Euhus et al., 2002), pancreatic (Wang et al., 2007), melanoma (Wang et al., 2010), gastrointestinal (Chen et al., 2006) and LFS (Peng et al., 2017). Recently, the Mendelian risk prediction model Famdenovo was developed to predict DNM status based on cancer-family history and several input genetic parameters (Gao et al., 2020). The Famdenovo framework was used to build two models, Famdenovo.*TP53* and Famdenovo.BRCA, to predict DNM status in families of LFS and hereditary breast and ovarian cancer (HBOC), respectively. The good performance of these two models was validated in LFS families from four US cohorts (Famdenovo.*TP53*) and families with HBOC from the Cancer Genetics Network (Famdenovo.BRCA).

Mendelian risk prediction models require as input several genetic parameters that are specific for the gene and disease of interest (Chen et al., 2006, 2004; Gao et al., 2020; Peng et al., 2017; Wang et al., 2010, 2007). The first genetic parameter would be the allele-frequency, which is necessary to calculate the founder probabilities. Next, is the de novo mutation rate, which is necessary to account for de novo mutations in the transmission probabilities. The transmission probabilities can otherwise be determined using Mendel's Laws of Heredity if the mode of inheritance is known. Finally, we need as input the disease-gene penetrance, which is the probability distribution of the latent time to disease given a genotype. Because of the broad range of cancer types in LFS, and because multiple events of cancer are common, disease-gene penetrance estimates in LFS require sophisticated statistical modeling and high-quality data. Fortunately, all the previously described inputs have been estimated previously for LFS (Gao et al., 2020; Gonzalez et al., 2009; Peng et al., 2017; Shin et al., 2020b, 2020a).

The need for a Mendelian model to predict de novo status

Due to the complexity of the LFS disease presentation and the prevalence of cancer in the general population, when the parental genotypes are unknown, predicting whether a patient with LFS is a DNM or FM is not trivial. Physicians can at most, rely on simple criteria to decide to decide their whether other family members are at also risk for disease. Mendelian modeling approaches have shown to exceed the predictive power of clinical criteria for LFS, in the context of diagnosis of LFS and also DNM status prediction (Gao et al., 2020; Peng et al., 2017). By having a robust AUC of 0.95 Famdenovo.*TP53* demonstrated great discrimination between DNMs and FMs. However, an OE ratio for perfectly calibrated model should be 1, which is not included in the confidence interval for Famdenovo.*TP53* OE ratio of 1.332 ([1.093, 1.633]) (Gao et al., 2020). A model calibration that is not maximized means that we have less confidence in the estimated probability and this limits the use of the model in research and clinical decision making.

To understand why model calibration and accuracy are important (especially in the clinical setting), we have to first consider that the historical LFS cohorts typically require collection of extensive pedigree data validated by more than one family member, which aids in making the input data less noisy, in comparison to data that would be collected in, for example, a single genetic counseling session. Moreover, strong discrimination means that we can identify an optimal cut off that separates FMs and DNMs well, although this cut off might not be known for a new data set, and it need not be intuitive. In noisier input data, more variability in the output predicted probability would be expected, especially in a less accurate and less calibrated model, which would then pose an issue when interpreting this value. In other words, we can think of the cut-off for maximal discrimination in a less calibrated and less accurate model as a moving. Great discrimination means that hitting the center of the target gets us perfect labels, increasing the accuracy makes the target move less (easier to hit), and higher calibration makes the center of the target bigger. Therefore, we are interested in developing a model that equals or exceeds the sensitivity and specificity of Famdenovo.*TP53*, while also having an improved model calibration and accuracy.

In order to develop an improved model, one possible approach is to feed more information to the probability calculation through the input penetrance estimates. The Famdenovo.*TP53* model uses penetrance estimates that are based on the time to first cancer diagnosis, but does not consider the type of cancer diagnosed (Gao et al., 2020; Peng et al., 2017). Because LFS is a disease were patients can be diagnosed with a wide range of cancer types, and because the age at diagnosis is somewhat dependent on the type of cancer developed, we hypothesize that by incorporating penetrance estimates that are specific for different types of cancers diagnosed (Shin et al., 2020b), we can generate a model with further improved calibration, accuracy and prediction.

Developing an improved model to predict DNM status has significance in both the research and clinical setting. In the research setting, identifying DNMs would allow large sequencing studies of DNMs, which, in turn, would help understand what risk factors are associated with acquisition of deleterious DNMs in TP53 and what genomic events downstream of a mutation in TP53 lead to tumorigenesis. Studying DNMs would also allow us to understand whether these patients have the same clinical phenotype as familial mutation carriers, whether they present a more diverse phenotype, or perhaps an attenuated phenotype. A different clinical phenotype would imply that these patients have different risk stratification, and might require different clinical management. In such case, where DNMs have a clinical picture that is different from

FMs, our new model would then become a tool that could aid in risk stratification and management of LFS patients, in addition to a tool that can be used to identify these patients for research purposes. If through research of DNMs we find that they do not have a different clinical phenotype from FMs, or if it simply found that they do not benefit from a different management, then our method could still be used to encourage identification of other family members at high risk of disease. For example, a LFS patient with a low probability of being DNM, and high probability of being FM, could be counseled to reach out to assymptomatic parents, siblings and/or nephews who might have be at high risk of a mutation in TP53 and would therefore benefit from close follow up.

Chapter 2 – Model development and evaluation

Cancer-specific Famdenovo model

In order to model such a complicated disease like LFS, sophisticated models are needed, and have been shown to outperform simple clinical criteria (Gao et al., 2020; Peng et al., 2017). To further improve upon current state of the art mendelian models used to predict DNM status we've developed a model that by incorporates penetrance estimates that are specific for different types of cancers diagnosed (Shin et al., 2020b). Because specific types of cancer are more likely to develop at a particular range of ages, these penetrance estimates should increase the information provided to the model, resulting in improved model performance. The increased accuracy and calibration of this cancer-specific model, should prove invaluable for research and clinical decision making.

The Cancer-Specific Famdenovo model (Famdenovo.CS) estimates the probability of a confirmed deleterious germline mutation being a *de novo* mutation on a counselee of interest. The probability calculation is based on the proband and his/her family member's disease history, as well as several input parameters. The input parameters needed for the model are the disease-gene penetrance, the allele frequency and de novo mutation rate. Let ***P*** be the family information, ***D*** be the disease information for the whole pedigree, ***G_c*** the genotype of the counselee, ***G_m*** the genotype of the mother, ***G_f*** the genotype of the father. We calculate the counselee's *de novo* mutation probability, $\Pr(G_c \text{ is } de\ novo \mid G_c \text{ is germline}, \mathbf{D}, \mathbf{P})$, as follows:

$$\Pr(G_c \text{ is } de\ novo \mid G_c \text{ is germline}, \mathbf{D}, \mathbf{P})$$

$$= \Pr\big(G_m = 0, G_f = 0 \,\big|\, G_c = 1, \mathbf{D}, \mathbf{P}\big)$$

$$= \frac{\Pr(G_c = 1 | G_m = 0, G_f = 0, \mathbf{D}, \mathbf{P}) \times \Pr\big(G_f = 0 \mid \mathbf{D}, \mathbf{P}\big) \times \Pr\,(G_m = 0 \mid \mathbf{D}, \mathbf{P})}{\Pr\,(G_c = 1 \mid \mathbf{D}, \mathbf{P})} \qquad (1)$$

where *G_m* and *G_f* are the genotype of the mother and father, respectively. We then apply a Mendelian modeling approach to derive the four probabilities in equation (1). Define **H** = (**P, D**) as the cancer history of the whole family. Let G_0 denote the genotype of a person of interest. In this case, G_0 can be G_m, G_f, or G_c. We calculate the probability $\Pr(G_0| \mathbf{H})$, the probability of a genotype given the family history, by updating the population prevalence, $\Pr\,(G_0)$, after incorporating the family and cancer history, ***H***. We can estimate it via the following formula:

$$\Pr(G_0|\mathbf{H}) = \Pr(G_0|\mathrm{H}_0, \mathrm{H}_1, \dots, \mathrm{H}_\mathrm{n}) = \frac{\Pr(G_0)\Pr(\mathbf{H}|G_0)}{\sum G_0 \Pr(G_0)\Pr(\mathbf{H}|G_0)} \quad (2)$$

$$\Pr(\mathbf{H}|G_0) = \sum\nolimits_{G_1, G_2, \dots, G_n} \Pr(\mathbf{H}|\boldsymbol{G})\Pr(G_1, G_2, \dots, G_n\,|G_0) =$$

$$\sum\nolimits_{G_1, G_2, \dots, G_n} \left[\prod\nolimits_{i=0}^{n} \Pr\,(\mathrm{H}_i|G_i)\right] \Pr(G_1, G_2, \dots, G_n\,|G_0). \quad (3)$$

Where *n* refers to the total number of individuals in the family pedigree of the counselee. $\Pr\,(\mathbf{H}|G_0)$ is the probability of the phenotypes for all family members, given the genotype of the counselee, which is calculated as the weighted average of the probabilities of family history given all the possible genotype configurations of all family members $\Pr\,(\mathbf{H}|\mathbf{G})$. The weights, $\Pr(G_1, G_2, \dots, G_n\,|G_0)$, can be estimated based on the probabilities of the genotype configuration that are given by the rules of Mendelian transmission. After assuming conditional independence, $\Pr\,(\mathbf{H}|\mathbf{G})$ are the products of all individual probability distributions of the penetrance $\Pr\,(H_i|G_i)$.

We calculate the posterior probability using the Elston-Stewart peeling algorithm (Elston and Stewart, 1971; Fernando et al., 1993). This algorithm uses a transmission matrix of the probability of the genotype of an individual, given the genotypes of his/her father and mother, $\Pr(\,G_s\,|\,G_f,\ G_m\,)$, to characterize Mendelian transmission (Elston and Stewart, 1971; Fernando et al., 1993).

Prevalence and de novo mutation rate

The prevalence of pathogenic *TP53* mutations (allele frequency) is specified as 0.0006, and was derived in previous studies (Peng et al., 2017). We assumed that the mutated *TP53* allele follows Hardy-Weinberg equilibrium. The frequencies for each genotype were 0.9988 for homozygous reference, 0.001199 for heterozygous, and 3.6x10-07 for homozygous variant. The assumption of Hardy-Weinberg equilibrium can be modified using user input if updated information is published regarding the homozygous genotype. We used 20% for the DNM rate among all mutations (Gao et al., 2020; Peng et al., 2017). Both priors, the allele frequency and the DNM rate, are then updated by family history in the posterior probability calculation. All priors were validated on independent, external study cohorts, that are not part of this study (Peng et al., 2017).

Cancer-specific penetrance

On Famdenovo.CS, we utilize cancer-specific estimates previously estimated in Seung Jun et al 2019. Seung Jun et al 2019 employed a Bayesian semi-parametric competing risk model that incorporates the family pedigree structure efficiently into the penetrance estimation and corrects for ascertainment bias, thereby also increasing the effective sample size in this rare population of LFS families (Shin et al., 2020b). First, let X account for the individual's sex (male, female), $\mathrm{T_k}$ denotes the time of the $\mathrm{k_{th}}$ type of event; where $k = 1, 2, 3, 4$ represents breast cancer, sarcomas, other cancers and death not related to cancer, respectively. We then define $\mathrm{T} = \min\limits_{\mathrm{k}}(\mathrm{T}_k)$ and $\mathrm{C} = \mathrm{observed(k)}$. If we observe the $\mathrm{k_{th}}$ type of event at time T, the cancer-specific penetrance $q_k(t)$ is:

$$q_k(t = T|\, G, X\,) = \; P(T < t, C = k\, |G, X) = \; \int_0^t \lambda_k(t \mid G, X\,) S(u \mid G, X) du \quad (4)$$

Where,

$$\lambda_k(t \mid G, X\,) = \lim_{h \downarrow 0} \frac{\Pr(t \leq T < t+h, C=k \mid T>t, G, X\,)}{h} \quad (5a)$$

$$\lambda_k(t \mid G, X\,) = \lambda_{0,k}(t)\xi \exp\,(\beta_1 G + \beta_2 X + \beta_3 G \times X) \quad (5b)$$

Where $\lambda_{0,k}(t)$ denotes the baseline hazard ratio and ξ denotes a family specific random frailty.

$$S(u \mid G, X) = \exp\{-\textstyle\sum_{k=1}^{K} \Lambda_k(t|G, X)\} \; (6)$$

$$\Lambda_k = \int_0^t \lambda_k\,(u|G, X)du \; (7)$$

The cancer-specific penetrance estimates from Seung Jun et al 2019 are available through the LFSPRO R-package (Peng et al., 2017) and were used on the Famdenovo.CS model to calculate the $\Pr\,(H_i|G_i)$ terms as follows. Let ∅ represent not observing any of the k_{th} events (null set).

For an individual with genotype G_i **presenting $\boldsymbol{k_{th}}$ event at age $\boldsymbol{T}$**,

$$\Pr(H_i|G_i) = P(C = k,\, T = t|\, G_i,\, X_i\,) = \, q_{k=C}(t + h \mid G_i\,,\, X_i) - \, q_{k=C}(t \mid G_i,\, X_i)$$

$$= \int_T^{T+h} \lambda_k(t = \tau|\, G_i,\, X_i) S(t = \tau | G_i,\, X_i) d\tau \; (8)$$

For an individual with genotype G_i who is **asymptomatic by age $\boldsymbol{T}$**,

$$\Pr(H_i|G_i) = P(C = no\ event,\, t = T|\, G_i,\, X_i\,) = 1 - \textstyle\sum_{k=1}^{4} q_{k=C}(t = T \mid G_i,\, X_i)$$

$$= 1 - \textstyle\sum_{k=1}^{4} \int_0^T \lambda_k(t = \tau|\, G_i,\, X_i) S(t = \tau | G_i,\, X_i) d\tau \; (9)$$

Study cohorts

We evaluated our method using data from five different cohorts of LFS families (**Table 1**). **(A)** The MD Anderson cohort (MDA) includes families that prospectively followed and were initially ascertained because they met the classic LFS criteria. These families were identified by trained MD Anderson personnel who screened for potential subjects by inspecting the electronic medical record system, patient referrals (from the institution or outside referrals), patient clinics, patients census and surgery schedules. Patients eligible for the study were contacted (with approval from the attending physician) to determine if they were interested in participating in the study. For our study, we limit ourselves to patients with a confirmed deleterious mutation in *TP53*, for a total of 140 families. Of these 140 families, we have a confirmed de novo status for 82 families, where we know the inheritance pattern of the mutation, and 58 families with a de novo status that is unknown (Bougeard et al., 2015; Chompret et al., 2001; Li et al., 1988; Shin et al., 2020a).

(B) A long term prospective cohort of LFS families (NCT01443468) collected by The National Cancer Institute (NCI) (Mai et al., 2016). The probands included in the NCI cohort were ascertained for one of the following: meeting classic LFS criteria, meeting Li-Fraumeni-like diagnostic criteria, they have a pathogenic germline mutation in *TP53*, have a first- or second-degree relative with a pathogenic mutation in *TP53*, or they have been diagnosed with adrenocortical carcinoma, choroid plexus carcinoma, or more than two primary events of cancer (Birch et al., 1994). Of the total 78 families included in this cohort, we know the de novo status for 66 families (12 families with unknown *de novo* status).

(C) A cohort of patients with LFS collected by the Dana Farber Cancer Institute (DFCI) through clinical genetics practice. Patients eligible for this study must have tested positive for a pathogenic mutation in *TP53*, meet LFS criteria or have a family member who meets LFS criteria, be an obligate carrier of a pathogenic mutation in *TP53*, or have been previously diagnosed with LFS. Patients in this cohort were initially only included people tested through single-gene testing, but since 2012 now also includes patients who tested positive for a pathogenic mutation in *TP53* through multi-gene panels (Gao et al., 2020). The patients included in this cohort also accepted enrollment into a surveillance protocol through whole-body MRI. The DFCI consists of 91 families of which 30 have a known *de novo* status.

(D) A cohort of patients from the Children's Hospital of Philadelphia (CHOP). This cohort includes cases ascertained to provide genetic counseling to the pediatric population found to have a genetic predisposition to cancer (Cancer Predisposition Program). Patients ascertained through this program include those with a family history of inherited cancer syndromes, children with an incidental finding of a deleterious mutation involving a cancer predisposition gene, children that present with an adult cancer (for example, soft tissue sarcomas), children found with a tumor that is associated with hereditary component such as adrenal corticoid carcinoma and children with multiple primary cancers. The CHOP cohort includes a total of 15 families, of which 8 have a known *de novo* status.

(E) Data from the Clinical Cancer Genetics program at MD Anderson Cancer Center (CCG-MDA) that is comprised of patients that are seen for genetic counseling through the Clinical Cancer Genetics (CCG) Department at MD Anderson Cancer Center. Personal and family history are collected in a counseling session and entered into a progeny database for tracking through the CCG department. This database includes patients counseled starting at year 1975. There is active accrual of patients currently being seen. For this study, patients that were identified to have a pathogenic or likely pathogenic mutation in *TP53* through single-gene testing or multi-gene panel were included. This cohort includes a total of 124 families, of which we know the de novo status of 25 families.

Table 1. *Overview of the LFS families in the five cohorts included in our study.*

Cohort	MDA	CHOP	DFCI	NCI	CCG
# Families	**140**	**15**	**91**	**78**	**101**
Largest Family	**151**	**45**	**107**	**130**	**75**
Smallest Family	**4**	**13**	**3**	**3**	**7**
Family Size - mean	**45**	**26**	**34**	**24**	**31**
Family Size - standard deviation	**31**	**12**	**25**	**17**	**14**
Age of diagnosis (AoD) - mean	**42.3**	**43.3**	**38.9**	**33.9**	**44.4**
AoD - standard deviation	**22.2**	**25.2**	**20.5**	**22.1**	**19.3**

Evaluation criteria

Combining the five study cohorts, we have a combined 211 families with a known de novo mutation status. However, for the MDA, NCI, DFCI and CHOP cohorts, it was common for each family with known de novo status to have more than one person with a *TP53* mutation and known de novo status. Although Famdenovo.CS estimates a de novo mutation probability per *TP53* mutation carrier, individuals belonging to the same family are not independent of each other. Therefore, for the families with known DNM status in MDA, DFCI, CHOP, NCI we estimated one family-wise DNM probability. On the other hand, this was never the case with the CCG cohort, where we have at most 1 person with known DNM status per family. We therefore divided the validation set of 211 families in two:

Validation Set A (VSA) – 186 families with possibly more than 1 individual with a mutation in *TP53* and with known DNM status. Since *TP53* mutation carriers in the same family are not independent of each other, we also used family-wise DNM or FM labels. A family was classified as DNM if it had at least one *TP53* mutation carrier that was confirmed to be DNM, otherwise, they were classified as FM. When applying the Famdenovo.CS model in this validation set, we used family-wise *de novo* probability and classification for evaluation, where we classified a family as DNM if at least one of the family members had a DNM probability over the a given cut-off, otherwise, the family was classified as FM.

Validation Set B (VSB) – 25 families with one proband with a known de novo status included in the CCG-MDA cohort. An individual in this validation set was considered de novo if both parents tested negative for *TP53* mutation. If either parent tested positive for *TP53* mutation, then the individual would be classified as familial. In total, in the CCG-MDA cohort, we had 20 families with known *de novo* status after filtering family units that were missing information required to apply the Famdenovo.CS model. Reasons for removing a family from the VSB were family unit with less than 4 members or lack of information for any member of the family besides the proband.

For both validation-sets, VSA and VSB, we used Famdenovo.*TP53* and Famdenovo.CS to estimate the DNM probability on all individuals with known DNM status. We used the concordance index (AUC), observed:expected ratios (OE) and Brier score (BS) to measure our model's discrimination, calibration and accuracy, respectively. A high AUC indicates that we can find a cut off value on the Receiver Operating Characteristic curve (ROC) corresponding to predictions with high sensitivity and specificity. The OE is the ratio between the number of observed true positive, in this case DNMs, and the sum of all of the estimated DNM probabilities. An OE = 1 indicates perfect calibration, where the number of estimated DNMs and the number of observed DNMs are equal. A perfect BS is indicated by the value 0, where the probability estimate is always 1 for DNMs and 0 for FMs. We estimated 95%

confidence intervals (CI) on our summary measures based on 1,000 bootstraps for each validation set. All of our analysis was performed using the open-source environment R (http://cran.r-project.org).

Mutation testing

Mutation testing for the MD Anderson cohort was done using blood samples collected from probands who had provided their informed consent. To determine mutation status, PCR sequencing of exons 2-11 of the *TP53* gene was performed (Hwang et al., 2003). If the proband was positive for a mutation in *TP53*, all first-degree relatives and all other family members who were at risk of carrying the mutation were also tested, even if they had not been affected by cancer. Since the extension of germline testing was done based on mutation status and not disease history or phenotype, this should not introduce bias into our analysis (Katki et al. 2008). If the proband tested negative for a mutation in *TP53*, other family members were not tested.

For the NCI cohort, individuals could be tested prior to or during enrollment. If an individual was tested prior to enrollment, the study team obtained copies of the clinical reports for the *TP53* mutation tests and verified them prior to enrolling the individual in the trial. If an individual was enrolled in the trial and was not tested previously, then genetic testing was performed after enrollment. If probands tested positive for a deleterious mutation in *TP53* before or after enrollment, at risk family members were offered site-specific genetic testing through the study. Testing was not offered to family members of probands who tested negative for deleterious mutation in *TP53*. The NCI cohort mutation testing included detection of large genomic events, such as deletions or large rearrangements (Gao et al., 2020).

The DFCI cohort used the Clinical Operations and Research Information System (CORIS) to search for patient information and identified eligible families who met the classic or updated Chompret criteria (Li and Fraumeni Jr, 1969; Rath et al., 2013; Tinat et al., 2009). Mutation testing for *TP53* in the DFCI cohort was done using exon aggregation analysis (EGAN) (Rath et al., 2013).

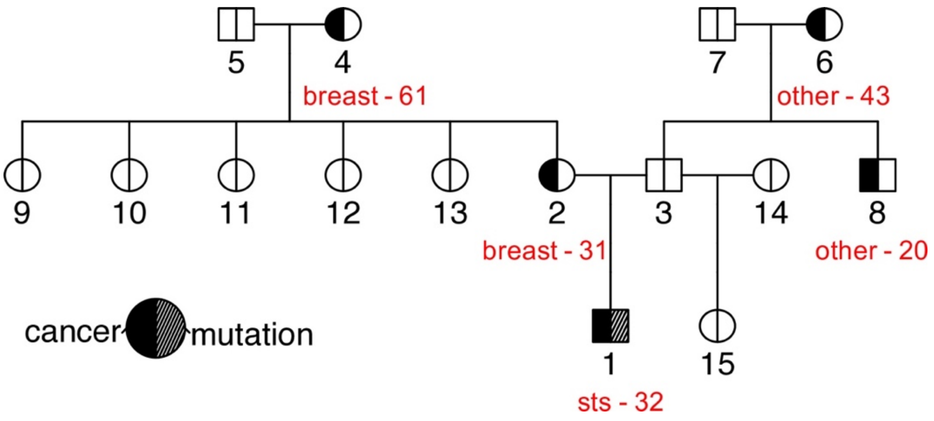
For the CHOP cohort, patients with a clinical history of pediatric cancer with primary tumors in the LFS spectrum, such as adrenal corticoid carcinoma, and patients that met classic or Chompret criteria were tested for *TP53* mutation in Ambry Genetics, The Hospital for Sick Children or the Genetic Diagnostic Laboratory of the University of Pennsylvania.

Probands in the CCG-MDA cohort were tested either via single-gene *TP53* testing or multigene panel tests that included *TP53*. Testing was performed in several CLIA/CAP certified laboratories. Most probands met the Classic or Chompret LFS criteria. Patients that did not meet Chompret or Classic LFS criteria were tested either because of clinical suspicion from a certified genetic counselor or they were identified on panel testing performed on suspicion for other hereditary cancer syndromes. Family members of the confirmed *TP53* mutation carrier were not required to undergo additional testing, however recommendations for family member testing were made during standard of care genetic counseling sessions.

Inputs and output of the cancer specific Famdenovo model

Famdenovo.CS requires 4 inputs. First, a vector of IDs of the individuals you want to analyze. These individuals need to be confirmed germline *TP53* mutation carriers. In the illustrative example in **Figure 1**, the individual with ID=1 is the only confirmed *TP53* mutation carrier and thus, the sole member of the input vector.

Figure 1. *Example pedigree*



Second, Famdenovo.CS requires a data frame with the family information, where each row describes a family member. Please see an example of an input family data on **Table 2**. This data frame consists of 5 columns:

1. id = Index of the person this row corresponds to.
2. fid = index of the father of the person this row corresponds to.
3. mid = index of the mother of the person this row corresponds to.
4. gender = biological sex of the person this row corresponds to.
    a. 0 = female
    b. 1 = male
5. age = age of the person this row corresponds to. Current age if the person is alive and age of death otherwise.

Table 2. *Example input of family data to Famdenovo.CS*

id	fid	mid	gender	age
1	3	2	1	32
2	5	4	0	35
3	7	6	1	53
4	NA	NA	0	77
5	NA	NA	1	81
6	NA	NA	0	78
7	NA	NA	1	81
8	7	6	1	47
9	5	4	0	20
10	5	4	0	20
11	5	4	0	20
12	5	4	0	20
13	5	4	0	20
14	NA	NA	0	20
15	3	14	0	17

Thirdly, Famdenovo.CS requires a data frame with the cancer information for the whole family being analyzed (**Table 3**). There should be one row for every event of cancer. The columns in this table consist of:

1. id = index of the person who was diagnosed with this event of cancer
2. cancer.type = type of cancer diagnosed. The type of cancer should be one of 11 cancer types according to the National Comprehensive Cancer Network guidelines version 1.2012 Li-Fraumeni Syndrome criteria.

3. diag.age = age of the individual when he diagnosed with this event of cancer.

Table 3. *Example input of cancer history to Famdenovo.CS*

id	cancer.type	diag.age
1	sts	32
2	breast	31
4	breast	61
6	non.lfs	43
8	non.lfs	20

Fourthly, Famdenovo.CS requires a data frame with the mutation information of everyone who has undergone genetic testing for a pathogenic germline mutation in *TP53* (**Table 4**). The table should include two columns:

1. id = index of this person
2. mut.state = genotype of this person.
    a. M = tested positive for a pathogenic germline mutation in *TP53*.
    b. W = tested negative for a pathogenic germline mutation in *TP53*.

Table 4. *Example of input mutation data for Famdenovo.CS*

id	mut.state
1	M

The output of Famdenovo.CS is a data frame containing 2 columns (**Table 5**).

1. id = the index of the person analyzed
2. prob.denovo = the calculated de novo probability

On this example Famdenovo.CS showed a very low de novo probability of 0.043. Interestingly, on this same example, the original Famdenovo.*TP53* estimates a 0.0085 de novo probability. Famdenovo.CS actually estimates a probability that is five times larger than the original model, consistent with a positive change in model calibration.

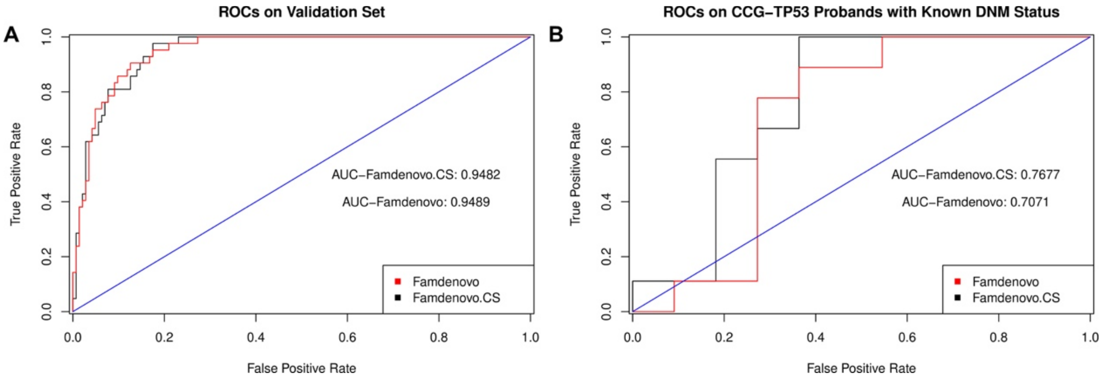
***Table 5**. Example outputs of Famdenovo.CS*

id	prob.denovo
1	0.04251107

Discrimination, calibration and accuracy of the cancer-specific Famdenovo model

***Figure 2.** ROC Curves*

Receiver Operating Characteristic curves of Famdenovo.CS on VSA and VSB. Famdenovo.*TP53* validation curve is provided for comparison.



Famdenovo.CS showed a good discrimination measured by an AUC of 0.95 (95% CI: [0.92 ,1.00]) in VSA and 0.77 (95% CI: [0.50, 0.96]) in VSB. This discrimination capacity was as good as the model with the overall penetrance estimates. Using the cancer-specific penetrance improved the calibration and accuracy of the prediction, demonstrated by the OE ratio and Brier score. The OE Ratio for Famdenovo.CS in VSA is 1.17 (95% CI [0.90, 1.46]), and now the confidence interval now includes 1 which is the optimal value. The OE ratio on VSB was equally as good in Famdenovo.CS as in the original model. In terms of accuracy, the Brier Score was as good in both models on VSA, however, Famdenovo.CS had a significantly decrease in Brier Score in VSB of -0.091 (95% CI: [-0.024, -0.19]), that demonstrates an increase in accuracy.

Table 6. *Comparison of the Cancer Specific Famdenovo model*

Metrics	Famdenovo.*TP53*	Famdenovo.CS	Validation Set
AUC	0.95 [0.92, 1]	0.95 [0.92, 1]	A
OE Ratio	1.48 [1.04, 1.79]	1.17 [0.90, 1.46]	A
Brier Score	0.27 [0.23, 0.33]	0.26 [0.22, 0.31]	A
AUC	0.71 [0.50, 0.93]	0.77 [0.50, 0.96]	B
OE Ratio	1.72 [0.92, 3.5]	1.60 [0.96, 2.6]	B
Brier Score	0.33 [0.17, 0.51]	0.24 [0.13, 0.37]	B

Table 7*. Performance metrics at different cut offs for Famdenovo.CS.*

Cut Off	F1-Score	Sensitivity	Specificity	PPV	NPV
0.2	0.74	0.86	0.87	0.66	0.96
0.25	0.73	0.81	0.88	0.67	0.94
0.3	**0.77**	**0.79**	**0.92**	**0.75**	**0.94**
0.35	**0.77**	**0.79**	**0.92**	**0.75**	**0.94**

0.4	0.76	0.77	0.93	0.76	0.93

We measured several metrics at different cut-offs for the Famdenovo.CS probability to choose the one cut-off that provides the optimal performance. Although robust performance is observed at different cut-off values for the DNM probability, we chose to move forward with 0.35, which grants high PPV, Sensitivity and F1-Score while keeping a strong NPV and Specificity.

Chapter 3 – Analysis of de novo *TP53* mutation carriers

Identification of DNMs and FMs

After validating the improvement in accuracy and calibration of the Famdenovo.CS model, we then applied our method to LFS families with unknown DNM status in order to augment the sample size of classified families and be able to characterize the clinical characteristics of DNM carriers. Previous studies have explored the 324 families available in the NCI, CHOP, MDA and DFCI cohort (Gao et al., 2020). However, 124 families available in the CCG-MDA cohort have not been further analyzed. This study is still actively accruing families, and follows the most recent standard of care practices for LFS, characteristics that should make this cohort more sensitive to identifying DNM patients, and the cohort should be more representative of the general population in LFS. As such, these characteristics make the CCG-MDA cohort an appropriate data set to make inference about the proportions of patients who present due to a de novo mutation, the types of cancer DNMs present, their ages of diagnosis, the most frequent mutations in *TP53* in DNMs, and other available clinical criteria.

Before applying Famdenovo.CS to the CCG-MDA cohort, we first excluded 23 probands due to not having extended pedigree available, or lack of key information such as age of diagnosis, age of last contact and no parents or other ancestors on the family pedigree (**Table 8**). Out of the remaining 101 families, we have 10 confirmed DNM carriers and 15 confirmed FMs carriers. These were proband whose DNM status had been previously confirmed through DNA testing of the parents. By applying Famdenovo.CS to the remaining 76 mutation carriers with unknown inheritance mode of the *TP53* mutation, and we predict 29 DNMs and 47 FMs. In total, we predict 39 DNMs and 62 FMs (38.6% overall DNM rate).

Table 8. *Effective sample size in CCG-MDA cohort.*

Family/Proband Characteristics	**Count**
Available in CCG-MDA	124
Family members < 4 or no parent information	17
Mosaic *TP53* mutation	1
Other missing information (i.e. missing ages)	5
Final effective sample size	101
Predicted DNMs	39
Predicted FMs	62

Estimation of an unbiased DNM rate

Many of the clinical criterion to ascertain LFS patients depend on family history, therefore, favoring ascertainment of FMs into cohorts. This can make the overall DNM rate previously estimated lower than the true DNM rate. To estimate an unbiased DNM rate, we look at subsets of mutation carriers that were ascertained through unbiased criteria, specifically early onset breast cancer (breast cancer before age 32) or multiple primary cancers (MPC) (**Table 9 and Table 10**). Using the early onset breast cancer criteria, we identify 10 FMs and 11 DNMs (52.3% DNM rate).

***Table 9**. Ascertainment bias correction using patients ascertained due to early onset breast cancers.*

	Overall	Early Onset Breast Cancer
Predicted DNMs	39	11
Predicted FMs	62	10

When we look at the MPC criteria divided into strata according to age of their first primary cancer **(Table 10)**, we see there is a similar ratio when we look at patients whose first primary cancer event was diagnosed before age (53% DNM rate). This ratio decreases as we increase the age of the first primary cancer event, consistent with a decreased chance of ascertaining a DNM carrier without family history who has later onset of cancer. To estimate a DNM rate based on the MPC ascertainment criteria, we apply a weighted average to the 4 strata of MPC probands (**Table 10**), and estimate a 43% DNM rate. Overall, we predict that likely more than 43-52% of *TP53* mutation carriers present due to a DNM: a proportion higher than previously estimated in other studies (Gonzalez et al., 2009) yet consistent with recent unbiased estimates (Gao et al., 2020).

***Table 10.** Ascertainment bias correction using patients ascertained due to multiple primary cancers.*

Age at first primary	0-20	21-40	41-60	60+

Predicted DNMs	9	7	2	0
Predicted FMs	7	15	7	1

Cancer Types and ages of diagnosis observed in DNMs and FMs

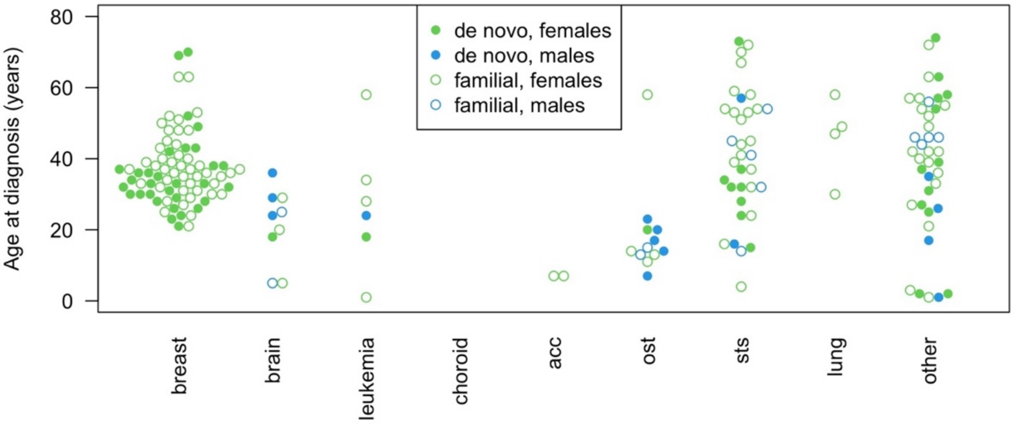
We interrogated the types of cancers diagnosed in the DNM and FM carriers and estimated the odds ratio (OR) of observing a particular cancer type in the DNM group vs the FM group (**Table 11**). Osteosarcoma (OST), breast and brain cancer have an OR > 1, while leukemia and soft tissue sarcomas (STS) had an OR < 1. However, the confidence interval for all these OR include 1, therefore, we conclude all cancer types are equally likely to be observed in either group, and differences observed in the CCG-MDA cohort are most likely due to sample size or ascertainment differences between DNMs and FMs. **Figure 3** shows the distribution of ages of diagnosis for male and female, DNM and FM probands in the CCG-MDA cohort. The distribution of age of diagnosis is similarly distributed amongst DNMs and FMs, after accounting for sex and cancer type. Some cancer types were not seen in DNMs, such as adrenal corticoid carcinoma (ACC) or lung cancer. This finding suggests that DNM carriers presenting these cancer types are not being ascertained to clinical cohorts with the current testing criteria.

***Table 11**. Spectrum of cancer types diagnosed in predicted DNMs and FMs.*

Cancer Type	DNMs	FMs	OR	OR Confidence Interval
ACC	0	2	.	.
Brain	5	5	1.67	[0.465, 5.97]
Breast	31	46	1.16	[0.64, 2.12]

Leukemia	2	4	0.804	[0.143, 4.51]
Lung	0	4	.	.
OST	6	6	1.68	[0.52, 5.42]
STS	17	24	0.62	[0.278, 1.39]
Other	10	24	1.19	[0.589, 2.42]

***Figure 3**. Spectrum of cancer types and corresponding age at diagnosis observed in DNMs and FMs in the CCG-MDA cohort.*



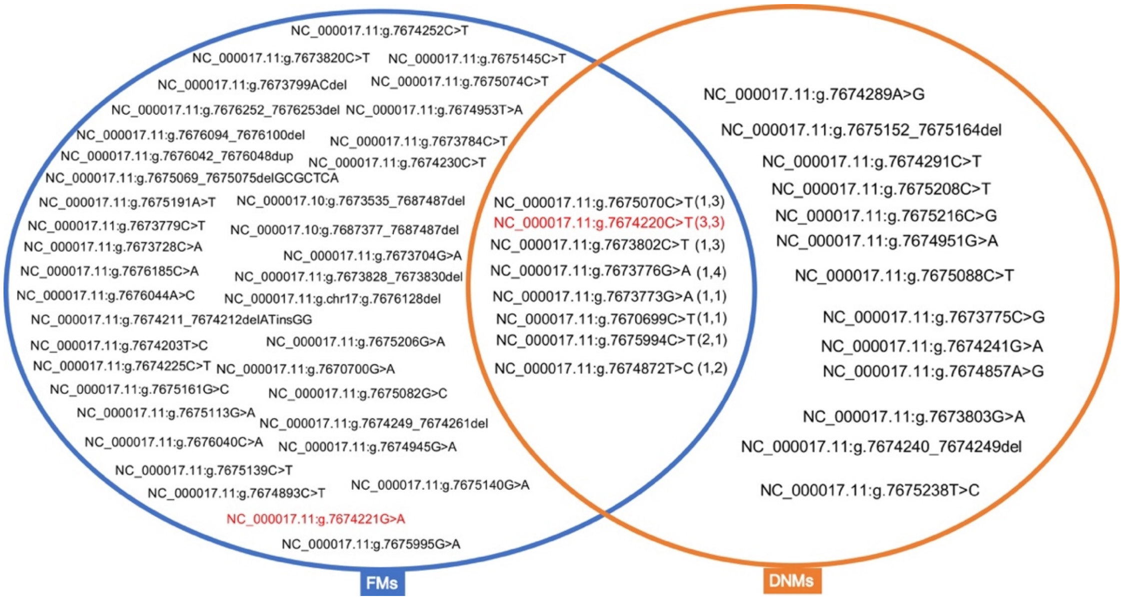
Mutation types observed in DNMs and FMs

A total of 63 different mutations were observed in the CCG-MDA cohort (**Figure 4**). Most mutations were observed only in DNMs or FMs except for 8 mutations that were observed in both groups. Previous studies identified that the mutation NC_000017.11:g.7674221G>A, which causes an amino acid change *TP53*-p.R248W, is not observed in DNMs even though it's in high enough frequency (Gao et al., 2020). This is also the case in the CCG-MDA cohort, where this variant was only found in FMs. Interestingly, if we combine the sample size of NC_000017.11:g.7674221G>A variants in previous studies with the variants found in CCG-MDA, a total of 8 FMs carrying the NC_000017.11:g.7674221G>A have been identified, compared to 0 DNMs. This is a significant observation if we assume a uniform distribution of DNM rate amongst variants (p-value = 0.035, Poisson test). One possible explanation for this finding is that DNMs with the *TP53*-p.R248W variant might not be ascertained under the current clinical criteria because they do not present the classic LFS phenotype, or because the mutation is less penetrant when the carrier is DNM. This could be a result of a genetic anticipation phenomena, that is stronger this particular variant than, for example, on the NC_000017.11:g.7674220C>T mutation that also causes a change on the Arginine 248 residue of the p53 protein (*TP53*-p.R248Q), but that is observed on both DNMs and FMs.

Due to the complex nature of LFS as a disease, and the variability in the presentation, it has been previously hard to definitively answer whether a genetic anticipation phenomena is present in LFS families and even harder to quantify it (Trkova et al., 2002). Due to the large number of deleterious variants in *TP53*, mutation-specific analysis will require a gigantic sample size, especially if we want to somehow quantify the presumed anticipation phenomena and test differences between 2 different pathogenic variants in *TP53*. Of course, this is merely speculative, and there are other possible explanations. Our sample size might currently be too low to identify DNM carriers of the NC_000017.11:g.7674221G>A variant, carriers of this variant might initially present as mosaic mutation carriers with involvement of sperm/oocytes (Azzollini et al., 2020). Regardless, the next step would be to identify more probands with DNMs and FMs in *TP53*, so we can reach an appropriate sample size and be able to perform robust mutation specific analysis.

***Figure 4**. Venn diagram of the deleterious mutations in TP53.*

The mutations are shown in HGVS format. For the mutations observed in both DNMs and FMs, the frequency of mutations observed in each group is shown in parenthesis. The mutations that cause the amino acid changes *TP53*-p.R248W and *TP53*-p.R248Q are shown in red.


NC_000017.11:g.7674252C>T
NC_000017.11:g.7673820C>T
NC_000017.11:g.7675145C>T
NC_000017.11:g.7673799ACdel
NC_000017.11:g.7675074C>T
NC_000017.11:g.7676252_7676253del
NC_000017.11:g.7674953T>A
NC_000017.11:g.7676094_7676100del
NC_000017.11:g.7673784C>T
NC_000017.11:g.7676042_7676048dup
NC_000017.11:g.7674230C>T
NC_000017.11:g.7675069_7675075delGCGCTCA
NC_000017.11:g.7675191A>T
NC_000017.10:g.7673535_7687487del
NC_000017.11:g.7673779C>T
NC_000017.10:g.7687377_7687487del
NC_000017.11:g.7673728C>A
NC_000017.11:g.7673704G>A
NC_000017.11:g.7676185C>A
NC_000017.11:g.7673828_7673830del
NC_000017.11:g.7676044A>C
NC_000017.11:g.chr17:g.7676128del
NC_000017.11:g.7674211_7674212delATinsGG
NC_000017.11:g.7674203T>C
NC_000017.11:g.7675206G>A
NC_000017.11:g.7674225C>T
NC_000017.11:g.7670700G>A
NC_000017.11:g.7675161G>C
NC_000017.11:g.7675082G>C
NC_000017.11:g.7675113G>A
NC_000017.11:g.7674249_7674261del
NC_000017.11:g.7676040C>A
NC_000017.11:g.7674945G>A
NC_000017.11:g.7675139C>T
NC_000017.11:g.7674893C>T
NC_000017.11:g.7675140G>A
NC_000017.11:g.7674221G>A
NC_000017.11:g.7675995G>A
NC_000017.11:g.7675070C>T(1,3)
NC_000017.11:g.7674220C>T(3,3)
NC_000017.11:g.7673802C>T (1,3)
NC_000017.11:g.7673776G>A (1,4)
NC_000017.11:g.7673773G>A (1,1)
NC_000017.11:g.7670699C>T(1,1)
NC_000017.11:g.7675994C>T(2,1)
NC_000017.11:g.7674872T>C (1,2)
NC_000017.11:g.7674289A>G
NC_000017.11:g.7675152_7675164del
NC_000017.11:g.7674291C>T
NC_000017.11:g.7675208C>T
NC_000017.11:g.7675216C>G
NC_000017.11:g.7674951G>A
NC_000017.11:g.7675088C>T
NC_000017.11:g.7673775C>G
NC_000017.11:g.7674241G>A
NC_000017.11:g.7674857A>G
NC_000017.11:g.7673803G>A
NC_000017.11:g.7674240_7674249del
NC_000017.11:g.7675238T>C
FMs
DNMs

Time to first cancer in DNMs and FMs

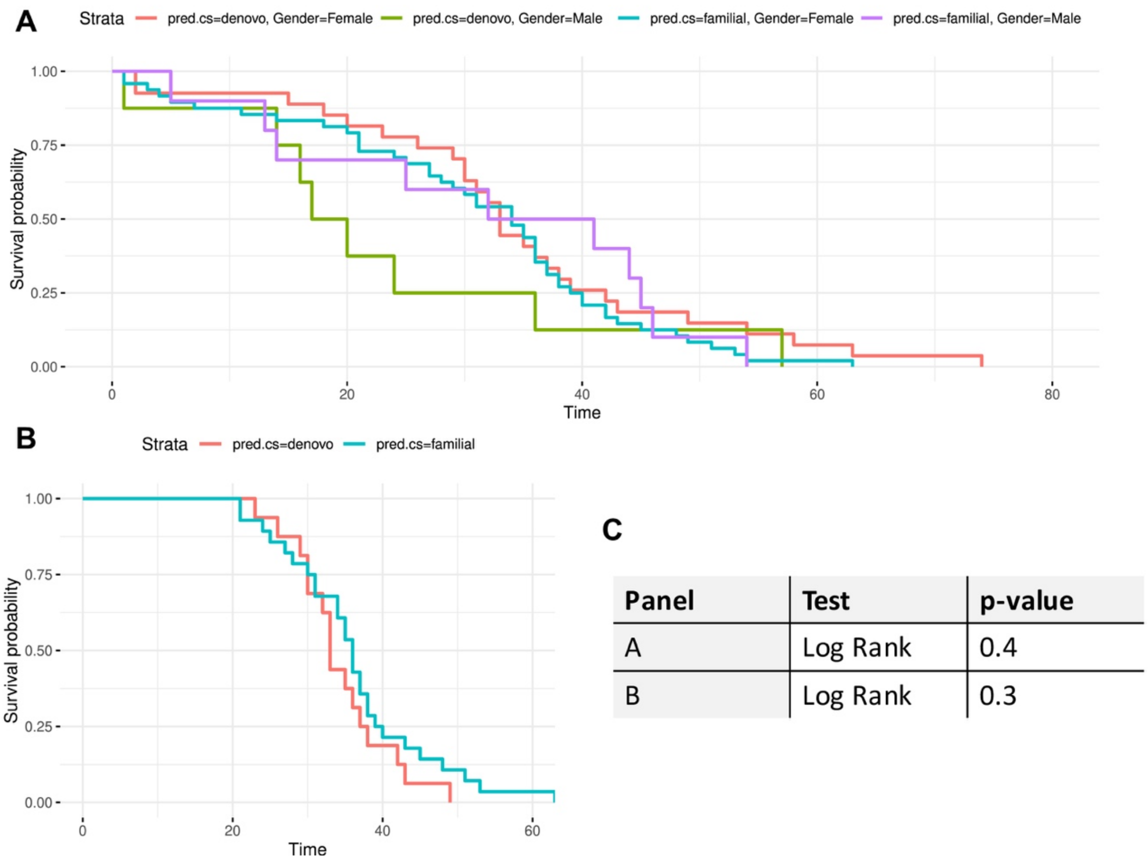
We then interrogated if there were any differences in the latent time fist cancer in DNMs and FMs in the CCG-MDA cohort. We chose to compare DNMs and FMs while controlling for two clinical covariates. First, we controlled for gender (male or female) since a difference in the lifetime risk has been cancer males and females has been reported (Schneider et al., 2019). **Figure 5A** shows a Kaplan-Meier curve of time to first cancer for all probands in CCG-MDA. Using the Log-Rank test, we compared the survival of four groups; de novo males, de novo females, familial males and familial females, and found that they were not significantly different (p-value = 0.4). Because the survival curves of de novo males and familial males do seem to have some separation, we also performed an additional comparison of only these two groups using the Log-Rank test, but as before, there was not a significant difference between them (p=0.6). However, in the context of LFS, this comparison has many caveats. First, females have a much higher risk of breast cancer than males, and early onset breast cancer is one of the few unbiased ascertainment criteria. This means that groups like males with a de novo mutation may be underrepresented in this cohort, or at least not a random sample of the LFS population. For ascertainment, de novo males need to present with early onset tumors that are strongly associated with LFS, as seen by an early dip in the de novo males' curve, mostly caused by early onset osteosarcoma, soft tissue sarcomas and brain cancers. The biggest dip for this group is between the ages of 15 and 25 years of age, consistent with a strong pediatric or young adult presentation of LFS. This is in contrast to males with a familial mutation that are more likely to be identified due to family history even if they lack pediatric cancers.

It is worth noting that this plot seems to have the opposite trend expected from male and female LFS patients, were females are expected to have a higher lifetime cancer risk than males (Schneider et al., 2019; Shin et al., 2020b, 2020a). This current analysis is focused only on probands ascertained clinically, who all present with a first cancer. This suggests that there are possibly many male DNMs that are simply not ascertained into cohorts due to a later onset of disease, and possible because they never present disease. This consistent with other observations, such as the lack of *TP53*-p.R248W mutations in DNMs, and strongly suggests that there is a need for more aggressive ascertainments of male DNMs.

The lack of a significant difference between the groups also calls into attention the relatively small number of events for some of the populations (males, in general). This once again highlights the importance of our statistical methods to confidently identify DNMs and FMs, and boost our sample size to appropriate sizes. Finally, different cancer types are likely to present at different ages, and therefore we controlled for the type of cancer diagnosed in our next analysis.

Figure 5*. Analysis of the latent time to first cancer of DNM and FM probands in CCG cohort.*

A) Comparison of the overall time to first cancer of any type. **B)** Comparison of time to first cancer in probands where the first cancer was breast cancer. **C)** Log-Rank test for differences in survival in the groups shown.



Panel	Test	p-value
A	Log Rank	0.4
B	Log Rank	0.3

As for previous analysis, we classified the cancers diagnosed by their classification in their LFS spectrum. We run a Cox Proportional Hazards model using DNM status, LFS spectrum cancer types and gender as the explanatory variables. The results of the model fit are listed in **Table 12**. The base variables in the model where “de novo” (DNM status), “adrenal corticoid carcinoma” (cancer type) and “female” (gender). The only significant variable in the model was the “cancer type” variable. We performed a test of proportionality on our fitted model and found positive evidence that the “cancer type” variable deviates from the assumption of proportional hazards (p-value = 0.00158). This highlights the statistical challenge of this problem and suggests that alternative approaches to control for the type of cancer diagnosed. We then proceeded to study the time to first cancer in patients whose first cancer was a breast cancer (**Figure 5B**). The survival curves of DNMs and FMs are not significantly different (Log-Rank test p-value = 0.3). Overall, we can conclude that the time to first cancer in DNMs and FMs ascertained into the CCG-MDA cohort is similar.

Table 12*. Summary of the model fit for time to first cancer.*

Variable	Coefficients	Hazard Ratios	Pr(>|z|)
DNM Status	---------	---------	---------
familial	0.170	1.185	0.473
Cancer type	---------	---------	---------
brain	-1.41	0.245	0.212
breast	-2.91	0.055	0.007
leukemia	-1.582	0.206	0.190
lung	-3.811	0.022	0.010
ost	-0.429	0.651	0.700
other	-3.273	0.038	0.003

sts	-3.100	0.045	0.007
Gender	---------	---------	---------
Male	-0.197	0.822	0.591

Chapter 4 – Discussion, model improvements and future directions

In this work we have developed a probabilistic method called Famdenovo.CS that can be used to systematically identify carriers of DNMs in *TP53*. We've demonstrated our method's excellent capacity to discriminate DNMs in different data sets with varying ascertainment criteria and data collection practices. We also showed that by incorporating penetrance estimates that are specific to the type of first cancer diagnosed, we were able to improve the prediction accuracy and the model calibration. We also demonstrated the utility of Famdenovo.CS by applying it to 101 LFS families of the CCG-MDA cohort. Of these 101 families, we predict a total of 39 DNMs and 62 FMs and using unbiased ascertainment criteria, we estimate a 1:1 ratio of DNMs compared to FMs, supporting the immense importance of these carriers to the total population of patients with LFS. In order to study and characterize DNMs in LFS, we compared several clinical characteristics between DNMs and FMs including: deleterious mutations in *TP53*, cancer types diagnosed, age of cancer diagnosis and time to first cancer.

We did not find differences between DNMs and FMs in the risk of developing each cancer type, nor did we find differences in the ages of diagnosis, or time to first cancer diagnosis. However, because of the various cancer types that are consistently observed in LFS, cancer-specific analysis is not trivial or straight forward, and more complicated statistical modeling than our current Cox Proportional Hazards models might be needed. Moreover, there are very few events for certain cancer types; for example, lung, adrenal corticoid and choroid plexus cancer. This observation indicates that we also need to continue collecting LFS patients and identifying DNMs, in order to increase the amount of high-quality data used for statistical inference, thereby coming full circle, and supporting the importance of the work done on this thesis, which is meant to fill this critical need.

We also studied the deleterious mutations in *TP53* observed in DNMs and FMs. Most mutations were unique (only observed in one family) with a few exceptions that were mostly mutations that have been previously determined to be hotspots, such as NC_000017.11:g.7675088C>T (*TP53*-p.R175H), NC_000017.11:g.7674220C>T (*TP53*-p.R248Q), NC_000017.11:g.7674221G>A (*TP53*-p.R248W) and NC_000017.11:g.7673776G>A (*TP53*-p.R282W) (Walerych et al., 2012). We limited our analysis to pathogenic and likely pathogenic mutations, and Famdenovo.CS assumes that all deleterious mutations in *TP53* are equally penetrant. However, differences in the penetrance of each specific mutation have been reported, for example in NC_000017.11:g.7670699C>T (*TP53*-p.R337H), commonly referred to as the “Brazilian variant” (Hahn et al., 2018; Pinto and Zambetti, 2020; Volc et al., 2020).

Differences in the penetrance of distinct *TP53* mutations should also be accounted for. However, besides the Brazilian variant, research on the annotation of different *TP53* mutations is currently ongoing (Leroy et al., 2017; Tikkanen et al., 2018), hence, we do not have a biologically driven way to group distinct *TP53* mutations. The lack of biologically motivated grouping of the mutations is a huge limitation, since a meaningful grouping is needed as each mutation is rarely repeated in distinct families, therefore not achieving the sample size required for mutation-specific penetrance estimation. One possibility might be to aid clustering of TP53 variants using variant annotation scores such as Sift or Polyphen (Adzhubei et al., 2013; Ng and Henikoff, 2003) however, all variants accepted as pathogenic in the clinical setting generally have deleterious Sift/Polyphen scores, and it is not clear whether further, reasonable segregation could be provided by clustering mutations according to these scores. As both annotation literature accumulates and the collection of LFS families continues, this analysis will be enabled. Of course, penetrance estimation that accounts for both cancer type and a mutation in *TP53* is very complicated, and will require novel sophisticated statistical models, in addition to a very large sample size of high-quality data.

Our analysis of the mutations observed in DNMs and FMs complemented previous reports in other LFS cohorts where the hotspot mutation causing in *TP53* the amino acid change p.R248W (NC_000017.11:g.7674221G>A) is only observed in FMs. With the addition of the families with this mutation identified in the CCG-MDA cohort, we now have stronger evidence to conclude a likely distinct mechanism for this mutation to establish itself in the germline of LFS families. Although we did not find DNMs with this variant, mosaic (somatic) mutation carriers of this variant have been reported in literature (Azzollini et al., 2020). The report of this variant as a mosaic mutation, also serves as evidence for mosaicism with inclusion of the germ cells, or mosaicism within the germ cells themselves, to be a mechanism of establishment of this variant in the germline of the next generation (Renaux-Petel et al., 2018).

Although mosaicism has been established as a confounder of positive *TP53* germline mutation testing (Weitzel et al., 2018), our cohorts are ascertained through clinical suspicion of LFS, increasing the likelihood of a mutation being germline. Moreover, we removed probands whose genetic testing reported a “likely mosaic” or “low allele frequency” variant (according to the genetic laboratory), as our model assumes all mutations in *TP53* are pathogenic germline variants. A reduced contribution of mosaicism to our cohort, might explain why some mutations in *TP53* are not observed as DNMs or FMs. Therefore, expanding our model to include mosaicism as an additional genotype, can be considered. However, it is important to consider the alternate hypothesis that the ascertainment criteria in our cohort simply did not identify carriers of some mutations.

Although for the majority of this work we focused on the research utility of the Famdenovo.CS model, we cannot disregard the clinical utility of our method. LFS is a familial syndrome where understanding whether a patient diagnosed with LFS is DNM or FM, might provide the physician or genetic counselor with additional information relevant in a counseling session. For example, the parents of a pediatric patient that is highly likely to be DNM, might prefer to disregard genetic testing for *TP53* if they have not shown a phenotype suspicious for LFS, especially if the testing is not covered by their medical insurance. This type of genetic counseling is greatly improved by a more accurate and meaningful probability calculation, which fortunately is an area where our model demonstrated improvement. As we continue studying and further understand DNMs in LFS, we expect that our method will gain even more clinical relevance.

Our current study design is limited to identification of DNMs and clinical characterization of said population. Genomic characterization of DNMs and FMs still remains. The gene *TP53* is involved in many molecular processes, and understanding how genomes change across generations will require DNMs and subsequent generations. Moreover, genomic characterization of the DNM carriers in other diseases such as Autism, has yielded increased understanding of disease mechanisms (Michaelson et al., 2012; Yuen et al., 2016). Genome sequencing analysis of the DNMs in LFS will also allow for identification of the most common risk factors and genomic mechanisms responsible for acquisition of a deleterious DNM in *TP53*. For example, a larger paternal contribution of genome-wide DNMs has been reported (Jónsson et al., 2018, 2017). Older maternal age has been associated with DNMs in tumor-suppressor genes, while older paternal age has been associated with DNMs in oncogenes (Acuna-Hidalgo et al., 2016). Our model will allow for the necessary identification of families with DNM carriers needed for the aforementioned research.

There are still areas where our model can be improved. As reiterated several times through this text, LFS is a complicated disease where patients are predisposed to a wide spectrum of cancers, and where each patient can have multiple events of cancer throughout their life. We've demonstrated improvement of the model through the introduction of cancer-specific penetrance estimates, however, these estimates were developed considering the time to the first event of cancer. Method to estimate penetrance that account for multiple events of cancer have been developed (Shin et al., 2020a), and it's possible that considering multiple events of cancer, on top of the type of cancer diagnosed in each event, would further improve our model. Although, at this current time, a model that jointly estimates penetrance while considering both the type of cancer diagnosed and multiple events of cancer remains undeveloped for LFS, if such penetrance estimates become available, they will likely increase our model's discrimination, calibration and accuracy, and should be incorporated. Another possible addition, is to include penetrance estimates that are specific to the mutation in *TP53*. Estimating this mutation-specific penetrance is currently not feasible due to the limited sample size of recurrent mutations in *TP53* in cohorts of LFS families. However, the ongoing annotation of *TP53* variants and accumulation of literature in the topic should enable this in the future.

Besides more sophisticated penetrance estimates, another area that can possibly improve our model's performance is to include a "mosaicism" as a separate genotype. Mosaicism should be low on our data, as it is clinically ascertained using LFS criteria, and we removed test results with low variant allele frequency (Weitzel et al., 2018). However, as a model that considers mosaic *TP53* mutation carriers as another genotype, could provide utility in clinical practice, especially in settings with a less strict ascertainment criteria. Mosaic carriers are less likely to pass the genotype to their offspring, and are unlikely to have anterior family members affected by a mutation in *TP53*. Moreover, a mosaic carrier can confound both FM and DNM classifications. Lastly, in this version of our model we assume there are no issues of non-paternity/maternity. For future iterations of our model, it might be useful to incorporate this into the probability calculation.

Finally, another possible direction is to extend this model to other diseases. The model is easily generalizable to other cancer syndromes with autosomal dominant inheritance patterns that depend on a single or small number of genes such as breast and ovarian cancer syndrome, adenomatous familial polyposis, Lynch syndrome, familial pancreatic cancer or familial malignant melanoma. Another possible direction for this sort of modeling approach would be to use in neurological disease, such as Neurofibromatosis, Alzheimer’s disease and perhaps autism spectrum disorder (ASD). However, neurological diseases, such as ASD, are usually a result of many complex genomic events. This would make the space of all possible genotypes for an individual gigantic, and calculating the likelihood for all possible combination of genotypes for a whole pedigree a gargantuan computational problem, and currently unsolvable. However, as knowledge of neurologic disorders expands, gene and variant selection will very likely increase the opportunity for application of mendelian models, even in such complex diseases.

References

Acuna-Hidalgo, R., Veltman, J.A., Hoischen, A., 2016. New insights into the generation and role of de novo mutations in health and disease. Genome Biol. 17. https://doi.org/10.1186/s13059-016-1110-1

Adzhubei, I., Jordan, D.M., Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. Chapter 7, Unit7.20-Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76

Anupindi, S.A., Bedoya, M.A., Lindell, R.B., Rambhatla, S.J., Zelley, K., Nichols, K.E., Chauvin, N.A., 2015. Diagnostic performance of whole-body MRI as a tool for cancer screening in children with genetic cancer-predisposing conditions. Am. J. Roentgenol. 205, 400–408. https://doi.org/10.2214/AJR.14.13663

Aubrey, B.J., Strasser, A., Kelly, G.L., 2016. Tumor-suppressor functions of the TP53 pathway. Cold Spring Harb. Perspect. Med. 6. https://doi.org/10.1101/cshperspect.a026062

Azzollini, J., Schiavello, E., Buttarelli, F.R., Clerici, C.A., Tizzoni, L., De Vecchi, G., Capra, F., Pisati, F., Biassoni, V., Runza, L., Carrabba, G., Giangaspero, F., Massimino, M., Pensotti, V., Manoukian, S., 2020. Pre- and post-zygotic TP53 de novo mutations in SHH-Medulloblastoma. Cancers (Basel). 12, 1–16. https://doi.org/10.3390/cancers12092503

Ballinger, M.L., Best, A., Mai, P.L., Khincha, P.P., Loud, J.T., Peters, J.A., Achatz, M.I., Chojniak, R., Balieiro da Costa, A., Santiago, K.M., Garber, J., O’Neill, A.F., Eeles, R.A., Evans, D.G., Bleiker, E., Sonke, G.S., Ruijs, M., Loo, C., Schiffman, J., Naumer, A., Kohlmann, W., Strong, L.C., Bojadzieva, J., Malkin, D., Rednam, S.P., Stoffel, E.M., Koeppe, E., Weitzel, J.N., Slavin, T.P., Nehoray, B., Robson, M., Walsh, M., Manelli, L., Villani, A., Thomas, D.M., Savage, S.A., 2017. Baseline Surveillance in Li-Fraumeni Syndrome Using Whole-Body Magnetic Resonance Imaging: A Meta-analysis. JAMA Oncol. 3, 1634–1639. https://doi.org/10.1001/jamaoncol.2017.1968

Barbosa, K., Li, S., Adams, P.D., Deshpande, A.J., 2019. The role of TP53 in acute myeloid leukemia: Challenges and opportunities. Genes Chromosom. Cancer 58, 875–888. https://doi.org/10.1002/gcc.22796

Baugh, E.H., Ke, H., Levine, A.J., Bonneau, R.A., Chan, C.S., 2018. Why are there hotspot mutations in the TP53 gene in human cancers? Cell Death Differ. 25, 154–160. https://doi.org/10.1038/cdd.2017.180

Birch, J.M., Hartley, A.L., Tricker, K.J., Prosser, J., Condie, A., Kelsey, A.M., Harris, M., Jones, P.H., Binchy, A., Crowther, D., et al., 1994. Prevalence and diversity of constitutional mutations in the p53 gene among 21 Li-Fraumeni families. Cancer Res 54, 1298–1304.

Bougeard, G., Renaux-Petel, M., Flaman, J.M., Charbonnier, C., Fermey, P., Belotti, M., Gauthier-Villars, M., Stoppa-Lyonnet, D., Consolino, E., Brugières, L., Caron, O., Benusiglio, P.R., Bressac-de Paillerets, B., Bonadona, V., Bonaïti-Pellié, C., Tinat, J., Baert-Desurmont, S., Frebourg, T., 2015. Revisiting Li-Fraumeni syndrome from TP53 mutation carriers. J. Clin. Oncol. 33, 2345–2352. https://doi.org/10.1200/JCO.2014.59.5728

Brandler, W.M., Sebat, J., 2015. From de novo mutations to personalized therapeutic interventions in autism. Annu. Rev. Med. 66, 487–507. https://doi.org/10.1146/annurev-med-091113-024550

Chen, S., Wang, W., Broman, K.W., Katki, H.A., Parmigiani, G., 2004. BayesMendel: An R environment for Mendelian risk prediction. Stat. Appl. Genet. Mol. Biol. 3.

https://doi.org/10.2202/1544-6115.1063

Chen, S., Wang, W., Shing Lee, M., Khedoudja Nafa, S., Lee, J., Kathy Romans, M., Patrice Watson, M., Gruber, S.B., David Euhus, M., Kinzler, K.W., Jass, J., Steven Gallinger, Ds., Lindor, N.M., Casey, G., Ellis, N., Giardiello, F.M., Offit, K., Giovanni Parmigiani, M., 2006. Prediction of Germline Mutations and Cancer Risk in the Lynch Syndrome. JAMA 296, 1479–1487.

Chompret, A., Abel, A., Stoppa-Lyonnet, D., Brugieres, L., Pages, S., Feunteun, J., Bonaiti-Pellie, C., 2001. Sensitivity and predictive value ofcriteria for p53 germline mutationscreening. J. Med. Genet. 38, 43–47.

Cioppi, F., Casamonti, E., Krausz, C., 2019. Age-Dependent De Novo Mutations During Spermatogenesis and Their Consequences, in: Baldi, E., Muratori, M. (Eds.), Genetic Damage in Human Spermatozoa. Springer International Publishing, Cham, pp. 29–46. https://doi.org/10.1007/978-3-030-21664-1_2

Elston, R.C., Stewart, J., 1971. A General Model for the Genetic Analysis of Pedigree Data. Hum. Hered. 21, 523–542. https://doi.org/10.1159/000152448

Euhus, D.M., Smith, K.C., Robinson, L., Stucky, A., Olopade, O.I., Cummings, S., Garber, J.E., Chittenden, A., Mills, G.B., Rieger, P., Esserman, L., Crawford, B., Hughes, K.S., Roche, C.A., Ganz, P.A., Seldon, J., Fabian, C.J., Klemp, J., Tomlinson, G., 2002. Pretest Prediction of BRCA1 or BRCA2 Mutation by Risk Counselors and the Computer Model BRCAPRO. J. Natl. Cancer Inst. 94, 844–851.

Fernando, R.L., Stricker, C., Elston, R.C., 1993. An efficient algorithm to compute the posterior genotypic distribution for every member of a pedigree without loops. Theor Appl Genet 87, 89–93.

Francioli, L.C., Polak, P.P., Koren, A., Menelaou, A., Chun, S., Renkens, I., Van Duijn, C.M., Swertz, M., Wijmenga, C., Van Ommen, G., Slagboom, P.E., Boomsma, D.I., Ye, K., Guryev, V., Arndt, P.F., Kloosterman, W.P., De Bakker, P.I.W., Sunyaev, S.R., 2015. Genome-wide patterns and properties of de novo mutations in humans. Nat. Genet. 47, 822–826. https://doi.org/10.1038/ng.3292

Gao, F., Pan, X., Dodd-Eaton, E.B., Recio, C.V., Montierth, M.D., Bojadzieva, J., Mai, P.L., Zelley, K., Johnson, V.E., Braun, D., Nichols, K.E., Garber, J.E., Savage, S.A., Strong, L.C., Wang, W., 2020. A pedigree-based prediction model identifies carriers of deleterious de novo mutations in families with Li-Fraumeni syndrome. Genome Res. 30. https://doi.org/10.1101/gr.249599.119

Goldmann, J.M., Veltman, J.A., Gilissen, C., 2019. De Novo Mutations Reflect Development and Aging of the Human Germline. Trends Genet. 35, 828–839. https://doi.org/10.1016/j.tig.2019.08.005

Gonzalez, K.D., Buzin, C.H., Noltner, K.A., Gu, D., Li, W., Malkin, D., Sommer, S.S., 2009. High frequency of de novo mutations in Li-Fraumeni syndrome. J. Med. Genet. 46, 689–693. https://doi.org/10.1136/jmg.2008.058958

Hahn, E.C., Bittar, C.M., Vianna, F.S.L., Netto, C.B.O., Biazús, J.V., Cericatto, R., Cavalheiro, J.A., De Melo, M.P., Menke, C.H., Rabin, E., Leistner-Segal, S., Ashton-Prolla, P., 2018. TP53 p.Arg337His germline mutation prevalence in Southern Brazil: Further evidence for mutation testing in young breast cancer patients. PLoS One 13. https://doi.org/10.1371/journal.pone.0209934

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: The next generation. Cell. https://doi.org/10.1016/j.cell.2011.02.013

Hanel, W., Marchenko, N., Xu, S., Xiaofeng Yu, S., Weng, W., Moll, U., 2013. Two hot spot mutant p53 mouse models display differential gain of function in tumorigenesis. Cell Death Differ. 20, 898–909. https://doi.org/10.1038/cdd.2013.17

Holmfeldt, L., Wei, L., Diaz-Flores, E., Walsh, M., Zhang, J., Ding, L., Payne-Turner, D., Churchman, M., Andersson, A., Chen, S.C., Mccastlain, K., Becksfort, J., Ma, J., Wu, G., Patel, S.N., Heatley, S.L., Phillips, L.A., Song, G., Easton, J., Parker, M., Chen, X., Rusch, M., Boggs, K., Vadodaria, B., Hedlund, E., Drenberg, C., Baker, S., Pei, D., Cheng, C., Huether, R., Lu, C., Fulton, R.S., Fulton, L.L., Tabib, Y., Dooling, D.J., Ochoa, K., Minden, M., Lewis, I.D., To, L.B., Marlton, P., Roberts, A.W., Raca, G., Stock, W., Neale, G., Drexler, H.G., Dickins, R.A., Ellison, D.W., Shurtleff, S.A., Pui, C.H., Ribeiro, R.C., Devidas, M., Carroll, A.J., Heerema, N.A., Wood, B., Borowitz, M.J., Gastier-Foster, J.M., Raimondi, S.C., Mardis, E.R., Wilson, R.K., Downing, J.R., Hunger, S.P., Loh, M.L., Mullighan, C.G., 2013. The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nat. Genet. 45, 242–252. https://doi.org/10.1038/ng.2532

Hwang, S.J., Lozano, G., Amos, C.I., Strong, L.C., 2003. Germline p53 mutations in a cohort with childhood sarcoma: sex differences in cancer risk. Am J Hum Genet 72, 975–983. https://doi.org/10.1086/374567

Jin, Z.B., Li, Z., Liu, Z., Jiang, Y., Cai, X.B., Wu, J., 2018. Identification of de novo germline mutations and causal genes for sporadic diseases using trio-based whole-exome/genome sequencing. Biol. Rev. 93, 1014–1031. https://doi.org/10.1111/brv.12383

Jónsson, H., Sulem, P., Arnadottir, G.A., Pálsson, G., Eggertsson, H.P., Kristmundsdottir, S., Zink, F., Kehr, B., Hjorleifsson, K.E., Jensson, B., Jonsdottir, I., Marelsson, S.E., Gudjonsson, S.A., Gylfason, A., Jonasdottir, Adalbjorg, Jonasdottir, Aslaug, Stacey, S.N., Magnusson, O.T., Thorsteinsdottir, U., Masson, G., Kong, A., Halldorsson, B. V., Helgason, A., Gudbjartsson, D.F., Stefansson, K., 2018. Multiple transmissions of de novo mutations in families. Nat. Genet. 50, 1674–1680. https://doi.org/10.1038/s41588-018-0259-9

Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., Hardarson, M.T., Hjorleifsson, K.E., Eggertsson, H.P., Gudjonsson, S.A., Ward, L.D., Arnadottir, G.A., Helgason, E.A., Helgason, H., Gylfason, A., Jonasdottir, Adalbjorg, Jonasdottir, Aslaug, Rafnar, T., Frigge, M., Stacey, S.N., Th. Magnusson, O., Thorsteinsdottir, U., Masson, G., Kong, A., Halldorsson, B. V., Helgason, A., Gudbjartsson, D.F., Stefansson, K., 2017. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. Nature 549, 519–522. https://doi.org/10.1038/nature24018

Katki, H.A., Blackford, A., Chen, S., Parmigiani, G., 2008. Multiple diseases in carrier probability estimation: accounting for surviving all cancers other than breast and ovary in BRCAPRO. Stat Med 27, 4532–4548. https://doi.org/10.1002/sim.3302

Kim, M.P., Lozano, G., 2018. Mutant p53 partners in crime. Cell Death Differ. 25, 161–168. https://doi.org/10.1038/cdd.2017.185

Kratz, C.P., Achatz, M.I., Brugieres, L., Frebourg, T., Garber, J.E., Greer, M.L.C., Hansford, J.R., Janeway, K.A., Kohlmann, W.K., McGee, R., Mullighan, C.G., Onel, K., Pajtler, K.W., Pfister, S.M., Savage, S.A., Schiffman, J.D., Schneider, K.A., Strong, L.C., Evans, D.G.R., Wasserman, J.D., Villani, A., Malkin, D., 2017. Cancer screening recommendations for individuals with Li-Fraumeni syndrome. Clin. Cancer Res. 23, e38–e45. https://doi.org/10.1158/1078-0432.CCR-17-0408

Leroy, B., Ballinger, M.L., Baran-Marszak, F., Bond, G.L., Braithwaite, A., Concin, N., Donehower, L.A., El-Deiry, W.S., Fenaux, P., Gaidano, G., Langerød, A., Hellstrom-Lindberg, E., Iggo, R., Lehmann-Che, J., Mai, P.L., Malkin, D., Moll, U.M., Myers, J.N., Nichols, K.E., Pospisilova, S., Ashton-Prolla, P., Rossi, D., Savage, S.A., Strong, L.C., Tonin, P.N., Zeillinger, R., Zenz, T., Fraumeni, J.F., Taschner, P.E.M., Hainaut, P., Soussi, T., 2017. Recommended guidelines for validation, quality control, and reporting of TP53 variants in clinical practice. Cancer Res. 77, 1250–1260. https://doi.org/10.1158/0008-5472.CAN-16-2179

Levine, A.J., 2020. p53: 800 million years of evolution and 40 years of discovery. Nat. Rev. Cancer 20, 471–480. https://doi.org/10.1038/s41568-020-0262-1

Li, F., Fraumeni Jr, J., 1969. Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome. J. Natl. Cancer Inst. 43, 1365–1373.

Li, F.P., Fraumeni, J.F., Mulvihill, J.J., Blattner, W.A., Dreyfus, M.G., Tucker, M.A., Miller, R.W., 1988. A cancer family syndrome in twenty-four kindreds. CANCER Res. 48, 5358–5362.

Macfarland, S.P., Zelley, K., Long, J.M., Mckenna, D., Mamula, P., Domchek, S.M., Nathanson, K.L., Brodeur, G.M., Rustgi, A.K., Katona, B.W., Maxwell, K.N., 2019. Earlier Colorectal Cancer Screening May Be Necessary In Patients With Li-Fraumeni Syndrome. Gastroenterology 156, 273–274. https://doi.org/10.1053/j.gastro.2018.09.036

Mai, P.L., Best, A.F., Peters, J.A., DeCastro, R.M., Khincha, P.P., Loud, J.T., Bremer, R.C., Rosenberg, P.S., Savage, S.A., 2016. Risks of first and subsequent cancers among TP53 mutation carriers in the National Cancer Institute Li-Fraumeni syndrome cohort. Cancer 122, 3673–3681. https://doi.org/10.1002/cncr.30248

Mai, P.L., Malkin, D., Garber, J.E., Schiffman, J.D., Weitzel, J.N., Strong, L.C., Wyss, O., Locke, L., Means, V., Achatz, M.I., Hainaut, P., Frebourg, T., Evans, D.G., Bleiker, E., Patenaude, A., Schneider, K., Wilfond, B., Peters, J.A., Hwang, P.M., Ford, J., Tabori, U., Ognjanovic, S., Dennis, P.A., Wentzensen, I.M., Greene, M.H., Fraumeni, J.F., Savage, S.A., 2012. Li-Fraumeni syndrome: Report of a clinical research workshop and creation of a research consortium. Cancer Genet. 205, 479–487. https://doi.org/10.1016/j.cancergen.2012.06.008

Malkin, D., 2011. Li-fraumeni syndrome. Genes and Cancer 2, 475–484. https://doi.org/10.1177/1947601911413466

Malkin, David, Li, Frederick P, Strong, L.C., Fraumeni, J.F., Nelson, C.E., Kim, David H, Kassel, Jayne, Gryka, Magdalena A, Bischoff, Farideh Z, Tainsky, Michael A, Friend, Stephen H, Malkin, D, Nclson, C.E., Kim, D H, Kassel, J, Gryka, M A, Friend, S H, Li, F P, Bischoff, F Z, Tainsky, M A, 1990. Germ Line p53 Mutations in a Familial Syndrome of Breast Cancer, Sarcomas, and Other Neoplasms. Science (80-.). 250, 1233–1238.

McBride, K.A., Ballinger, M.L., Schlub, T.E., Young, M.A., Tattersall, M.H.N., Kirk, J., Eeles, R., Killick, E., Walker, L.G., Shanley, S., Thomas, D.M., Mitchell, G., 2017. Psychosocial morbidity in TP53 mutation carriers: is whole-body cancer screening beneficial? Fam. Cancer 16, 423–432. https://doi.org/10.1007/s10689-016-9964-7

Michaelson, J.J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., Wu, W., Corominas, R., Peoples, Á., Koren, A., Gore, A., Kang, S., Lin, G.N., Estabillo, J., Gadomski, T., Singh, B., Zhang, K., Akshoomoff, N., Corsello, C., McCarroll, S., Iakoucheva, L.M., Li, Y., Wang, J., Sebat, J., 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell 151, 1431–1442. https://doi.org/10.1016/j.cell.2012.11.019

Ng, P.C., Henikoff, S., 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 31, 3812–3814. https://doi.org/10.1093/nar/gkg509

Nicolas, G., Veltman, J.A., 2019. The role of de novo mutations in adult-onset neurodegenerative disorders. Acta Neuropathol. 137, 183–207. https://doi.org/10.1007/s00401-018-1939-3

Ohno, M., 2019. Spontaneous de novo germline mutations in humans and mice: rates, spectra, causes and consequences. Genes Genet. Syst. 94, 13–22. https://doi.org/10.1266/ggs.18-00015

Olivier, M., Hollstein, M., Hainaut, P., 2010. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb. Perspect. Biol. 2. https://doi.org/10.1101/cshperspect.a001008

Oren, M., Rotter, V., 2010. Mutant p53 gain-of-function in cancer. Cold Spring Harb. Perspect. Biol. 2. https://doi.org/10.1101/cshperspect.a001107

Peng, G., Bojadzieva, J., Ballinger, M.L., Li, J., Blackford, A.L., Mai, P.L., Savage, S.A., Thomas, D.M., Strong, L.C., Wang, W., 2017. Estimating TP53 mutation carrier probability in families with li-fraumeni syndrome using LFSPRO. Cancer Epidemiol. Biomarkers Prev. 26, 837–844. https://doi.org/10.1158/1055-9965.EPI-16-0695

Pinto, E.M., Zambetti, G.P., 2020. What 20 years of research has taught us about the TP53 p.R337H mutation. Cancer. https://doi.org/10.1002/cncr.33143

Rath, M.G., Masciari, S., Gelman, R., Miron, A., Miron, P., Foley, K., Richardson, A.L., Krop, I.E., Verselis, S.J., Dillon, D.A., Garber, J.E., 2013. Prevalence of germline TP53 mutations in HER2+ breast cancer patients. Breast Cancer Res. Treat. 139, 193–198. https://doi.org/10.1007/s10549-012-2375-z

Renaux-Petel, M., Charbonnier, F., Théry, J.C., Fermey, P., Lienard, G., Bou, J., Coutant, S., Vezain, M., Kasper, E., Fourneaux, S., Manase, S., Blanluet, M., Leheup, B., Mansuy, L., Champigneulle, J., Chappé, C., Longy, M., Sévenet, N., Paillerets, B.B. De, Guerrini-Rousseau, L., Brugières, L., Caron, O., Sabourin, J.C., Tournier, I., Baert-Desurmont, S., Frébourg, T., Bougeard, G., 2018. Contribution of de novo and mosaic TP53 mutations to Li-Fraumeni syndrome. J. Med. Genet. 55, 173–180. https://doi.org/10.1136/jmedgenet-2017-104976

Ronemus, M., Iossifov, I., Levy, D., Wigler, M., 2014. The role of de novo mutations in the genetics of autism spectrum disorders. Nat. Rev. Genet. 15, 133–141. https://doi.org/10.1038/nrg3585

Schneider, K., Zelley, K., Nichols, K.E., Garber, J., 2019. Li-Fraumeni Syndrome., in: Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J.H., Mirzaa, G., Amemiya, A. (Eds.), . Seattle (WA).

Schon, K., Tischkowitz, M., 2018. Clinical implications of germline mutations in breast cancer: TP53. Breast Cancer Res. Treat. https://doi.org/10.1007/s10549-017-4531-y

Shin, S.J., Dodd-Eaton, E.B., Gao, F., Bojadzieva, J., Chen, J., Kong, X., Amos, C.I., Ning, J., Strong, L.C., Wang, W., 2020a. Penetrance estimates over time to first and second primary cancer diagnosis in families with Li-Fraumeni syndrome: A single institution perspective. Cancer Res. 80, 347–353. https://doi.org/10.1158/0008-5472.CAN-19-0725

Shin, S.J., Dodd-Eaton, E.B., Peng, G., Bojadzieva, J., Chen, J., Amos, C.I., Frone, M.N., Khincha, P.P., Mai, P.L., Savage, S.A., Ballinger, M.L., Thomas, D.M., Yuan, Y., Strong, L.C., Wang, W., 2020b. Penetrance of different cancer types in families with Li-Fraumeni syndrome: A validation study using multicenter cohorts. Cancer Res. 80, 354–360. https://doi.org/10.1158/0008-5472.CAN-19-0728

Stein, Y., Rotter, V., Aloni-Grinstein, R., 2019. Gain-of-function mutant p53: All the roads lead to tumorigenesis. Int. J. Mol. Sci. 20. https://doi.org/10.3390/ijms20246197

Tikkanen, T., Leroy, B., Fournier, J.L., Risques, R.A., Malcikova, J., Soussi, T., 2018. Seshat: A Web service for accurate annotation, validation, and analysis of TP53 variants generated by conventional and next-generation sequencing. Hum. Mutat. 39, 925–933. https://doi.org/10.1002/humu.23543

Tinat, J., Bougeard, G., Baert-Desurmont, S., Vasseur, S., Martin, C., Bouvignies, E., Caron, O., BrigitteBressac-De Paillerets, B., Berthet, P., Dugast, C., Bonaïti-Pellié, C., Stoppa-Lyonnet, D., Frébourg, T., 2009. 2009 Version of the Chompret Criteria for Li Fraumeni Syndrome. J. Clin. Oncol. 27. https://doi.org/10.1200/JCO.2009.22.7967

Trkova, M., Hladikova, M., Kasal, P., Goetz, P., Sedlacek, Z., 2002. Is there anticipation in the age at onset of cancer in families with Li-Fraumeni syndrome? J Hum Genet 47, 381–386.

Turner, T.N., Eichler, E.E., 2019. The Role of De Novo Noncoding Regulatory Mutations in Neurodevelopmental Disorders. Trends Neurosci. 42, 115–127. https://doi.org/10.1016/j.tins.2018.11.002

Valdez, J.M., Nichols, K.E., Kesserwan, C., 2017. Li-Fraumeni syndrome: a paradigm for the understanding of hereditary cancer predisposition. Br. J. Haematol. https://doi.org/10.1111/bjh.14461

Vieler, M., Sanyal, S., 2018. P53 isoforms and their implications in cancer. Cancers (Basel). 10. https://doi.org/10.3390/cancers10090288

Villani, A., Shore, A., Wasserman, J.D., Stephens, D., Kim, R.H., Druker, H., Gallinger, B., Naumer, A., Kohlmann, W., Novokmet, A., Tabori, U., Tijerin, M., Greer, M.L.C., Finlay, J.L., Schiffman, J.D., Malkin, D., 2016. Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: 11 year follow-up of a prospective observational study. Lancet Oncol. 17, 1295–1305. https://doi.org/10.1016/S1470-2045(16)30249-2

Volc, S.M., Ramos, C.R.N., Galvao, H.D.C.R., Felicio, P.S., Coelho, A.S., Berardineli, G.N., Campacci, N., Sabato, C.D.S., Abrahao-Machado, L.F., Santana, I.V.V., Campanella, N., Lengert, A.V.H., Vidal, D.O., Reis, R.M., Dantas, C.F., Coelho, R.C., Boldrini, E., Serrano, S.V., Palmero, E.I., 2020. The Brazilian TP53 mutation (R337H) and sarcomas. PLoS One 15. https://doi.org/10.1371/journal.pone.0227260

Walerych, D., Napoli, M., Collavin, L., Del Sal, G., 2012. The rebel angel: Mutant p53 as the driving oncogene in breast cancer. Carcinogenesis 33, 2007–2017. https://doi.org/10.1093/carcin/bgs232

Wang, W., Chen, S., Brune, K.A., Hruban, R.H., Parmigiani, G., Klein, A.P., 2007. PancPRO: Risk assessment for individuals with a family history of pancreatic cancer. J. Clin. Oncol. 25, 1417–1422. https://doi.org/10.1200/JCO.2006.09.2452

Wang, W., Niendorf, K.B., Patel, D., Blackford, A., Marroni, F., Sober, A.J., Parmigiani, G., Tsao, H., 2010. Estimating CDKN2A carrier probability and personalizing cancer risk assessments in hereditary melanoma using melaPRO. Cancer Res. 70, 552–559. https://doi.org/10.1158/0008-5472.CAN-09-2653

Weitzel, J.N., Chao, E.C., Nehoray, B., Van Tongeren, L.R., LaDuca, H., Blazer, K.R., Slavin, T., Facmg, D.A.B.M.D., Pesaran, T., Rybak, C., Solomon, I., Niell-Swiller, M., Dolinsky, J.S., Castillo, D., Elliott, A., Gau, C.L., Speare, V., Jasperson, K., 2018. Somatic TP53 variants frequently confound germ-line testing results. Genet. Med. 20, 809–816. https://doi.org/10.1038/gim.2017.196

Wu, C.C., Shete, S., Amos, C.I., Strong, L.C., 2006. Joint effects of germ-line p53 mutation and sex on cancer risk in Li-Fraumeni syndrome. Cancer Res 66, 8287–8292. https://doi.org/10.1158/0008-5472.can-05-4247

Yates, A.D., Achuthan, P., Akanni, W., Allen, James, Allen, Jamie, Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J.C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., Giron, C.G., Gil, L., Grego, T., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., Kay, M., Lavidas, I., Le, T., Lemos, D., Martinez, J.G., Maurel, T., McDowall, M., McMahon, A., Mohanan, S., Moore, B., Nuhn, M., Oheh, D.N., Parker, A., Parton, A., Patricio, M., Sakthivel, M.P., Abdul Salam, A.I., Schmitt, B.M., Schuilenburg, H., Sheppard, D., Sycheva, M., Szuba, M., Taylor, K., Thormann, A., Threadgold, G., Vullo, A., Walts, B., Winterbottom, A., Zadissa, A., Chakiachvili, M., Flint, B., Frankish, A., Hunt, S.E., Iisley, G., Kostadima, M., Langridge, N., Loveland, J.E., Martin, F.J., Morales, J., Mudge, J.M., Muffato, M., Perry, E., Ruffier, M., Trevanion, S.J., Cunningham, F., Howe, K.L., Zerbino, D.R., Flicek, P., 2020. Ensembl 2020. Nucleic Acids Res. 48, D682–D688. https://doi.org/10.1093/nar/gkz966

Yuen, R.K.C., Merico, D., Cao, H., Pellecchia, G., Alipanahi, B., Thiruvahindrapuram, B., Tong, X., Sun, Y., Cao, D., Zhang, T., Wu, X., Jin, X., Zhou, Z., Liu, X., Nalpathamkalam, T., Walker, S., Howe, J.L., Wang, Z., Macdonald, J.R., Chan, A.J.S., D’Abate, L., Deneault, E., Siu, M.T., Tammimies, K., Uddin, M., Zarrei, M., Wang, M., Li, Y., Wang, Jun, Wang, Jian, Yang, H., Bookman, M., Bingham, J., Gross, S.S., Loy, D., Pletcher, M., Marshall, C.R., Anagnostou, E., Zwaigenbaum, L., Weksberg, R., Fernandez, B.A., Roberts, W., Szatmari, P., Glazer, D., Frey, B.J., Ring, R.H., Xu, X., Scherer, S.W., 2016. Genome-wide characteristics of de novo mutations in autism. npj Genomic Med. 1. https://doi.org/10.1038/npjgenmed.2016.27

VITA

Carlos Christian Vera Recio was born in San German, Puerto Rico, the son of Awilda Recio Camacho and Carlos Felipe era Muñoz. He was raised in Mayagüez, Puerto Rico, were he attended elementary, middle and high school. He would eventually go on to received his Bachelor of Science with a major in Biology and a curriculum sequence in Applied Math from the University of Puerto Rico – Mayagüez Campus, in May, 2014. In August of 2104, he enrolled in a Physician Scientist program through a U54 Partnership in Excellence grant between the University of Puerto Rico School of Medicine and The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences.