# A-TO-I RNA EDITING EVENTS, POTENTIAL BIOMARKERS FOR PROGNOSIS AND CHEMOSENSITIVITY IN GASTRIC CANCER

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

By
Isli Çela
September, 2022

A-TO-I RNA EDITING EVENTS, POTENTIAL BIOMARKERS FOR PROGNOSIS AND CHEMOSENSITIVITY IN GASTRIC CANCER

By Isli Çela

September 2022

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Ali Osmay Güre (Advisor)

Özlen Konu Karakayalı

Onur Emre Onat

Approved for the Graduate School of Engineering and Science:

ORHAN ARIKAN

Director of the Graduate School

# ABSTRACT

## A-TO-I RNA EDITING EVENTS, POTENTIAL BIOMARKERS FOR PROGNOSIS AND CHEMOSENSITIVITY IN GASTRIC CANCER

ISLI ÇELA
M.Sc. in Molecular Biology and Genetics

Advisor: Ali Osmay GÜRE

September 2022

Gastric cancer (GC) is one of the leading causes of cancer mortality, and it frequently presents in advanced stages with a poor prognosis and response to treatment. Although extensive research has identified many potential biomarkers in GC, the heterogeneity of the disease is an impediment to validation, so only a small number find limited application in clinics. RNA editing is an epigenetic modification that results in nucleotide changes in the RNA sequence. Adenosine to Inosine (A-to-I) substitutions are the most common editing events in humans, and they are mediated by Adenosine deaminases acting on RNA (ADAR) enzymes. Inosine (I) mimics Guanosine (G) and creates pairs with Cytidine (C), resulting in changes in RNA structure and stability, amino acid substitutions, alternative splicing, or gene expression regulation via miRNA target site modifications.RNA editing dysregulations have been found in breast, lung, kidney, brain, and gastric cancers, but the utility of specific editing events as biomarkers is largely unexplored. In this study we investigate the potential of A-to-I editing events as chemosensitivity and prognostic biomarkers in GC. Across multiple datasets, our analysis shows that RNA editing events at 305 unique positions correlate with drug sensitivity measures of 17 approved chemotherapeutics in GC cell lines.The most significant editing event-drug sensitivity correlations indicate that higher editing levels are associated with higher chemosensitivity. Interestingly, the expression levels of genes with identified editing events have a weaker or no correlation with drug sensitivity, implying that editing events are biomarkers independent of transcript levels. We show that, while ADAR enzymes mediate editing events, ADAR expression levels are not interchangeable with editing frequencies as chemosensitivity biomarkers in GC. We discovered a non-synonymous editing event in the C11orf80 coding sequence, resulting in an amino acid substitution (S.p133G). Also, we identified an editing event in the 3'UTR of SOGA1 that correlates with increased SOGA1 expression. The presence of this editing site in a putative target site of miR-9-5p suggests that gene expression might be regulated by miRNA target site modifications.

In the TCGA and Singapore cohorts, the prognostic role of editing events in GC was investigated. Overall, higher levels of editing are associated with better survival in GC patients. In both cohorts, we found an editing event in the CLPX gene at the position

65442098 to be an independent good prognostic factor. We chose editing events that would best categorize our patients into "High" and "Low" edited groups using the Log Rank Multiple Cut-off (LRMC) plot distribution. In each dataset, we propose two editing events, one good and one bad prognostic factor that independently correlate with survival in GC patients. In the Singapore Cohort, high editing levels in ZNF587 are associated with a good prognosis, while those in DCAF16 are associated with a poor prognosis.High editing levels in CTSB correlate with better overall survival (OS) in the TCGA cohort, while those in NUP43 correlate with worse OS. Because transcript levels do not correlate with survival, the prognostic effects of these editing events are unaffected by gene expression levels.We believe that editing levels at specific positions can be used as prognostic biomarkers in a significant way, providing a more cost-effective and applicable alternative to prognostic editing signature models. Our findings suggest that editing events could be used as independent biomarkers for chemosensitivity and prognosis in gastric cancer, however more investigation is required to elucidate the mechanisms underlying the observed relationships.

*Keywords:* A-to-I, ADAR, chemosensitivity, Gastric cancer, prognosis, RNA editing

# ÖZET

## A-I RNA DÜZENLEME OLAYLARI, GASTRİK KANSERİNDE PROGNOZ VE KEMOSENSİVİTE İÇİN POTANSİYEL BİYOMARKERLER

İSLİ ÇELA

Yüksek Lisans, Moleküler Biyoloji ve Genetik

Tez Danışmanı: Ali Osmay GÜRE

Eylül 2022

Mide kanseri (MK), kanser ölümlerinin önde gelen nedenlerinden biridir ve sıklıkla ileri evrelerde kötü prognoz ve tedaviye yanıt ile ortaya çıkar.Kapsamlı araştırmalar MK'de birçok potansiyel biyobelirteç tanımlamış olsa da, hastalığın heterojenliği doğrulama için bir engeldir, bu nedenle kliniklerde yalnızca küçük bir sayıda, sınırlı uygulamalar görülür. RNA düzenleme, RNA dizisinde nükleotid değişiklikleriyle sonuçlanan epigenetik bir modifikasyondur. Adenozin'den İnozin'e (A-I) ikameleri, insanlarda en yaygın düzenleme olaylarıdır ve RNA (ADAR) enzimleri üzerinde etkili olan Adenozin deaminazları aracılık eder.İnosin(I), Guanozin'i (G) taklit eder ve Sitidin (C) ile çiftler oluşturarak RNA yapısı ve stabilitesinde, amino asit ikamelerinde, alternatif birleştirmede veya miRNA hedef bölgesi modifikasyonları yoluyla gen ekspresyonu düzenlemesinde değişikliklere neden olur. Meme, akciğer, böbrek, beyin ve mide kanserlerinde RNA düzenleme düzensizlikleri bulunmuştur, ancak biyobelirteçler olarak spesifik düzenleme olaylarının faydası büyük ölçüde keşfedilmemiştir. Bu çalışmada, MK'de A-I düzenleme olaylarının kemosensitivite ve prognostik biyobelirteçler olarak potansiyelini araştırıyoruz. Birden fazla veri setinde, analizimiz 305 benzersiz konumdaki RNA düzenleme olaylarının, MK hücre hatlarında bilinen 17 kemoterapötik maddenin ilaç duyarlılığı ölçümleriyle korele olduğunu göstermektedir. En önemli düzenleme olayı-ilaç duyarlılığı korelasyonları, daha yüksek düzenleme seviyelerinin daha yüksek kemosensitivite ile ilişkili olduğunu gösterir. İlginç bir şekilde, tanımlanmış düzenleme olayları olan genlerin ekspresyon seviyeleri, ilaç duyarlılığı ile daha zayıf bir korelasyona sahiptir veya hiç korelasyon göstermez, bu da düzenleme olaylarının transkript seviyelerinden bağımsız biyobelirteçler olduğunu ima eder. ADAR enzimleri düzenleme olaylarına aracılık ederken, ADAR ekspresyon

seviyelerinin MK'de kemosensitivite biyobelirteçleri olarak düzenleme frekansları ile değiştirilemez olduğunu gösterdik. C11orf80 kodlama dizisinde, bir amino asit ikamesi (S.p133G) ile sonuçlanan yanlış anlamlı bir düzenleme olayı keşfettik. Ayrıca, SOGA1'in 3'UTR'sinde artan SOGA1 ifadesi ile ilişkili olan bir düzenleme olayı belirledik. Bu düzenleme bölgesinin miR-9-5p'nin varsayılan bir hedef bölgesinde bulunması, gen ekspresyonunun miRNA hedef bölgesi modifikasyonları tarafından düzenlenebileceğini düşündürmektedir.

TCGA ve Singapur kohortlarında, MK'deki olayları düzenlemenin prognostik rolü araştırıldı. Genel olarak, daha yüksek düzenleme seviyeleri, MK hastalarında daha iyi sağkalım ile ilişkilidir. Her iki kohortta, CLPX geninde 65442098 konumundaki bir düzenleme olayının bağımsız bir iyi prognostik faktör olduğunu bulduk. Log Rank Multiple Cut-off (LRMC) çizim dağılımını kullanarak hastalarımızı "Yüksek" ve "Düşük" düzenlenmiş gruplara, en iyi şekilde kategorize edecek düzenleme olaylarını seçtik. Her veri setinde, MK hastalarında sağkalım ile bağımsız olarak korelasyon gösteren bir iyi ve bir kötü prognostik faktör olmak üzere iki düzenleme olayı öneriyoruz. Singapur Kohortunda, ZNF587'deki yüksek düzenleme seviyeleri iyi bir prognoz ile ilişkiliyken, DCAF16'dakiler kötü bir prognoz ile ilişkilidir. CTSB'deki yüksek düzenleme seviyeleri, TCGA kohortunda daha iyi genel hayatta kalma (OS) ile ilişkilidir, NUP43'tekiler ise daha kötü OS ile ilişkilidir. Transkript seviyeleri hayatta kalma ile korelasyon göstermediğinden, bu düzenleme olaylarının prognostik etkileri, gen ekspresyon seviyelerinden etkilenmez. Belirli pozisyonlardaki düzenleme seviyelerinin, prognostik düzenleme imza modellerine daha uygun maliyetli ve uygulanabilir bir alternatif sağlayarak, prognostik biyobelirteçler olarak önemli bir şekilde kullanılabileceğine inanıyoruz.Bulgularımız, düzenleme olaylarının mide kanserinde kemosensitivite ve prognoz için bağımsız biyobelirteçler olarak kullanılabileceğini gösteriyor, ancak gözlemlenen ilişkilerin altında yatan mekanizmaları aydınlatmak için daha fazla araştırmaya gerek duyulmaktadır

*Anahtar Kelimeler:* A-I, ADAR, kemosensitivite, Mide kanseri, prognoz, RNA düzenleme

***… To those who are driven by their passion for science***

# ACKNOWLEDGMENTS

First and foremost, I would like to extend my deepest gratitude to my mentor, Prof. Dr. Ali Osmay Güre, for guiding me with patience throughout my master journey. I would like to thank him for entrusting me with this project, for his valuable lessons, the constant support and most importantly for believing in me. It has been my privilege to work under the supervision of one of the most inspiring scientist, I have had the pleasure to meet.

Secondly I would like to thank my thesis committee members, Assoc. Prof. Özlen Konu and Assist. Prof. Onur Emre Onat for their counseling, invaluable input and kind words during this stressful, yet exciting period.

I would like to all past and present members of AOG lab; Shila Azizolli, Noor Niaz, Ronak Naeemaee, Farid Ahadli, Baris Küçükkaraduman, Ege Dedeoglu, Muhammad Waqas Akbar, Seçil Demirkol, Marzana Ishraq and Murat Isbilen. I am grateful to have been part of this wonderful team of talented scientist, and lucky to have worked in the healthy and happy work environment they have created. I appreciate their unsparing help and cherish all the good times we had together.

I would like to thank Noor, for being a caring friend and colleague. Thank you Noor for all the thoughtful little notes, the meaningful gifts, the unintentional jokes and late night conversations. You truly are one of the kindest souls I have ever met, and I am lucky to call you my friend. I would also like to thank Ronak, for always being there for me in the important moments, putting aside her projects and dog-sitting duties, just to give an honest advice, to play board games or help in whichever way possible. You have forever changed my gift-wrapping expectations and have given a special meaning to my perception of hot headed red-head.

I would like to extend my sincerest gratitude to MSc. Shila Azizolli. Thank you for always being by my side and making me believe in myself even when at my lowest. This experience would have not been the same without you. I am lucky to have shared the tears, the laughters, the highs and the lows with a wonderful person like you. I admire your drive, energy and passion you put in your work and will be impatiently waiting for you to reach the excellence peaks you are deserving of.

I would also like to thank my friends Aldo Tali, Franc Gripshi and Arli Lika for being a constant source of good energy and lifting my spirits even though miles apart. I would like to thank my dear friend Olivia Hafizi for always believing in me and knowing how to put a smile on my face. I am thankful for my friend Rezarta Shani who with her caring and uplifting talks was able to make even the bad days a little bit brighter.

I would like to thank my Turkish friends who have always made me feel at home around them. I thank Sezin Ünlü that with her bubbly personality has created some of the most beautiful memories I have in Turkey, which I will cherish forever. I admire her

determination and cannot wait to see her achievements. I would also like to thank Mustafa Akkaya, for all the fun times, the coffee runs, the endless conversations and the belly laughs. I am grateful for our friendship and truly believe you will become the best Doctor of Mathematics.

Last, but not least I would like to thank my family, to whom I owe all my achievements. I would like to thank my parents for their constant emotional support and for always believing in me. I thank them for engraving in me the desire to expand my horizons and teaching me that nothing is unreachable, if you put your mind to it. I would like to thank my sister for her unconditional love and for being an inspiring role model to me. I would like to thank my aunts for their encouraging words, and my grandmother for her endearing phone calls. I would like to also thank my grandfather, for his constant interest in every academic step of mine and for always keeping me in his wishes and prayers.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1

# 1. Introduction

The gastrointestinal tract or GI tract is a tube spanning the human body, responsible for supplying nutrients and water from the external environment. An essential organ of the GI tract is the stomach, a baglike organ that stores the food, mechanically and chemically digests it and protects the body from environmental pathogens that are swallowed with the food. The GI tract is characterized by a rapid turnover of the cells to produce new epithelium for the crypts and gastric glands, increasing the cancer susceptibility in these organs [1].

Gastric Cancer (GC) is a malignancy characterized by the uncontrolled growth of cells in the stomach [2]. Gastric adenocarcinomas, arising from glands in the stomach lining or mucosa comprise almost 90% of the gastric cancers, which is why the two terms are used interchangeably unless specified otherwise [3]. Other types of gastric cancer include those originating from mucosa associated lymphoid tissue (MALT-lymphomas) and other rare types namely leiomyosarcomas, originating from the muscle tissue surrounding the stomach wall and carcinoid tumors [4].

## 1.1 Gastric adenocarcinoma classifications

The stomach is subdivided into five parts; cardia, fundus, body, pylorus and the atrium. Based on their location, gastric adenocarcinomas can be classified as proximal (cardia) and distal (non-cardia) [3].Majority of the gastric cancer occurs in the distal area and only 18% of gastric cancers occur in the proximal area [5]. This site classification of gastric tumors is of relevance as the two types display different demographic patterns, associated risk factors and etiologies [6].

On a histological basis, Lauren`s classification subdivides gastric adenocarcinomas into two main subtypes: intestinal, characterized by the presence of neoplastic cells in gland-like tubular structures and diffused, in which cells infiltrate individually without forming a cohesive mass or structure [7]. One of the limitations of Lauren`s classification is that it cannot categorize 15-20% of the adenocarcinomas [8]. However, the fact that the diffused

and intestinal subtypes fail to transform into the other indicates that Lauren`s classification reflects their biological distinctions [7]. Intestinal and diffused subtypes also display differences in epidemiology and prognosis. Intestinal carcinomas are more prominent in high-prevalence areas (East Asia), display better prognosis and have a higher prevalence in older, male population. Diffused carcinomas are more frequent in low prevalence areas (North America, Western Europe), shows a worse prognosis and has a higher prevalence in younger individuals [7, 9]. Gastric cancer can be also classified based on the World Health Organization (WHO) classification system which recognizes four subtypes: papillary, mucinous, tubular and poorly cohesive (signet-ring cell carcinoma included) [10]. Aside from Lauren`s classification, WHO classification is one of the most frequently used.

The heterogeneous nature of gastric adenocarcinomas limits the clinical utility of these histological classifications, especially when it comes to treatment strategy selection. In an attempt to develop a more inclusive classification system, The Cancer Genome Atlas (TCGA) in 2014 suggested a model that divides gastric cancer into 4 molecular subtypes: Epstein–Barr positive (9% of the studied tumors), Microsatellite unstable (MSI, 22%), Chromosome Instable (CIN, 50%) and Genomic stable (GS, 20%) tumors. All subtypes are characterized by distinct molecular alteration, some of which could open new possibilities for targeted treatment strategies (e.g., EBV-positive tumors displayed a higher mutation rate in PIK3CA gene as compared to other subtypes, suggesting this subtype might be responsive to PIK3CA inhibitors) [11]. The Asian Cancer Research Group (ACRG) undertook a similar initiative in 2015 to classify gastric cancer into distinct molecular subtypes with clinical relevance. They identified 4 molecular subtypes: the mesenchymal-like (63%) displaying the poorer prognosis, the microsatellite instable (22%) with the best prognosis and TP53 active and inactive types with intermediate prognosis. Even though the TCGA and ACRG classification systems present similar subtypes (e.g., MSI), they also display differences in prognosis, molecular processes and driver genes [12]. These observed differences may arise from the cohort diversities (only 25% of the TCGA cohort was of Asian origin) or the different approaches used for the classification [13]. Although the mentioned classifications have provided insight into the molecular

landscape of gastric cancer, and the usefulness of subtyping in targeted therapy development, their benefits have yet to be seen in clinical trials.

## 1.2 Etiology

Gastric cancer, despite being one of the few cancer types directly linked to an infectious agent, has a complex etiology affected by life-style habits, genetic and environmental factors [3, 14].

### 1.2.1 Helicobacter pylori

H Pylori is a gram-negative bacterium that populates the gastric mucosa and it is able to initiate an immune response in the host. H pylori is mostly acquired at an early age possibly through oral ingestion, and it is persistent throughout the life-span of the carrier, when not treated [14]. H pylori infects more than half of the human population and it was listed as carcinogenic to humans in 1994 by the International Agency for Research on Cancer after considering several studies linking its infection to gastric cancer [15]. The inflammation induced by H pylori is always persistent after the infection as the bacteria tries to create a suitable environment and the host attempts to reduce the inflammatory response to avoid tissue damage. The outcomes of the infection vary from gastritis to gastric adenocarcinomas, as they are affected by bacteria strain virulence, host genetic factors, age of infection and nutrition [16]. Studies have reported that H pylori infection increases the risk of gastric cancer development anywhere from 2.1 to 16.7-fold, with these risks being higher in individuals infected by H pylori strains carrying the cytotoxin-associated gene A (cagA) [6]. It is reported that 65-80% of gastric adenocarcinomas are caused by H pylori, however the infectious agent is a major risk factor for distal gastric cancer and not the proximal type [17].

### 1.2.2 Lifestyle

Consumption of salt and salt-preserved foods has been reported as one of the probable causes of gastric cancer by the American institute for Cancer Research [18]. Based on case-control study in 2012, high intake of salt leads to a 22% higher risk of gastric cancer [19]. Besides salt, smoked food and meat consumption have also been associated with gastric cancer as potential risk factors. Diets rich in fruit and vegetable, anti-oxidants and fibers

on the other hand have been associated with a lower risk of gastric cancer. Similarly, to other cancer types, smoking and tobacco use have been reported as risk factors for gastric cancer as well [20]. Despite studies being contradictory on the status of alcohol consumption as a risk factor, a more recent systemic meta-analysis reports that alcohol consumption can increase the risk of gastric cancer [21]. Obesity, although not associated with distal gastric cancer, is a major risk factor for proximal GC as obese individuals tend to develop gastroesophageal reflux disease (GERD), which is a risk factor for both esophageal and gastric cancer. Increased physical activity on the other hand is reported as a possible risk reduction factor [3].

### 1.2.3 Genetic factors

Genetic syndromes such as hereditary nonpolyposis colorectal cancer or Li-Fraumeni syndrome are established risk factors for gastric cancer. Not only individuals positive for these conditions but also individuals related to them, are at a higher risk of developing the disease [6, 22]. Family history of gastric cancer itself is an added risk factor, especially in first-degree relatives. Apart from the inherited genetic conditions, genetic polymorphisms at cytokine coding genes have been reported as a risk factor for gastric cancer [20].

Other potential risk factors include infection with Epstein-Barr virus, type A blood and previous gastric surgery [6].

## 1.3 Demographic Trends

As of 2020, gastric cancer remains one of the most common cancer types ranking 5$^{th}$ and one of the leading causes of cancer mortality ranking 4$^{th}$ on a global scale [23]. While food preservation practices, dietary habits, and economic advancements have contributed to a decrease in gastric cancer incidence and mortality over the last decades, the fatality rates remain high, with 1.27 million deaths expected by 2040 [24]. However, these global trends are variable for gastric cancer. For the most part, distal GC incidence is higher than proximal GC incidence, although some variations are observed in different populations [25]. Geographically, Eastern Asia, Central-Eastern Europe and South America have the highest incidence rates while North America and Africa have the lowest in both men and women. Mortality rates show similar trends, as East Asian countries rank the highest and

North American countries the lowest [26]. Variations are observed between genders and age groups as well. GC incidences for males are twice as high as for females, and older age groups (age ≥ 60) account for most of the incidences [24]. Contradictory to the decreasing trends in gastric cancer, an increased incidence has been reported in younger age groups (age <50) in both, high and low incidence areas, suggesting that research on preventive and diagnostic strategies is still required [27].

## 1.4 Prognosis

Depending on the invasiveness level of the cancerous cells through the stomach layers, gastric cancers can be classified into early-stage and advanced stage. If the lesions are present in the mucosa or submucosa layer it is considered early-stage gastric cancer, and if the lesions extend beyond the submucosa to invade the muscular layer and beyond to metastasize to other organs are considered advanced-stage gastric cancer. The stage of the cancer will condition the treatment strategy and response, eventually affecting the patient survival. Diagnosis of GC at an early-stage allows for radical surgical treatment, with a positive post-operative 5-year survival of 90%. The lack of symptoms in the early stages of the disease, however, has resulted in a low diagnosis rate, and more than 70% of patients present with advanced disease [28]. Consequently, the prognosis in gastric cancer patients remains poor. In most parts of the world the 5-year relative survival rate is 20% and in metastatic GC patients the median OS is less than 1 year [14, 29]. Interestingly, GC survival displays better rates in areas of high prevalence when compared to low prevalence areas [30]. These differences might be a reflection of the effectiveness of mass screening programs employed by high incidence countries such as Japan or the prognostic differences of the tumor location, as proximal tumors show a lower 5-year overall survival (OS) rate than distal tumors, which remains the predominant type in Japan [3, 25]. Some other prognostic factors are tumor grade and histology, as diffused type, higher grade tumors display a poorer prognosis. However, these factors are dependent on the disease stage, and fail to be very informative on the patient prognosis [13]. Thus, there is still a need for prognostic biomarkers to improve the treatment and subsequently the survival rate of gastric cancer patients.

## 1.5 Treatment Strategies

Treatment regimens of gastric cancer do not follow a certain protocol but are guided by disease stage, biomarker presence and physician suggestions. Staging based on TNM system, which describes the invasion of through the stomach layers (T), to lymph nodes (N) and metastasis presence (M), has significantly improved decision making in gastric cancer treatment [31]. Typically, surgery is the initial line of treatment for gastric cancer, however surgical resection alone has proven insufficient especially for advance stage patients. In early-stage patients surgery is still a successful treatment option with a reported 5-year survival rate of 83% in European populations and over 90% worldwide. As the tumor infiltrates through the submucosa layer, the survival rate decreases as shown by the 25% 5-year OS rate in European patients with advanced GC (T stage III-IV N0 or any T stage N+) who undergo surgery alone [13].

### 1.5.1 Chemotherapy

Addition of chemotherapeutic treatment pre- or post- surgery has shown a positive outcome in patient survival. Survival rates increased by 10% in stage II-II patients that received adjuvant chemotherapy as compared to those who underwent surgery only [32]. The MAGIC (Medical Research Council Adjuvant Gastric Infusional Chemotherapy) clinical trial in UK reports similar trends as stage III patients who received Epirubicin/Cisplatin/ Fluorouracil (ECF) regimen had a 36% 5-year OS meanwhile patients treated with surgical resection only displayed a 23% 5-year OS. Based on the positive outcomes of many clinical trials, administration of pre- or perioperative chemotherapy has become standard practice for locally advanced gastric cancer [33]. Due to the limited effect of single chemotherapeutic usage, a combination of two or three agents is the general procedure. The most frequently used combinatorial chemotherapeutic regimens include; ECF, EOX (Epirubicin/Oxaliplatin/Capecitabine) and FLOT (Docetaxel/5FU/Leucovorin/Oxaliplatin) [31].

About 50% of gastric cancer cases present at a metastatic stage and treatment consists mainly in palliative chemotherapy. First-line treatment in metastatic GC is a combination of a platin-agent (Oxaliplatin) and a cytotoxic compound (5-FU) [13]. In case of disease

progression, a second-line treatment of paclitaxel in combination with Ramucirumab (VEGFR-2 antibody) is suggested [31]. Unfortunately, despite the advances in chemotherapeutic treatments the median OS remains dismal, ranging from 10 to 12 months [13].

Even though improvements in chemotherapeutic regimens have been made, the survival rates in GC are still low mainly due to the high heterogeneity of the disease and the chemo-resistance most advanced GC display [31]. Besides not benefiting from it, patients unresponsive to adjuvant chemotherapy are unable to tolerate second- or third-line treatment, therefore more research must be conducted to be able to identify chemorefractory tumors [13].

### 1.5.2 Targeted therapy

As our understanding of cancer biology evolves, it has become apparent that developing small molecules to target tumorigenic pathway components is a promising treatment strategy. Molecular targeted therapies that have been assessed in gastric cancer include growth factor inhibitors, cell cycle inhibitors, angiogenesis and matrix metalloproteases inhibitors [28]. However, so far only two targeted therapies have been approved by FDA for gastric cancer, Trastuzumab and Ramucirumab, inhibitors of HER2 and VEGFR2 respectively [13].

HER2 (Human Epidermal Growth Factor receptor 2) is a member of growth factor receptor family and is overexpressed in 20% of gastric cancer tumors. Trastuzumab (HER2 monoclonal antibody) addition to chemotherapy regimens has shown improvements in HER2+ gastric cancer patient survival. The ToGA trial reported a better relative risk (RR) (47% vs 35%), progression free survival (PFS) (6.7 months vs 5.5 months) and OS (13.8 months vs 11.8 months) in GC patients treated with Fluorouracil/Cisplatin with vs without Trastuzumab addition [13].

Vascular epithelial growth factor (VGFR) is a preferred target since new vessel formation (angiogenesis) is crucial for tumor growth and metastasis formation. Ramucirumab is a monoclonal antibody for VEGF receptor 2 (VEGFR2) which has been approved as a second line treatment after the REGARD phase III trial reported improved median OS of

1.4 months in treated patients. Despite the efficacy in monotherapy treatment, Ramucirumab has been beneficial in combination with chemotherapy as well [34].

In spite of the positive outcomes of targeted therapy in advanced GC treatment, many other molecular drugs have failed clinical trials, making treatment a challenge. Furthermore, these therapies are effective to selective groups such as HER2+ GC, which cover only a small percentage of the patients.

## 1.6 Gastric Cancer Biomarkers

Research has shown biomarkers are useful indicators of disease progression, treatment response and prognosis in different cancer types. In GC, CEA and CA19-9 are the most commonly used biomarkers by physicians. CEA (carcinoembryonic antigen), is widely used as a biomarker for GI tract tumors and in gastric cancer it is an independent predictor of liver metastasis recurrence. CEA has been reported as a recurrence biomarker for peritoneal metastasis as well [35]. Elevated serum levels of CEA in preoperative GC patients have been associated with poor prognosis in different studies [36, 37]. CA19-9 (carbohydrate antigen 19-9) although mainly recognized as a diagnostic biomarker for pancreatic cancer, it is associated with TNM and disease stage in gastric cancer [35]. CA19-9 is most frequently reported as a lymph node involvement marker with a predictive value ranging from 78% to 96% [36]. CA19-9 is also considered a recurrence biomarker with a sensitivity value of 56% and a specificity value of 74%. Other diagnostic biomarkers in GC include CA72-4, alpha-fetoprotein (AFP) and CA125 which are mainly associated with recurrence, liver metastasis and peritoneal dissemination respectively [35].

As for molecular biomarkers, HER2 is the first one to find application in clinics for GC [35]. Together with VEGFR2 they are used as predictive biomarkers for response to targeted therapy due the development of targeting agents against them [35]. Besides being a predictive biomarker, HER2 overexpression has been linked to poorer survival in GC. Several studies report HER2 as an independent prognostic factor for disease free survival, OS and in early-stage carcinomas it is considered as the 2$^{nd}$ strongest indicator of poor survival after lymph node involvement [37].

FGFR2 (Fibroblast growth factor receptor) and MET are both transmembrane receptors that upon binding with their respective ligands, initiate important signaling pathways governing cell proliferation, angiogenesis, migration and invasion [35]. Overexpression of FGFR2 and MET has been reported in 31.1% and 24.9% of gastric adenocarcinomas, a higher percentage than HER2 overexpression [38]. High levels of FGFR2 have been associated with higher staging, poorer OS and metastasis [35]. In grade 3+ patients who have received curative surgery and adjuvant chemotherapy, FGFR2 expression correlates positively with recurrence post 5-year post-op mark [39]. MET expression is also related to poorer OS, and might be a promising indicator of chemo-resistance as the usage of MET inhibitor has been shown to have an anti-proliferative effect in irinotecan resistant cells [40].

E-cadherin, is a transmembrane molecule important for cellular adhesion and maintenance of epithelial state, thus it is crucial tumor suppressor [41]. The protein is coded by CDH1 gene, which is frequently mutated in hereditary diffused gastric cancer (HDGC) an autosomal dominant condition accounting for a low percentage of GC (around 2%) [42]. However, a study in which both familial and sporadic GC cases were included, reported that deletions and mutations in CDH1 gene relate with a poorer survival rate [43]. Epigenetic modifications on CDH1 promoter have been suggested as a diagnostic in gastric cancer as well [44].

PD-L1 (programmed death-ligand 1) is an important regulator of the immune response as upon binding on PD1 (programmed death 1) it inhibits the immune response of activated T cells, a mechanism frequently exploited by tumors including gastric cancer [45]. PD-L1 expression is reported in 15-70% of cases and it is associated with poor survival rates [46]. Apart from having a prognostic value, PD-L1 is considered a promising target in GC treatment. Phase II trial KEYNOTE-059, reported a better response to Pembrolizumab (PD-L1 monoclonal antibody) in the PD-L1 expressing arm compared to the PD-L1 negative [47]. PD-L1 expression could possibly be a therapy response biomarker for Pembrolizumab therapy [48].

Numerous biomarkers have been reported in Gastric Cancer, however due to the heterogeneous nature of the disease, there is a lack of validation and only a few of them

are currently used in clinics. Early-stage detection of the disease is still a major burden due to the fact that advance GC patients respond poorly to the available treatments. The presented facts indicate a prevailing need for diagnostic, prognostic and treatment-response predictive biomarkers.

## 1.7 RNA Editing

In 1986, an unprecedented RNA modification was observed in a mitochondrial mRNA transcript making it different from its coding DNA sequence. This post-transcriptional modification could reverse an existing mutation and restore the function of the mRNA, and based on the ability to edit the RNA sequence, it was named RNA editing [49]. In the early days, the identification of RNA editing events was difficult and time-consuming due to the lack of advanced methodologies. However, with the evolution of technology in science and invention of high throughput next generation sequencing, many RNA editing sites were identified in several organisms, from bacteria to humans [50]. As indicated by its presence in different organisms, RNA editing is an evolutionary conserved process, however the reasons for its existence are still debated. The most apparent reasons are to increase protein diversity and reverse occurring genomic mutations [51]. In vertebrates the most prominent editing events are deamination of cytidine and adenosine bases, making deaminases the key RNA editors [50].

### 1.7.1 C-to-U Editing

Deamination of cytidine is carried out by activation-induced cytidine deaminase/apolipoprotein B editing complex (AID/APOBEC) family of enzymes which convert cytidine to uracil or deoxycytidines to deoxy uridines, causing a nucleotide change in the RNA or DNA sequence [52]. There are 11 AID/APOBEC deaminases present in humans, expressed in different tissues and cellular locations [50]. Some cytidine deaminases show preference for specific sequences such as APOBEC1 for AU-rich regions in the 3'UTR of mRNA transcripts or APOBEC3 enzymes for SINE/LINE regions [53, 54]. AID/APOBEC deaminases are implicated in important immune processes such as antibody diversity and antiviral response, mainly by acting on DNA sequences [50].

### 1.7.2 A-to-I Editing

A-to-I editing is the most prominent type of RNA editing and millions of A-to-I editing events have been identified in humans [55, 56]. Hydrolytic deamination of adenosine at the C6 position by the Adenosine deaminases acting on RNA (ADAR) family of enzymes, results in its conversion to inosine, which is recognized by the translational machinery as guanosine and why it is often referred to as A-to-G editing. Inosine is then paired with cytidine in double stranded structures [57]. Vertebrates have 3 ADAR enzymes; ADAR1, ADAR2 and ADAR3. ADARs share some functional domains such as the double stranded RNA (dsRNA) binding domain and the deaminase domain at the C-terminal, but they also have specific unique domains whose functionality is still unclear. ADAR1 and ADAR2 have been shown to have enzymatic properties and they require homodimerization via their dsRNA binding domain to exert their A-to-I editing activity. ADAR3 deaminase activity has not been reported yet, and the fact that ADAR3 lacks the ability to homodimerize might be the reason why. In humans ADAR1 is expressed in all tissues, while ADAR2 is expressed in some tissues, but most highly expressed in the brain. ADAR3 is exclusively expressed in the brain. ADARs predominantly locate at the nucleus, however ADAR1 is found in the cytoplasm as well. ADAR1p150 isoform is the ADAR1 isoform mainly found in the cytoplasm, where it translocates by binding to exportin-1 and edits sites on the 3'UTR of dsRNAs. ADAR1p110 isoform is mostly located in the nucleus and is constitutively expressed. ADAR2 is degraded by ubiquitination in the cytoplasm, so it can be considered nuclear [58].

ADARs bind to dsRNAs longer than 20 nucleotides in length and can edit 50% of the adenosines in a fully base-pair matched sequence, but only a few sites in partially complimentary sequences, suggesting a possible selectivity for certain editing sites based on the secondary structure of RNA [59, 60]. No conserved motif has been identified in the ADAR editing sites however preferences for sites neighboring 5` uridine and 3` guanosines has been noticed. Some sites can be edited by both ADAR enzymes and some others are specifically edited by ADAR1 or ADAR2 [58]. Generally, ADAR1 is the main enzyme acting on repeat elements and ADAR2 acts on non-repetitive coding regions [61]. A-to-I editing mainly occurs in non-coding regions and just a few in the coding regions result in

amino acid sequence alterations. Most editing sites are located on introns or UTRs (untranslated region) of coding genes, especially in short repetitive elements such as ALU repeats [62]. ADAR proteins are responsible for editing events in long non-coding RNAs and microRNAs as well [50]. A-to-I editing events can regulate the transcriptomic landscape by affecting several processes. Editing in pre-mRNA transcripts can create new splice sites or undo existing ones, thus affecting the alternative splicing process and the expression of certain isoforms [63]. The implication of RNA editing in alternative splicing has been reported in different organisms and more recently it has been hypothesized that editing in intronic repetitive sequences is involved in a Back-splicing, a special type of alternative splicing, which gives rise to circular RNAs [64]. Besides alternative splicing, A-to-I editing regulates transcript stability by modifying the secondary structure of RNAs and the binding ability with processing molecules[65]. Editing events can also influence transcript levels by modifying miRNA binding sites on the 3'UTR regions or miRNA sequences [66]. Studies in different model organisms and high levels of ADAR proteins in brain tissue suggest that RNA editing plays an important role in neurotransmission, neural development, and signal transduction [50].

## 1.8 A-to-I Editing in Cancer

Due to the functional consequences RNA editing has on the transcriptome, it comes as no surprise that research has shown that editing is implicated in tumorigenesis and cancer progression. Differences in editing levels have been shown in different cancers when compared to their matched normal tissue. Cancer tissues can be hypo-edited, hyper-edited or mis-edited in case no global pattern is observed, and these differing editing levels can yield different outcomes in cancer progression [50]. Editing events and ADAR levels have been associated with tumorigenesis, cell proliferation and migration, metastasis, apoptosis inhibition, chemo-resistance and prognosis in different cancer types such as breast, lung, esophageal and glioblastoma [67]. In 2015, Han.et al elucidated the pan cancer A-to-I editome landscape of 17 cancer types from the TCGA dataset. This study showed that editing events, including those in non-coding regions, were associated with tumor subtype, stage and patient survival in 12 of the cancer types. Also, non-synonymous A-to-I editing events have an effect on drug sensitivity [68]. Overall, these findings indicate that RNA

editing should be researched as a promising biomarker that could be applied in clinics for different cancer types.

RNA editing has been investigated in gastric cancer as well. GC tissue displays a mis-edited phenotype and these editing events are mainly mediated by ADAR1, as studies show that ADAR1 knockdown results in a decrease in editing. ADAR expression is dysregulated in GC, as ADAR1 is overexpressed and ADAR2 is downregulates. Patients displaying this pattern of expression of ADAR1 and ADAR2 had a shorter OS and transition from healthy to cancerous tissue was associated with an increase in editing levels in ADAR1 sites and a decrease in ADAR2 sites, indicating that ADAR1 has an oncogenic role while ADAR2 acts as a tumor-suppressor [69]. ADAR1 can also regulate transcript levels by editing the 3'UTR of transcripts to create miRNA seed sites in GC [70]. The potential of editing events as biomarkers has not been explored much in GC. Elevated editing levels in AZIN1 gene were identified as an independent prognostic factor, relating to poorer OS and DFS [71]. Recently, An and colleagues reported an RNA editing signature composed of 50 editing sites that could serve as a prognostic and predictive biomarker in advanced GC patients [72]. Even though these results are promising in regards to the effectiveness of utilizing RNA editing events as biomarkers, their application in clinics does not seem feasible due to the cost of detecting several editing events.

# CHAPTER 2

# 2. Aim

In this study we aim to show the potential of editing events as independent chemosensitivity and prognostic biomarkers in Gastric Cancer. Firstly, we demonstrate clear associations between drug sensitivity and editing levels at specific positions in GC cell lines. Based on the fact that the effect of editing events is closely related to their location we want to identify editing event-drug sensitivity combinations consistently correlating in different drug response datasets. Furthermore, we want to show that editing events are chemosensitivity biomarkers, independent of ADAR or gene expression levels.

Lastly we aim to identify editing events associated to survival in GC patients, and that can be used independently as prognostic biomarkers. We also want to show that even single editing events can be utilized to predict survival outcomes.

# CHAPTER 3

# 3. Methods

## 3.1 RNA Editing Analysis

Cancer Cell Line RNA-sequencing was performed using Ilumina HiSeq2000 or HiSeq 2500 with sequencing coverage > 100 million paired, 101 base pair long reads per sample [73]. The sequencing data obtained in FASTQ format are aligned to the reference genome using STAR align and undergo a two-quality check, pre-alignment quality assessment using FASTQC, and post-alignment using Picard Tool as described in the Genomic Data Commons (GDC) documentation (https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/). RNA sequencing data, aligned in BAM format for 37 gastric cancer cell lines were downloaded from the GDC data portal, using the GDC Data Transfer tool as suggested by the manual (https://gdc.cancer.gov/access-data/gdc-data-transfer-tool). BAM files were sorted, indexed and duplicates were removed using samstools v1.3.1 [74]. RediKnown.py script of REDItools v1.0.3 was used to quantify and identify previously reported editing events, using their genomic positions and the gastric cancer cell lines BAM files as input [75]. Genomic coordinates of known editing events were obtained by combining DARNED (https://darned.ucc.ie/download/) and REDIportal (http://srv00.recas.ba.infn.it/atlas/download.html) locations and given as an input in a TAB file [76, 77]. The hg19 assembly genome was obtained from UCSC database in FASTA format and used as reference. The number of A, T,C,G bases encountered at the positions matching with the coordinates of known RNA editing events in the reference list is reported. In sites where coverage was less than 10 reads and read quality score less than 25 (accuracy >99%), NA instead of number of each nucleotide was reported, as recommended by the developer [78]. To avoid misalignment errors, minimum mapping quality was set to 255 (value used to mark uniquely mapped reads in RNA-sequencing) and exclusion of multiple mapping reads was applied [79]. Editing events identified in a 4 nucleotide proximity to splice sites or in homopolymeric regions longer than 5 nucleotides in length are not included to rule out, false positive outcomes as a result of alignment or sequencing

mistakes. To avoid sequencing biases at the read ends, the first 5 nucleotides on each end were trimmed and editings in such positions were excluded. Filters were not applied on the minimum number of edited reads identified or minimum number of editing frequency since such parameters might be restricting and are mainly recommended for editings in non-Alu regions [78]. AnnotateTable.py script of REDItools was used to annotate identified editing positions using RefSeq Gene annotations from downloaded from UCSC [79]. The base counts, type of nucleotide changes observed, and editing frequencies, calculated by the ratio of number of edited nucleotides over overall nucleotide counts at a specific site, were reported for all identified positions across all samples. To rule out the possibility of undetected SNPs by REDItools, genomic coordinates of editing sites selected in further analysis were manually checked for reported variations in dbSNP (https://www.ncbi.nlm.nih.gov/snp/).

## 3.2 Drug Sensitivity Data

Drug sensitivity values measured by AUC (Area Under the Curve) and Activity Area metrics for all tested drugs were downloaded from the respective portals, GDSC (https://www.cancerrxgene.org/downloads), CTRP (https://ocg.cancer.gov/programs/ctd2/data-portal/) and CCLE (https://depmap.org/portal/download/). The data were filtered for selection of interested samples and compounds, and no transformation was applied to the published measures.

## 3.3 Correlation Analysis

Pearson and Spearman Correlations were computed using cor.test () function of R base package (https://www.R-project.org). In the drug sensitivity-editing frequency correlation analysis, an additional filter was applied using R logical operators in which correlations with less than 6 paired samples were not computed. To avoid correlations that do not reflect associations in the sample population, correlations with fewer than six observations were assigned a not applicable ("NA") value .Correlation p-values were adjusted using generic function p.adjust (), using Benjamini-Hochberg and Bonferroni correction methods. The scatter plots for the correlations were drawn using ggplot () function from ggplot2 package in R.

## 3.4 RANKSUM Analysis

The RANKSUM analysis was performed using a self-developed R script. To retrieve common editing events in between datasets, intersect () function from R base package was used and logical operators were used for the selection. A rank value was assigned based on descending absolute value of Spearman rho using rank () and order () functions from BiocGenerics package. The "Rank" vectors in each dataset were summed and ordered in an ascending fashion.

## 3.5 Gene Expression Data

Level 3, TPM normalized RNA sequencing data for GC cell lines were downloaded from CCLE portal (https://depmap.org/portal/download). The TPM expression levels were log2 transformed using R. miRNA expression data for GC cell lines were downloaded from CCLE, via Broad DepMap Portal (https://depmap.org/portal/download/). Level 3, TCGA-STAD RSEM normalized gene expression data were retrieved from Firebrowse portal (http://firebrowse.org).

## 3.6 Log-Rank Multiple Cut-Off (LRMC) Analysis

Using an in-house R script, editing frequency levels for each sample are used as cut-off to categorize patients into "high" and "low" edited groups. After survival curves are created, the differences between them are calculated using the log rank test. The negative logarithm with base 10 of the p-values are plotted on the y-axis, against the corresponding editing frequency used as a threshold to stratify the patients. The horizontal dashed line corresponds to p-value significance cut-off 0.05 and the vertical lines correspond to the 1$^{st}$, 2nd and 3$^{rd}$ quartile. Blue colored points represent positive and red colored points represent negative correlation to survival.

## 3.7 Survival Analysis

Univariate and Multivariate Cox regression analysis were performed using coxph () function from the survival package in R. Multiple logistic regression analysis was performed using SPSS Statistics 19 (IBM, Chicago, IL, USA). GraphPad Prism 8.0.1 was

used to plot Kaplan-Meier curves and compute the survival differences via Log Rank (Mantel-Cox) test ($p<0.05$ was used as significance threshold).

## 3.8 Editing Frequency Data from Survival Analysis Cohorts

Singapore Cohort RNA samples were sequenced using Illumina HiSeq 3000 (150 bp pair-end). After trimming raw reads using Trimmomatic (version 0.38), reads equal or longer than 35 nucleotides and with an average read quality higher than 20 were retained. Trimmed reads were aligned to hg19 human reference genome using Burrows–Wheeler aligner (BWA v.0.7.17) and its default parameters. Samtools (v 1.9) was used to remove duplicates and a mapping quality filter of 20 was applied. Substitutions supported in at least 2 reads were selected and reported SNPs were removed after crosschecking their presence in different cohorts (1000 Genomes Project, NHLBI GO Exome Sequencing Project, dbSNP). Editing events in the first 6 bases of the reads were excluded to avoid priming errors. Editings position in 4 nucleotide proximity to splices sites or in simple repeats and homopolymeric regions were discarded, if the reported site is located in non-Alu repeats. Events reported in reads mapping to non-unique genomic regions were removed using BLAST-like alignment tool. Lastly, only A-to-G selected editing events were annotated using Annovar and USCS *refGene* table [72].

The TCGA stomach adenocarcinoma (STAD) project samples were submitted to Illumina HiSeq 2500 (75 bp pair-end) sequencing. Reads were aligned using BWA (v.0.5.7) with default parameters except Smith-Waterman alignment [80]. Picard MarkDuplicated tool was used to remove PCR duplicates [81]. Editing frequencies at 1.4 million editing sites reported in RADAR were measured and only sites with base quality $\geq 20$ were considered. A site was considered edited if the substitution was present in at least 3 reads in 3 samples per tissue. The positions were re-annotated using Annovar and around 4000 sites reported in dbSNP (v 137), COSMIC and TCGA somatic mutations were excluded [68]. Other filtering conditions such as mapping quality $\geq 20$, exclusion of editings in non-unique genomic region using BLAST-like alignment tool, and for editings in non-ALU regions, exclusion of positions in homopolymeric regions longer than 5 nucleotides, in simple repeats, in hypervariable genomic parts and positions located within 4 bp of a slice site

were applied [81]. For further analysis, editing events with a coverage ≥ 10 in at least 30 samples/ tumor type were selected [68].

## 3.9 Prediction Tools

To predict the location of hsa-miR-9-5p target site in the 3'UTR region of SOGA1 TargetScan website was used (v7.2; targetscan.org). To predict the functional consequences of the non-synonymous amino acid substitutions caused by editing events, PredictSNP tool was used (https://loschmidt.chemi.muni.cz/predictsnp1/) [82].

## 3.10 Network and Pathway Enrichment Analysis

To analyze the potential functional association of the genes edited in positions whose levels correlate with sensitivity measures of the same drug, STRING database (v 11.5) was used (https://string-db.org/) [83]. Multiple protein analysis was conducted by giving a list of protein names as input and only pathways with a p-value < 0.05 were considered to be significantly enriched.

To explore the miRNA-target gene interactions of the identified genes with edited positions in the 3'UTR showing a significant correlation with patient prognosis, miRNET visualization platform was used (https://www.mirnet.ca/). The list of genes was given as a query and networks based on miRNA-target interactions in gastric tissue were visualized. The functional implication of the interactions was explored using miRNA Function and Disease database [84].

# CHAPTER 4

## 4. Results

### 4.1 Identification and Selection of A-to-I Editing Events in Gastric Cancer Cell Lines

#### 4.1.1 Landscape of identified A-to-I editing events in Gastric Cancer cell lines

RNA sequencing files of 37 Gastric Cancer (GC) cell lines were submitted to an editing analysis to explore the editing potential of known editing events, using RediKnown.py function of REDItools package with pre-set filtering conditions (**Appendix A**). REDItools was able to identify a total of 553934 editing sites from the GC cell line RNA-sequencing data, out of which 107850 were found to be edited in at least one of the samples. On average 25161 edited sites were identified per sample.

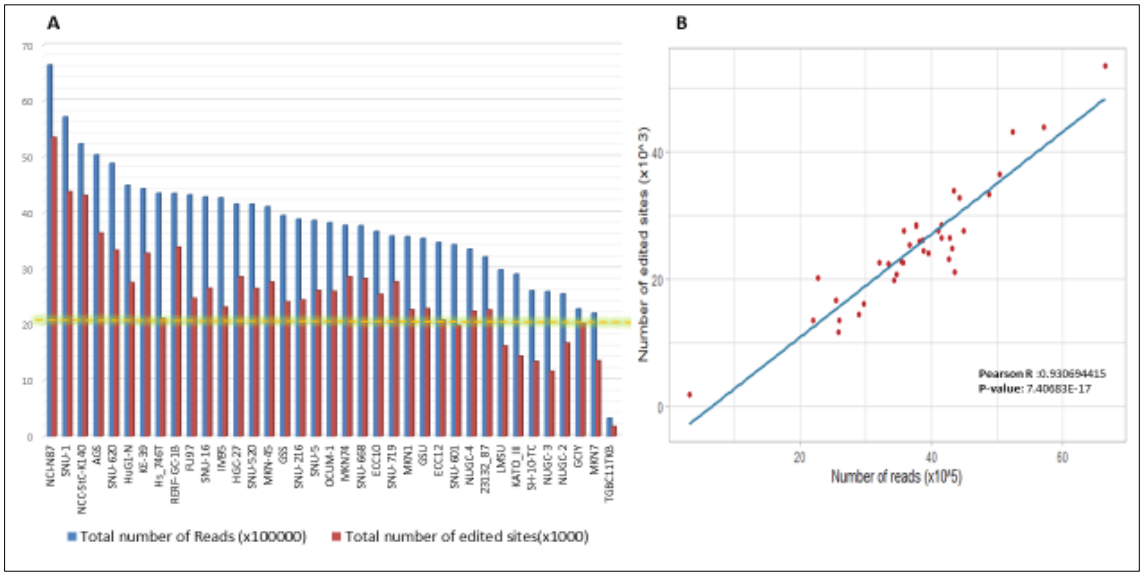



**Figure 4.1:** Overview of A-to-I editing events in GC cell lines. (A) Distribution of reads and edited sites in GC cell lines. Applied cut-off depicted in dashed yellow line. (B) Correlation between total number of reads and number of identified edited sites across different cell lines.

A strong correlation was observed between the number of reads and number of identified edited sites per sample (**Fig4.1B**, Pearson r= 0.93, p-value = 7.40683e-17). In order to maximize the information gathered from editing events in further analysis, cell lines with a total number of edited sites lower than 1$^{st}$ quartile (20.000) were discarded and 29 cell lines remained (**Fig4.1A**).

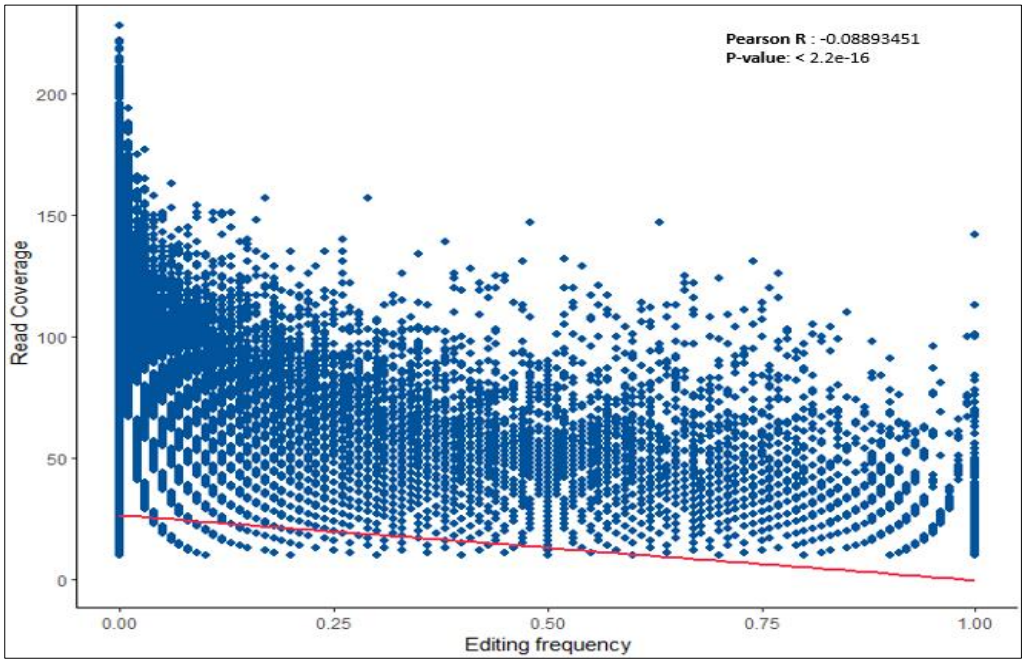



**Figure 4.2**: Correlation between read coverage and editing frequency in GC cell lines.

Editing frequency, the ratio of the number of edited reads to the overall number of reads at a specific position, was used as a measuring unit for editing levels. Despite the fact that the total number of edited sites is correlated to the number of reads, we observed that the editing levels were poorly correlated to the coverage (number of reads) at that specific site (Pearson R=-0.089, p< 2.2E-16) (**Fig4.2**).

### 4.1.2 Selection of editing events in GC cell lines

To provide an inclusive editing landscape in our analysis, we selected sites edited in more than 50% of the GC cell lines.

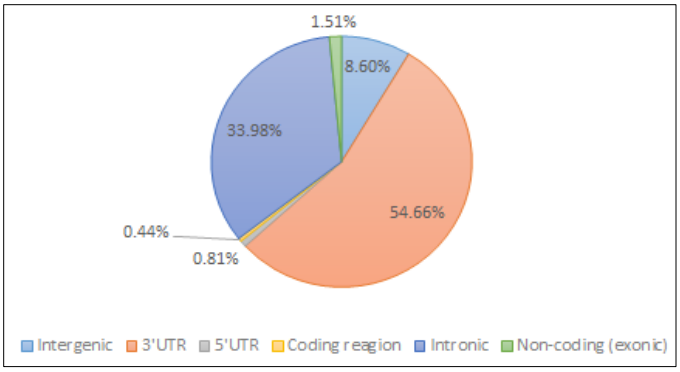



**Figure 4.3**. Location distribution of the 10537 selected A-to-I editing

A total 10537 editing events were selected, most of which were located in non-coding regions like 3'UTR and introns (**Fig4.3**).

ECC10, ECC12 and MKN1 cell lines were also omitted from the analysis due to them originating from rare gastric cancer types and failure to classify as stomach adenocarcinoma according to TCGA tumor classification (Appendix B). The chemosensitivity analysis was performed with editing data from the 26 selected GC cell lines and a pool of 10537 editing events (**Fig4.4**).

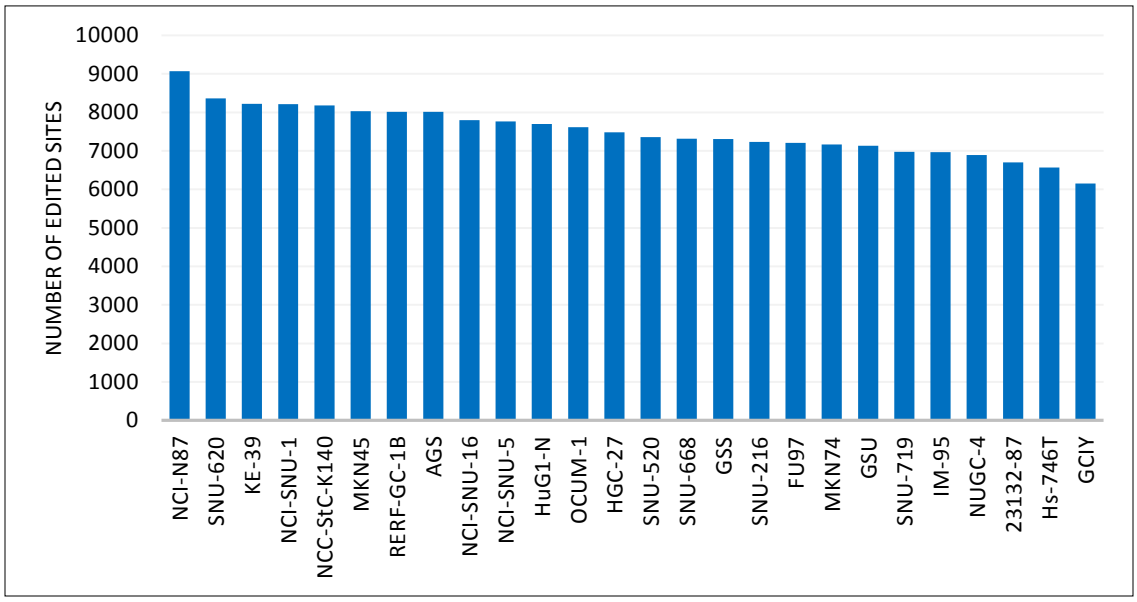



**Figure 4.4**: Distribution of number of edited sites across selected GC cell

## 4.2 A-to-I Editing Events as Chemosensitivity Biomarkers in GC Cell Lines

### 4.2.1 Editing events frequencies correlate with drug sensitivity in different drug response datasets.

The role of A-to-I editing events as predictive biomarkers for chemotherapy response was tested in GC cell lines due to the higher availability of drug response data. For our observations to be translatable in clinics, we decided that the focus of our analysis would be a selected list of 29 Drugs, FDA approved as treatment options for gastric cancer or other cancer types (**Table4.3**). Three independent drug response datasets, GDSC, CTRP and CCLE were used in the analysis to ensure the validity of the observed

associations. Both versions of GDSC dataset were used and due to the differences in screening, GDSC1 and GDSC2 were considered as two separate datasets (**Table4.1**).

**Table 4.1:** Summary of the drug sensitivity datasets used in the analysis. GC Cell Line = Number of the GC cell lines used; Drugs Total = Total number of drugs tested; Approved Drugs = Number of selected approved drugs tested; Editing Cell Line = Number of Cell Lines that have editing information; Available measure = Drug sensitivity measures available; Measure = Drug sensitivity measures that were used. AA = Activity Area; GDSC1 = GDSC1 Version 1; GDSC2 = GDSC Version 2

| Dataset | GC Cell Line | Drugs Total | Approved Drugs | Editing Cell Line | Available Measure | Measure |
|---|---|---|---|---|---|---|
| **CCLE** | 19 | 24 | 6 | 13 | AA<br>Amax<br>IC50<br>EC50 | AA |
| **CTRP** | 32 | 545 | 29 | 23 | AUC | AUC |
| **GDSC1** | 28 | 345 | 29 | 15 | AUC<br>IC50 | AUC |
| **GDSC2** | 28 | 175 | 18 | 15 | AUC<br>IC50 | AUC |

To check whether editing frequencies at specific sites associate with drug sensitivity, a correlation analysis was performed between the sensitivity measures (Area under the drug dose-response curve, AUC and Activity Area for CCLE) of the 29 selected drugs and editing frequencies at the 10537 selected sites. The analysis was repeated for all the datasets and a list of drugs, whose sensitivity significantly correlated with editing frequencies at least at one editing site was obtained (Benjamini-Hochberg or Bonferroni adjusted p-value <0.05). The lists from each dataset were then crosschecked for same Drug-Editing events combination, however we were not able to identify any (**Table 4.2**).We observed that the response to certain drugs such as Dasatinib, Fulvestrant and Pazopanib, is consistently related to editing frequency in GC cell lines in different datasets, even though not to the editing frequencies in the same position.

**Table 4.2:** Summary of the editing events significantly correlating with drug sensitivity in all datasets

| Dataset | Drug | Editing event | Pearson R | Adjusted p-value | Spearman Rho | Adjusted p-value |
|---|---|---|---|---|---|---|
| GDSC1 | Bexarotene | chr12\|53572046\|transcript\|CSAD | | | -1 | 1.42407E-35 |
| | Fulvestrant | chr7\|65620143\|-\|- | | | 0.970719794 | 0.028106805 |
| | | chr16\|28969225\|transcript\|NFATC2IP | | | 0.912624507 | 0.028106805 |
| | | chr8\|6500879\|transcript,exon,3UTR&transcript\|MCPH1&MCPH1-AS1 | | | -0.983333333 | 0.009898804 |
| | | chr17\|29861303\|transcript,exon,3UTR\|RAB11FIP4 | | | -0.963636364 | 0.024952322 |
| | | chr2\|203163866\|transcript\|NOP58 | | | -1 | <2.2E-16 |
| | Pazopanib | chr1\|40214148\|transcript\|PPIE | | | -1 | <2.2E-16 |
| | Ruxolitinib | chr8\|94731256\|transcript\|FAM92A | | | 1 | <2.2E-16 |
| | Temsirolimus | chr17\|73094640\|transcript\|SLC16A5 | | | -0.981725568 | 0.002408491 |
| | Vorinostat | chr1\|179044105\|transcript,exon,3UTR\|FAM20B | -0.912147588 | 0.022641342 | -0.918699283 | 0.013903789 |
| GDSC2 | Dasatinib | chr19\|13883862\|transcript,exon,3UTR\|MRI1 | -0.932340477 | 0.036050666 | | |
| | Fulvestrant | chr12\|27188059\|-\|- | | | -0.915109747 | 0.04699305 |
| | | chr19\|59076412\|transcript&transcript\|MZF1-AS1&MZF1 | -0.90932461 | 0.027394211 | | |
| | Tamoxifen | chr16\|81063389\|transcript,exon,3UTR\|CENPN | | | -0.958681324 | 0.002518587 |
| CTRP | Cabozatinib | chr1\|93621294\|transcript\|TMED5 | | | 0.943685372 | 0.038313846 |
| | | chr12\|124007386\|transcript\|RILPL1 | | | 0.879198985 | 0.038313846 |
| | Dasatinib | chr1\|173827747\|-\|- | -0.879287525 | 0.03516473 | | |
| | Docetaxel | chr7\|99712522\|transcript\|TAF6 | | | 1 | <2.2E-16 |
| | | chr6\|147886688\|transcript,exon,3UTR\|SAMD5 | | | -1 | <2.2E-16 |
| | | chr9\|37775695\|transcript\|TRMT10B | | | 1 | |
| | Doxorubicin | chr16\|31134397\|transcript\|KAT8 | | | -0.939178605 | 0.002231907 |
| | Fulvestrant | chr9\|37778400\|transcript,exon,3UTR\|TRMT10B | | | 0.955889667 | 0.020771882 |
| | | chr7\|74296346\|-\|- | | | 1 | <2.2E-16 |
| | Lapatinib | chr1\|53290057\|transcript,exon,3UTR\|ZYG11B | -0.832734196 | 0.008673247 | | |
| | Mitomycin | chr19\|14593693\|transcript,CDS,exon\|GIPC1 | | | -0.845266003 | 0.014902625 |
| | | chr17\|62479428\|transcript\|POLG2 | -0.893229524 | 0.033558929 | | |
| | Nilotinib | chr10\|43014136\|exon,transcript\|ZNF37BP | | | 0.957770229 | 0.010381165 |
| | | chr19\|14680184\|transcript\|NDUFB7 | -0.914131832 | 0.019747731 | -0.90665588 | 0.016677095 |
| | Pazopanib | chr16\|28969640\|transcript\|NFATC2IP | -0.867079752 | 0.007921339 | | |
| CCLE | Paclitaxel | chr1\|53289892\|transcript,exon,3UTR\|ZYG11B | | | -0.924347504 | 0.016066805 |
| | | chr7\|86816767\|transcript\|DMTF1 | | | -0.892350139 | 0.046651777 |
| | | chr11\|125545903\|transcript,3UTR,exon&transcript\|CHEK1&ACRV1 | | | -0.923925619 | 0.02261362 |
| | | chr14\|20919249\|transcript\|OSGEP | | | -1 | 1.44693E-35 |
| | | chr19\|23441380\|exon,transcript\|IPO5P1 | | | -0.991031209 | 0.02261362 |
| | | chr19\|48956734\|transcript,exon,3UTR\|GRWD1 | | | 0.952054692 | 0.006291133 |
| | Crizotinib | chr20\|5981051\|transcript\|MCM8-AS1 | | | -0.991031209 | 0.037327772 |
| | | chr15\|75645225\|transcript,exon,3UTR\|NEIL1 | | | -0.983333333 | 0.013196643 |

### 4.2.2 The RANKSUM approach

In order to identify the same Drug-Editing event combinations in more than one dataset we decided to employ a different Rank-based approach, schematically demonstrated in Figure 5. Firstly editing frequencies at the 10537 selected sites and drug sensitivity measures of the 29 chemotherapeutics were submitted to a Spearman correlation analysis. Editing events whose frequencies were significantly correlating with chemosensitivity were selected (nominal $p<0.05$). For each drug, the list of the selected editing events from each dataset were intersected, and editing events common in two or more datasets were chosen. Editing events were then ranked independently in each of the reported datasets based on the strength of the correlation (Spearman Rho) to the sensitivity of the respective drugs in an ascending order. A value was assigned to each editing event based on its ranking. The rank values from each dataset were summed and a RankSum value was assigned to the editing event. Using the RankSum value the editing events were sorted from smallest to largest. The list of common editing events for each drugs were analyzed and it was observed that in some of the cases, editing frequencies at the same site would show opposite correlation direction in different datasets. To ensure the consistency of the observed associations, only drugs that showed the same correlation direction with the editing events throughout the datasets were considered for further analysis.

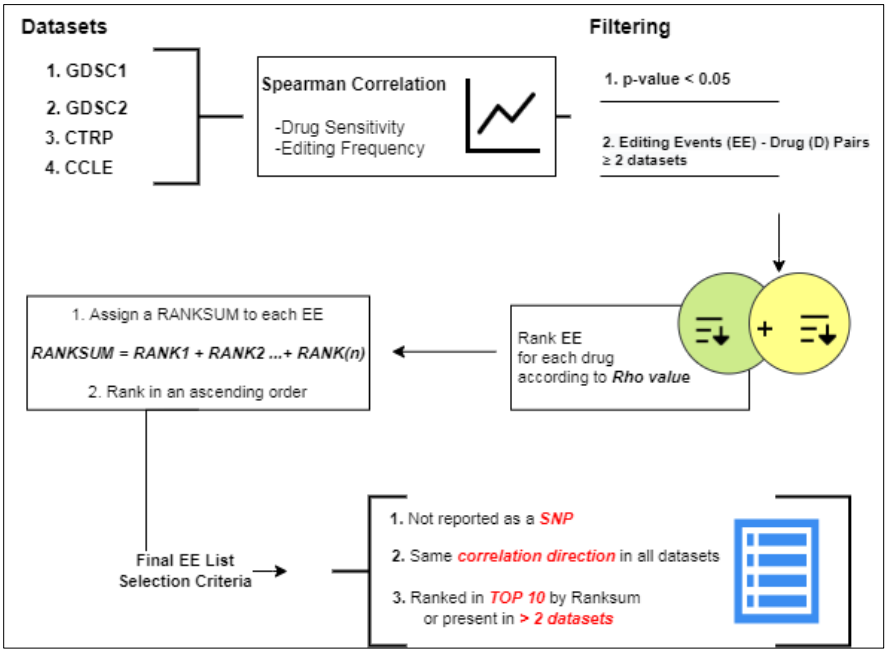



**Figure 4.5**. RANKSUM analysis workflow

**Table 4.3.** Summary of RANKSUM analysis results.

| | Nr of significantly correlating editing events | | | | Nr of common editing events/Nr of editing events displaying opposite correlations | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Drug | GDSC1 | GDSC2 | CTRP | CCLE | GDSC1&GDSC2 | GDSC1&CTRP | GDSC2&CTRP | CTRP&CCLE | GDSC2&CCLE | GDSC1&CCLE | CTRP&GDSC2&CCLE | GDSC1&GDSC2&CTRP | CCLE&CTRP&GDSC1 |
| 5-FU | 473 | 559 | 474 | NA | 10/10 | 8/6 | 21/15 | | | | | | |
| Axitinib | 461 | NA | 498 | NA | | 17/0 | | | | | | | |
| Bexarotene | 479 | NA | 451 | NA | | 11/9 | | | | | | | |
| Bleomycin | 421 | NA | 588 | NA | | 41/40 | | | | | | | |
| Bleomycin (10uM) | 553 | NA | 588 | NA | | 13/9 | | | | | | | |
| Bleomycin (50uM) | 631 | NA | 588 | NA | | 20/17 | | | | | | | |
| Bortezomib | 0 | 431 | 447 | NA | | | 11/7 | | | | | | |
| Cabozatinib | 503 | NA | 867 | NA | | 71/0 | | | | | | | |
| Cisplatin | 503 | 638 | 714 | NA | 23/0 | 25/19 | 23/17 | | | | | | |
| Crizotinib | 0 | 425 | 478 | 875 | | | 19/2 | 56/0 | 47/0 | | 4/0 | | |
| Dasatinib | 0 | 758 | 501 | NA | | | 61/2 | | | | | | |
| Docetaxel | 478 | 490 | 372 | NA | 61/0 | 13/1 | 27/2 | | | | | 2/0 | |
| Doxorubicin | 522 | NA | 547 | NA | | 18/8 | | | | | | | |
| Erlotinib | 0 | 492 | 489 | 479 | | | 36/1 | 52/0 | 23/0 | | 4/0 | | |
| Etoposide | 513 | NA | 530 | NA | | 21/8 | | | | | | | |
| Fulvestrant | 699 | 450 | 536 | NA | 126/0 | 54/0 | 28/0 | | | | | 13/0 | |
| Gemcitabine | 453 | 603 | 472 | NA | 98/0 | 13/6 | 44/0 | | | | | 4/0 | |
| Lapatinib | 0 | 473 | 569 | 415 | | | 35/0 | 94/0 | 14/0 | | 4/0 | | |
| Methotrexate | 646 | NA | 823 | NA | | 112/0 | | | | | | | |
| Mitomycin | 443 | NA | 507 | NA | | 18/5 | | | | | | | |
| Nilotinib | 835 | 534 | 556 | 567 | 110/0 | 94/0 | 42/0 | 52/3 | 9/9 | 86/0 | | 19/0 | 13/0 |
| Olaparib | 448 | 554 | 454 | NA | 209/0 | 61/0 | 54/1 | | | | | 30/0 | |
| Paclitaxel | 0 | 472 | 540 | 979 | | | 10/9 | 116/0 | 22/4 | | 1/0 | | |
| Pazopanib | 478 | NA | 502 | NA | | 17/6 | | | | | | | |
| Rapamycin | 0 | 816 | 490 | NA | | | 28/7 | | | | | | |
| Ruxolitinib | 399 | 529 | 546 | NA | 20/0 | 15/11 | 32/20 | | | | | | |
| SN-38 | 567 | NA | 525 | NA | | 23/6 | | | | | | | |
| Sorafenib | 0 | 604 | 818 | 714 | | | 69/1 | 80/0 | 152/0 | | 26/0 | | |
| Tamoxifen | 496 | 510 | 547 | NA | 96/0 | 43/1 | 27/1 | | | | | 7/1 | |
| Temsirolimus | 572 | NA | 532 | NA | | 54/1 | | | | | | | |
| Vorinostat | 497 | 453 | 525 | NA | 42/0 | 27/1 | 42/0 | | | | | 3/0 | |

About 60% (17/29) of the drugs were shown to be consistently correlated with editing event frequency in different datasets in the same fashion (**Table 4.3**). The cells highlighted in green in Table3 show the lists of common editing events which were selected for further consideration. Priority was given to editing events present in the intersection of 3 datasets. For drugs in which the number of editing events present in 3 datasets was less than 15, the top 10 editing events ranked by RankSum in the intersection of two datasets were considered. The same strategy was followed for drugs tested in two datasets only. All selected editing events were manually checked and those reported in dbSNP were discarded. After applying these criteria we obtained a pool of 305 unique editing events significantly correlated with chemosensitivity of 17 approved chemotherapeutics in GC cell lines in more than one dataset (**Appendix C**).

### 4.2.3 Editing events consistently associate with drug sensitivity in different datasets

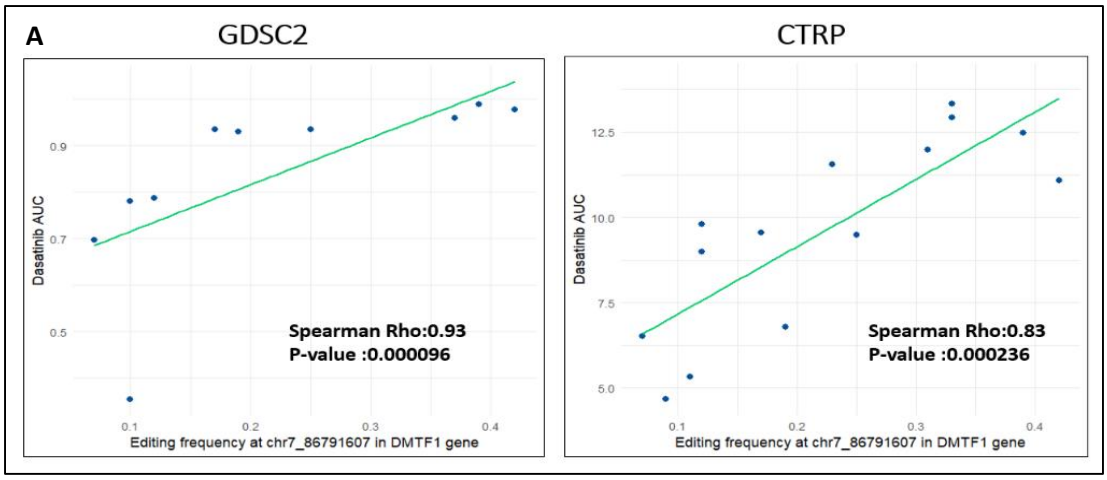
To further on investigate the implication of these editing events in chemosensitivity of GC cell lines, we decided to focus on events that show the most significant correlations with drug sensitivity (nominal p<0.0001).

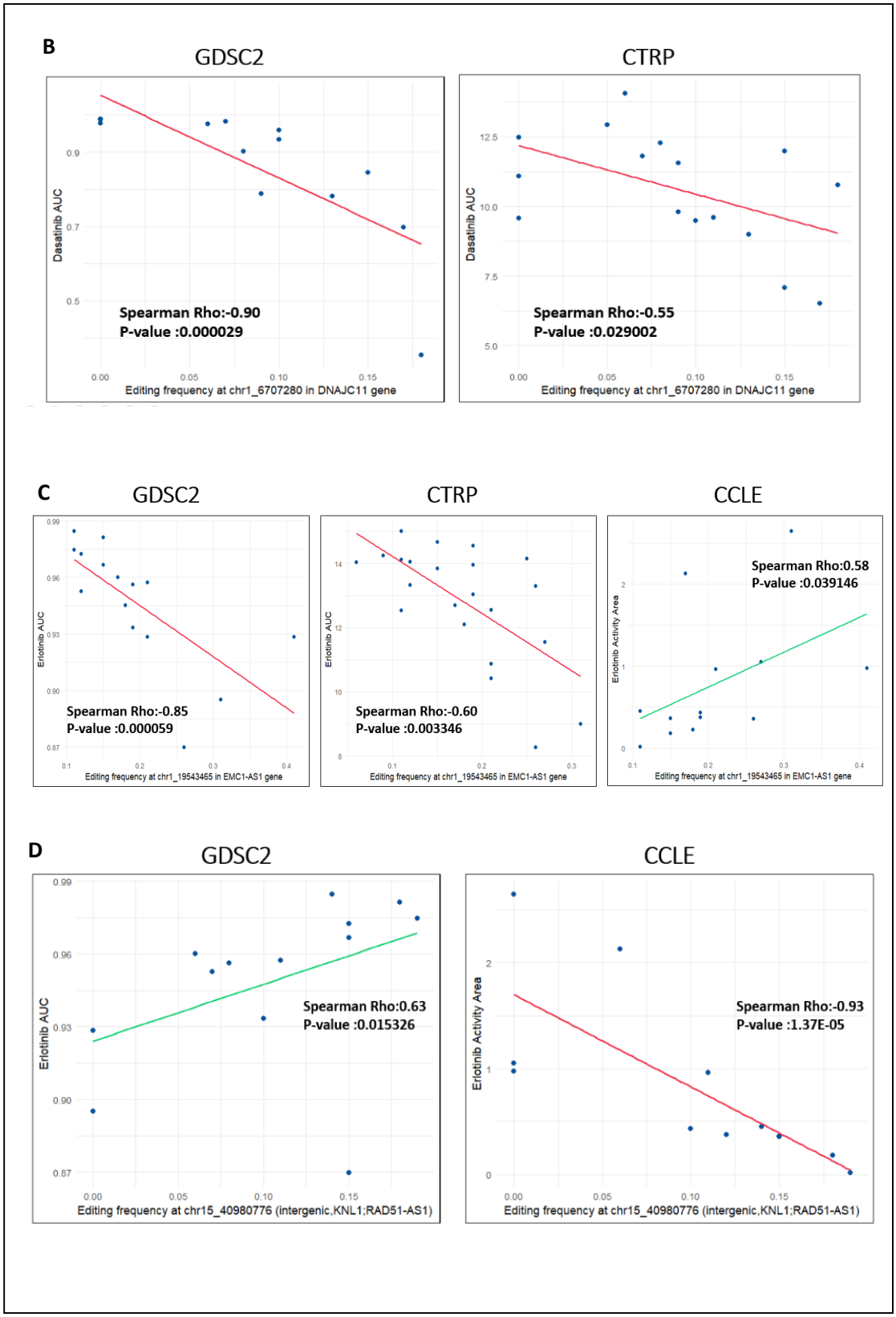
**Table 4.4:** Most significantly correlated editing events from the RANKSUM analysis

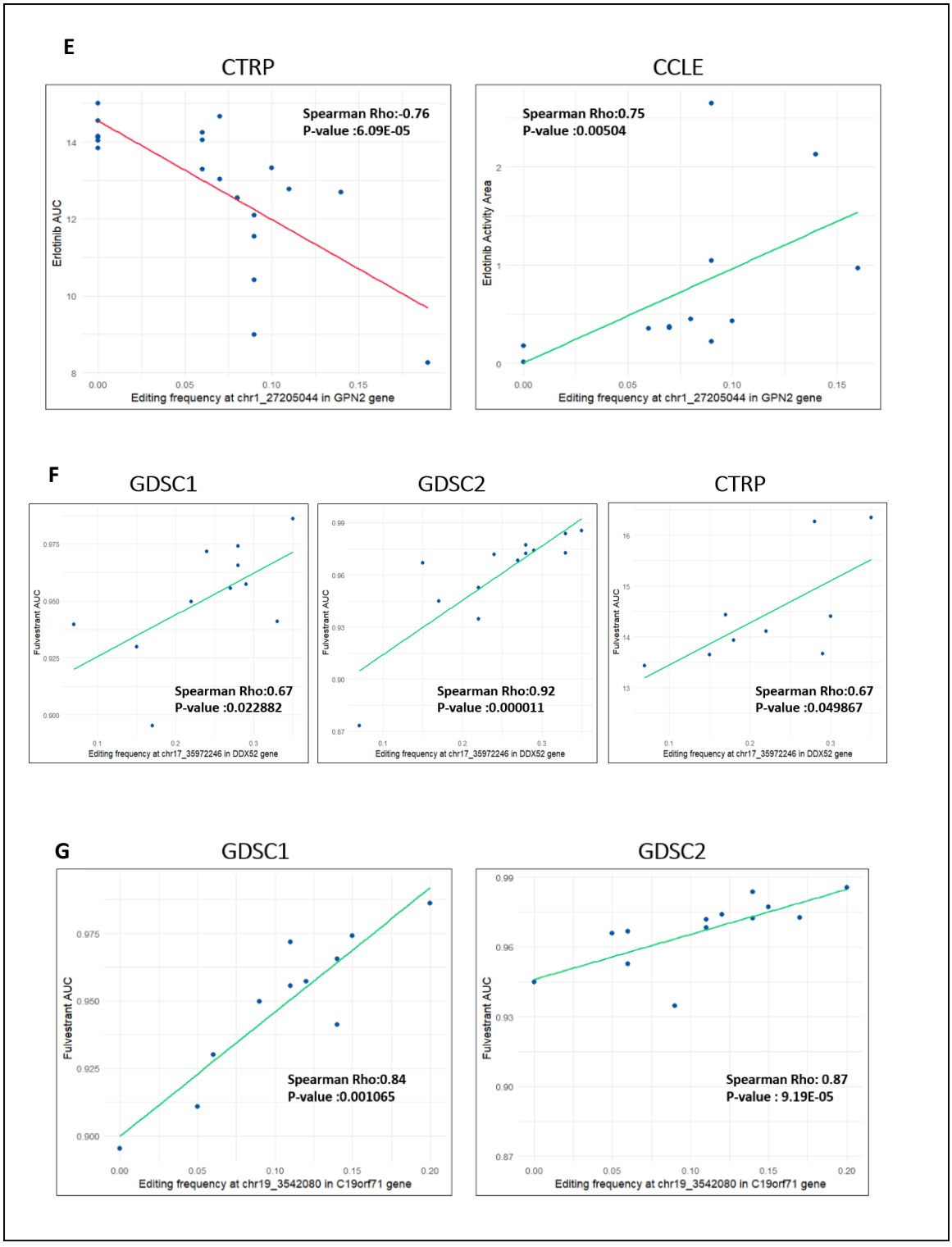
| Drug | Gene | Genomic coordinates | Region | GDSC1 | GDSC2 | CTRP | CCLE |
|---|---|---|---|---|---|---|---|
| Dasatanib | DMTF1 | ch7:86791607 | intronic | | 0.93*** | 0.83** | |
| Dasatanib | DNAJC11 | chr1:6707280 | intronic | | -0.90*** | -0.55 | |
| Erlotinib | EMC1-AS1 | chr1:19543465 | ncRNA_intronic | | -0.85*** | -0.6 | 0.58 |
| Erlotinib | KNL1<br>RAD51-AS1 | chr15:40980776 | intergenic | | 0.63 | | -0.93*** |
| Erlotinib | GPN2 | chr1:27205044 | 3UTR | | | -0.76 *** | 0.75 |
| Fulvestrant | DDX52 | chr17:35972246 | 3UTR | 0.67 | 0.92*** | 0.67 | |
| Fulvestrant | C19orf71 | chr19:3542080 | intronic | 0.84* | 0.87*** | | |
| Fulvestrant | MED21<br>C12orf71 | chr12:27188059 | intergenic | -0.81* | -0.92*** | | |
| Tamoxifen | CENPN | chr16:81063389 | 3UTR | -0.79* | -0.96*** | | |

In Table 4.4 we can observe that the most significant correlations between editing events and drug sensitivity measures are obtained from GDSC2 drug sensitivity data (7 out of 9). Editing frequencies at chr12:27188059 position in intergenic region between MED21 and

C12orf71 and at chr16:81063389 position in 3'UTR of CENPN gene were also identified in our initial correlation analysis (**Table 4.2**). Despite the fact that the most significantly correlated editing events are at unique position, they are correlated to the drug sensitivity of four drugs in total. The majority of these editing events are strongly correlated (Rho>0.7) with the sensitivity of the respective drugs. Unlike other datasets, Activity Area, the inverse measure of AUC, is used to measure sensitivity in CCLE. Therefore the opposite direction of correlation between CCLE and the other datasets indicates the same association between editing frequency and drug response. Correlation graphs for each of the editing sites in Table4 are shown in Fig6 in the same order (A-I). Based on our analysis, overall high editing levels associate with a chemosensitive phenotype in GC cell lines, although the pattern changes according to the drug. In our initial analysis and the RANKSUM approach, 62% and 56% of the identified editing events respectively, positively correlate with increased sensitivity towards chemotherapeutics (**Table 4.2** and **Table 4.4**).




B
GDSC2
CTRP
Dasatinib AUC
Spearman Rho:-0.90
P-value :0.000029
Spearman Rho:-0.55
P-value :0.029002
Editing frequency at chr1_6707280 in DNAJC11 gene
C
GDSC2
CTRP
CCLE
Erlotinib AUC
Erlotinib Activity Area
Spearman Rho:-0.85
P-value :0.000059
Spearman Rho:-0.60
P-value :0.003346
Spearman Rho:0.58
P-value :0.039146
Editing frequency at chr1_19543465 in EMC1-AS1 gene
D
GDSC2
CCLE
Spearman Rho:0.63
P-value :0.015326
Spearman Rho:-0.93
P-value :1.37E-05
Editing frequency at chr15_40980776 (intergenic,KNL1;RAD51-AS1)


E
CTRP
Spearman Rho:-0.76
P-value :6.09E-05
Erlotinib AUC
Editing frequency at chr1_27205044 in GPN2 gene
CCLE
Spearman Rho:0.75
P-value :0.00504
Erlotinib Activity Area
Editing frequency at chr1_27205044 in GPN2 gene
F
GDSC1
Spearman Rho:0.67
P-value :0.022882
Fulvestrant AUC
Editing frequency at chr17_35972246 in DDX52 gene
GDSC2
Spearman Rho:0.92
P-value :0.000011
Fulvestrant AUC
Editing frequency at chr17_35972246 in DDX52 gene
CTRP
Spearman Rho:0.67
P-value :0.049867
Fulvestrant AUC
Editing frequency at chr17_35972246 in DDX52 gene
G
GDSC1
Spearman Rho:0.84
P-value :0.001065
Fulvestrant AUC
Editing frequency at chr19_3542080 in C19orf71 gene
GDSC2
Spearman Rho: 0.87
P-value : 9.19E-05
Fulvestrant AUC
Editing frequency at chr19_3542080 in C19orf71 gene

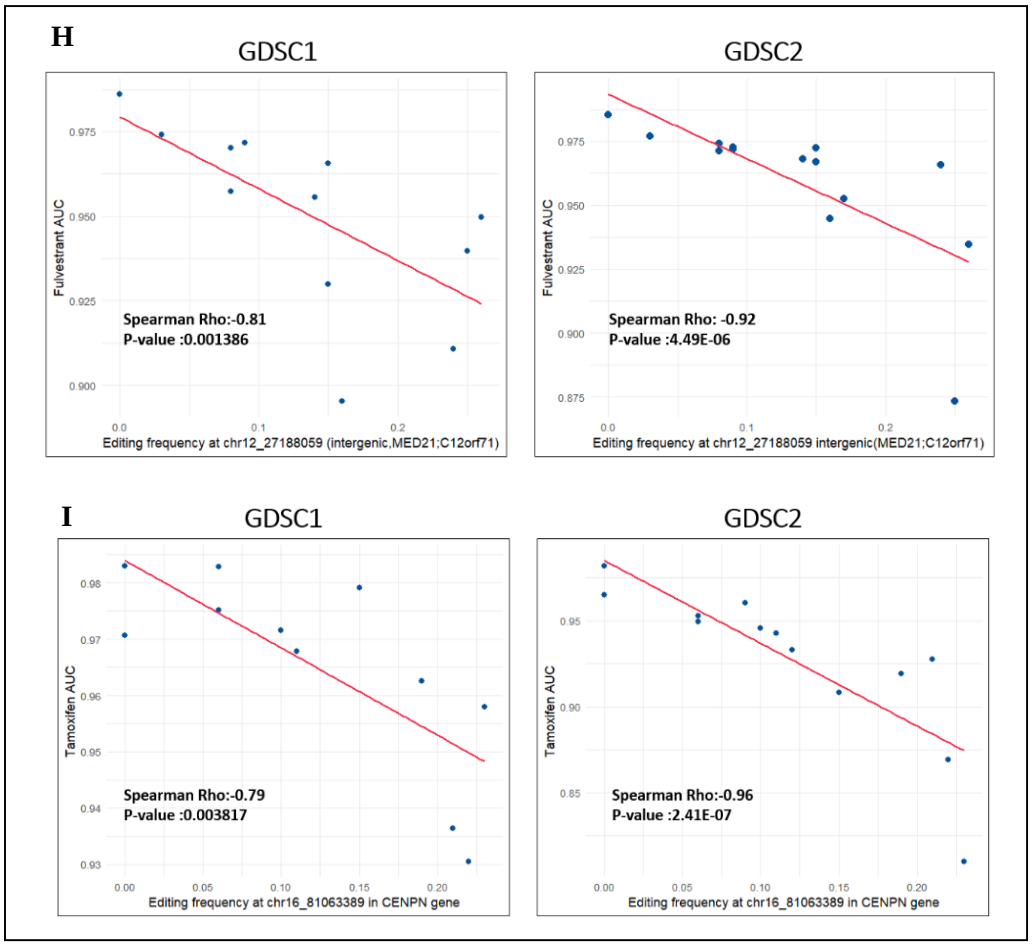



**Figure 4.6**. Scatter plots for editing event correlations listed in Table 4 in the same order (A-I)

To rule out the possibility that the observed correlation occur due to the expression levels of the genes they are located in, Pearson correlation of drug sensitivity and gene expression was performed for the 12 genes and 5 drugs in Table 4.4. Three of the gene transcripts significantly correlate with the drug sensitivity values ($p<0.05$).

**Table 4.5**. Summarized results of drug sensitivity and gene expression correlation. Significant correlations highlighted in red.

| Gene | Genomic coordinates | Location | Drug | GDSC1 *r* | GDSC1 p-value | GDSC2 *r* | GDSC2 p-value | CTRP *r* | CTRP p-value | CCLE *r* | CCLE p-value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DMTF1 | Ch7:86791607 | intronic | Dasatinib | NA | NA | -0.57 | 0.03 | -0.13 | 0.56 | NA | NA |
| DNAJC11 | Chr1:6707280 | intronic | Dasatinib | NA | NA | -0.17 | 0.55 | 0.23 | 0.30 | NA | NA |
| RP1-43E13.2 (EMC1-AS1) | Chr1:19543465 | ncRNA_intronic | Erlotinib | NA | NA | -0.04 | 0.90 | 0.00 | 0.99 | -0.20 | 0.52 |
| RAD51-AS1 | Chr15:40980776 | intergenic | Erlotinib | NA | NA | -0.19 | 0.51 | NA | NA | 0.03 | 0.92 |
| CASC5 (KNL1) | Chr15:40980776 | intergenic | Erlotinib | NA | NA | 0.07 | 0.80 | NA | NA | -0.45 | 0.12 |
| GPN2 | Chr1:27205044 | 3UTR | Erlotinib | NA | NA | NA | NA | 0.03 | 0.89 | -0.05 | 0.88 |
| DDX52 | Chr17:35972246 | 3UTR | Fulvestrant | -0.47 | 0.11 | -0.14 | 0.62 | 0.35 | 0.29 | NA | NA |
| C19orf71 | Chr19:3542080 | intronic | Fulvestrant | 0.56 | 0.05 | 0.49 | 0.07 | NA | NA | NA | NA |
| C12orf71 | Chr12:27188059 | intergenic | Fulvestrant | 0.22 | 0.46 | 0.14 | 0.62 | NA | NA | NA | NA |
| MED21 | Chr12:27188059 | intergenic | Fulvestrant | -0.17 | 0.59 | -0.26 | 0.36 | NA | NA | NA | NA |
| CENPN | Chr16:81063389 | 3UTR | Tamoxifen | -0.60 | 0.04 | -0.49 | 0.06 | NA | NA | NA | NA |
| C11orf80 | Chr11:66523904 | CDS –exonic | Lapatinib | NA | NA | -0.05 | 0.86 | 0.07 | 0.74 | -0.10 | 0.74 |

DMTF1 transcript levels correlate with Dasatinib sensitivity of GC cell lines in GDSC2 dataset, in an opposite direction to the editing frequencies. C19orf71 and CENPN transcript levels correlate with Fulvestrant and Tamoxifen sensitivity respectively in GDSC1. The observed correlations between transcript levels and drug sensitivity are weaker and less significant than the observed correlations with editing frequencies at editing sites located in these transcripts.

### 4.2.4 Correlation of ADAR enzymes expression to editing frequencies

ADAR enzymes, ADAR1 and ADAR2, are the main regulators of A-to-I editing events in vertebrates, so we decided to investigate if there is a correlation between expression levels of their coding genes and editing levels in GC cell lines. Out of the 305 editing events, 4 of them significantly correlate with ADAR gene transcript levels, which codes for ADAR1 enzyme. Editing levels at all four editing sites, show positive correlation to ADAR expression levels, with rho values ranging from 0.73 to 0.77 (Fig 4.7). It can also be noticed that these 4 distinct positions are edited in almost half of the transcripts, which is a relatively higher editing level as compared to average editing level in the GC cell lines (~12%). The majority of the other editing events show no correlation with ADAR levels (148 out of 305).Correlation analysis was performed with ADARB1 gene expression levels, which codes for ADAR2 protein, however none of the editing events significantly correlated (**Appendix D).**

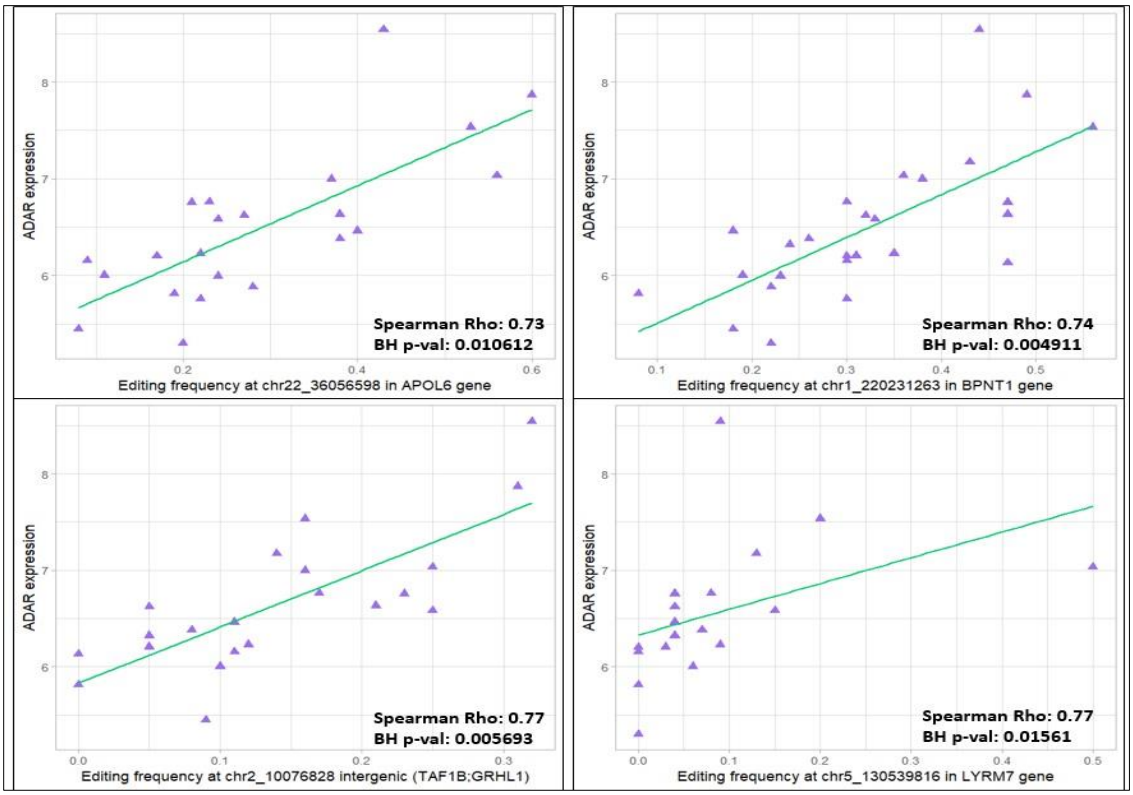



**Figure 4.7**: Correlation graphs of editing events significantly correlating with ADAR1 expression.

### 4.2.5 Functional implication of editing events on gene expression and protein sequence

Moving forward, we decided to investigate if the editing events have an impact on transcript levels of the genes they are located in. Editing frequencies at the 305 editing sites and the respective gene transcript levels were Spearman correlated. After multi-test correction (BH p-value adjustment), none of the transcripts showed a significant correlation to editing events. Based on the nominal p-value, a few transcripts correlate with the editing frequencies at editing sites mainly located in the 3'UTR or intronic regions (**Appendix E**). SOGA1 transcript levels show the highest significant correlation with the editing event at position 35411013 in the 3'UTR. SOGA1 transcript levels show moderate positive correlation with editing frequencies (Spearman Rho: 0.66, **Fig 4.8**). Using TargetScan software we were able to identify that the editing site is located in the seed

region of a poorly conserved target site of hsa-miR-9-5p, in a less prevalent transcript of SOGA1 (ENST00000456801.2) (**Fig 4.9).**

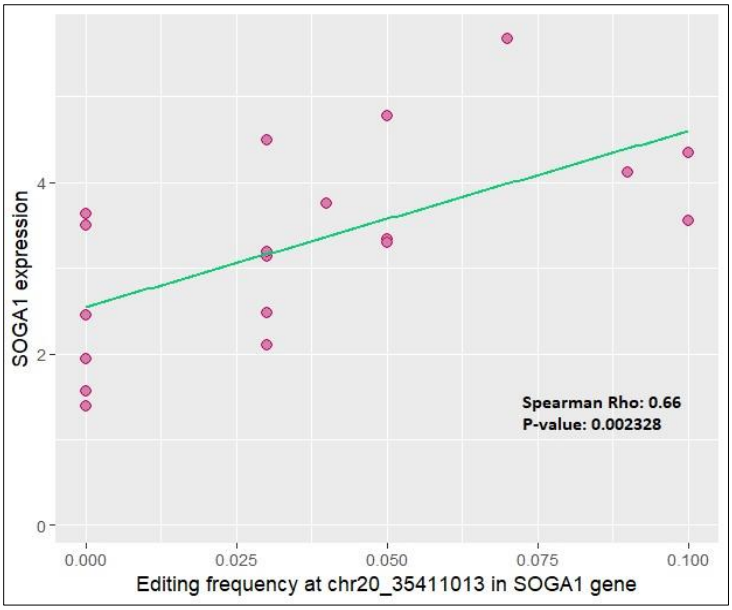



**Figure 4.8**: Correlation graph of editing frequencies in SOGA1 UTR and its transcript levels

Using miRNA expression data from CCLE, the expression of hsa-miR-9 was confirmed in GC cell line, however a correlation to SOGA1 gene expression was not detected (**Appendix F**).

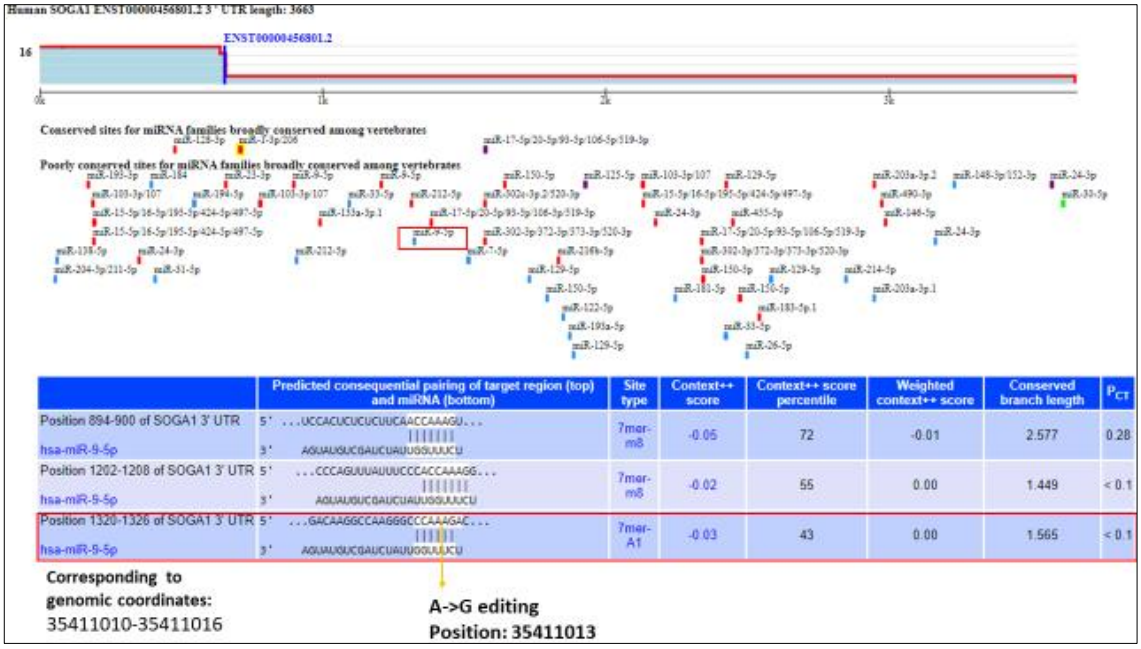



**Figure 4.9:** Schematic depiction of hsa-miR-9-5p target site in SOGA1 3'UTR. Target site overlapping the editing event location is highlighted in a red box. Yellow arrow indicates editing site.

We also decided to consider an editing event located in the coding region, due to it being a rarely encountered event in our analysis. The editing event at 66523904 position in the coding sequence of C11orf80 (TOP6BL) gene showed a moderate positive correlation with Lapatinib drug sensitivity in three different datasets (**Fig 4.10**). The A-to-G editing occurs in type 2 DNA topoisomerase 6 subunit B-like isoform a, on the 397$^{th}$ nucleotide in the CDS, causing a nonsynonymous change in the 133th amino acid from Serine to Glycine (p.S133G). To observe the effect this amino acid change could potentially have on the protein function, we used PredictSNP web-tool, which reports a prediction based on 6 different computational tools (**Fig 4.11**). Five of the tools integrated in PredictSNP predict a Neutral effect of the mutation on the protein with an accuracy ranging from 55% to 71%.

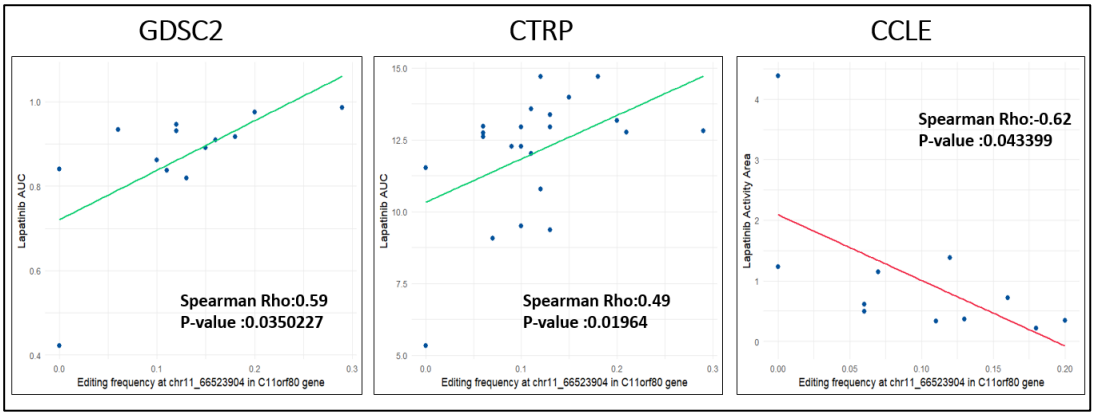



**Figure 4.10**: Editing events in CDS of C11orf80 gene, correlate with Lapatinib sensitivity in GC cell lines.

However, based on SIFT prediction tool, the mutation has a deleterious effect on the protein, with the highest prediction accuracy at 79%. The 3D conformation of this isoform was unavailable in protein data bank (PDB), thus we were not able to identify the effect this mutation could have on protein structure and stability.

| | % Expected Accuracy | | | *Neutral* | *Deleterious* | |
|---|---|---|---|---|---|---|
| **Mutation** | **PredictSNP** | **PhD-SNP** | **PolyPhen-1** | **PolyPhen-2** | **SIFT** | **SNAP** |
| **S133G** | 71 % | 55 % | 67 % | 71 % | 79 % | 61 % |

**Figure 4.11**: Functional consequences of Serine to Glycine substitution in position 133 of TOP6BL protein caused by A-to-I editing event in CDS as predicted by PredictSNP

## 4.3 A-to-I Editing Events as Prognostic Biomarkers in GC Patients

### 4.3.1 Selection of editing events associated to prognosis

To study the prognostic implication of A-to-I editing events in GC, RNA editing frequencies in patient tissue were obtained from two different datasets summarized in Table 4.6. The datasets belong to two completely different and independent studies, leading to discrepancies in the population pool and the editing identification approach.

**Table 4.6:** Summary table for cohort characteristics.

| | **Singapore Cohort** | **TCGA Cohort (STAD)** |
|---|---|---|
| **Patients Number** | 104 | 285 |
| **Stage** | III-IV | II-IV |
| **Survival Data** | 54 | 224 |
| **Editing Events Reported** | 780 | 26,389 |
| **Clinical Parameters** | Gender , Treatment , Lauren`s Histology, Overall Response | Gender , Stage, Primary Response |
| **Study** | Singapore Gastric Cancer Consortium (SGCC) | The Cancer Genome Atlas – Stomach Adenocarcinoma |

Nevertheless due to the scarce number of datasets providing editing frequency data we decided to proceed with these datasets, which will be referred to further on as the Singapore and TCGA cohort.

Univariate Cox Regression Analysis was performed on all the reported editing events in both datasets to investigate their association with overall survival (OS). In the Singapore Cohort 59 out of 780 editing events significantly associate to prognosis ($p<0.05$). Based on the Hazard Ratio (HR), majority of the editing events (52/59) were good prognostic factors ($HR<1$) and only a few (7/59) were classified as bad prognostic factors. In the TCGA Cohort 1311 out of 26389 editing events were significantly associated to OS. Around 65% of the editing events were identified as bad prognostic factors (854/1311) and 35% as good prognostic factors (457/1311) (**Appendix F**). We decided to screen the editing events with

a significant prognostic effect in both datasets for common occurrences. Two editing events at the exact same position and 15 common edited genes associate with OS in both datasets (**Table 4.7**).

**Table 4.7:** List of editing events in genes associated to prognosis in both cohorts.

| Gene | Singapore Cohort | HR | p-value | TCGA Cohort | HR | p-value |
|---|---|---|---|---|---|---|
| ACOX1 | chr17\|73940167\|ACOX1 | 438.5848 | 0.013012 | chr17\|73973261\|ACOX1 | 3.62E-05 | 0.023298 |
| | | | | chr17\|73939058\|ACOX1 | 4.76465E+18 | 9.06E-05 |
| | | | | chr17\|73940953\|ACOX1 | 0.000127677 | 0.04985 |
| | | | | chr17\|73941003\|ACOX1 | 3148.322635 | 0.001526 |
| ARSD | chrX\|2824979\|ARSD | 0.013484 | 0.018043 | chrX\|2824214\|ARSD | 2351.297795 | 0.007669 |
| | | | | chrX\|2825003\|ARSD | 1047324.975 | 0.002394 |
| BPNT1 | chr1\|220231276\|BPNT1 | 0.016067 | 0.034386 | chr1\|220231492\|BPNT1 | 1504954.219 | 0.049571 |
| CCDC125 | chr5\|68578277\|CCDC125 | 0.072637 | 0.027733 | chr5\|68578171\|CCDC125 | 0.056446601 | 0.048719 |
| | | | | chr5\|68578306\|CCDC125 | 0.074910648 | 0.030607 |
| | | | | chr5\|68578307\|CCDC125 | 0.016951892 | 0.027554 |
| CLPX | | | | chr15\|65442476\|CLPX | 0.085268551 | 0.020301 |
| | | | | chr15\|65441807\|CLPX | 0.004364127 | 0.004061 |
| | **chr15\|65442098\|CLPX** | **0.010973** | **0.006228** | **chr15\|65442098\|CLPX** | **0.010290843** | **0.004961** |
| | | | | chr15\|65442587\|CLPX | 0.002127272 | 0.014887 |
| | | | | chr15\|65442678\|CLPX | 0.058015087 | 0.040914 |
| CPT1A | chr11\|68523337\|CPT1A | 0.101026 | 0.042808 | chr11\|68523425\|CPT1A | 1.43E-08 | 0.000824 |
| | | | | chr11\|68523497\|CPT1A | 0.027168705 | 0.027051 |
| CTSB | chr8\|11701429\|CTSB | 0.009616 | 0.008163 | chr8\|11700339\|CTSB | 6.70E+22 | 0.00746 |
| | | | | chr8\|11700418\|CTSB | 0.004344069 | 0.015043 |
| | | | | chr8\|11700426\|CTSB | 8.68E-06 | 0.012019 |
| | | | | chr8\|11700595\|CTSB | 0.001713497 | 0.029787 |
| | | | | chr8\|11701473\|CTSB | 0.000243609 | 0.047094 |
| | | | | chr8\|11712134\|CTSB | 247.1239486 | 0.042162 |
| | | | | chr8\|11712983\|CTSB | 65.79382818 | 0.016144 |
| CTSS | chr1\|150704659\|CTSS | 0.004524 | 0.040567 | chr1\|150703647\|CTSS | 20.43326228 | 0.041669 |
| | | | | chr1\|150703715\|CTSS | 44.57324534 | 0.016171 |
| | | | | chr1\|150703716\|CTSS | 73.13131071 | 0.004441 |
| | | | | chr1\|150705145\|CTSS | 1.17E-11 | 0.046021 |
| | | | | chr1\|150705264\|CTSS | 1.65E-19 | 0.007254 |
| | | | | chr1\|150705289\|CTSS | 1.40447E+15 | 0.013699 |
| DCAF16 | chr4\|17802824\|DCAF16 | 0.000496 | 0.000609 | chr4\|17802849\|DCAF16 | 208.3420623 | 0.024636 |
| | chr4\|17802889\|DCAF16 | 0.00999 | 0.025716 | chr4\|17802922\|DCAF16 | 548.2884945 | 0.044514 |
| | chr4\|17803559\|DCAF16 | 156.3706 | 0.021154 | chr4\|17802931\|DCAF16 | 0.046348324 | 0.016671 |
| | | | | chr4\|17803568\|DCAF16 | 2170.748015 | 0.018355 |
| | | | | chr4\|17803578\|DCAF16 | 7435.584653 | 0.001334 |
| | | | | chr4\|17804654\|DCAF16 | 5.92E-08 | 0.031141 |
| | | | | chr4\|17804715\|DCAF16 | 346190.4543 | 0.002651 |
| GNPNAT1 | chr14\|53242651\|GNPNAT1 | 0.077923 | 0.0462 | chr14\|53242425\|GNPNAT1 | 0.029063153 | 0.042972 |
| NDUFS1 | chr2\|206986978\|NDUFS1 | 0.080156 | 0.02678 | chr2\|206987127\|NDUFS1 | 0.037507775 | 0.038198 |
| | chr2\|206987231\|NDUFS1 | 0.047011 | 0.015055 | chr2\|206987632\|NDUFS1 | 0.015604819 | 0.016522 |
| | chr2\|206987634\|NDUFS1 | 0.004202 | 0.012648 | | | |
| NUP155 | chr5\|37291446\|NUP155 | 0.034621 | 0.025572 | chr5\|37291526\|NUP155 | 1.79E-07 | 0.023455 |
| | chr5\|37291548\|NUP155 | 0.03562 | 0.02086 | | | |
| NUP43 | chr6\|150045858\|NUP43 | 0.071004 | 0.039099 | chr6\|150045989\|NUP43 | 0.00774927 | 0.022514 |
| | | | | chr6\|150046473\|NUP43 | 4059.742618 | 0.000126 |
| | | | | chr6\|150046557\|NUP43 | 0.000145261 | 0.007941 |
| | | | | chr6\|150047455\|NUP43 | 278.3879819 | 0.016841 |
| TEP1 | chr14\|20834426\|TEP1 | 0.018875 | 0.041289 | chr14\|20834521\|TEP1 | 126.075719 | 0.038785 |
| | | | | chr14\|20835178\|TEP1 | 0.077024671 | 0.028658 |
| | | | | chr14\|20835357\|TEP1 | 0.059616572 | 0.02852 |
| | | | | chr14\|20835465\|TEP1 | 0.009145759 | 0.019014 |
| | | | | chr14\|20835942\|TEP1 | 9.33E-05 | 0.039708 |
| ZDHHC20 | | | | chr13\|21949149\|ZDHHC20 | 4.56E-06 | 0.006562 |
| | **chr13\|21948495\|ZDHHC20** | **0.059345** | **0.034022** | **chr13\|21948495\|ZDHHC20** | **0.145004518** | **0.043581** |
| | | | | chr13\|21948496\|ZDHHC20 | 4.68E-07 | 0.001919 |
| | | | | chr13\|21948624\|ZDHHC20 | 0.097840209 | 0.021562 |
| | | | | chr13\|21948625\|ZDHHC20 | 0.155840085 | 0.026162 |
| | | | | chr13\|21948646\|ZDHHC20 | 0.013850254 | 0.03581 |
| | | | | chr13\|21949038\|ZDHHC20 | 0.068303555 | 0.014658 |
| | | | | chr13\|21949047\|ZDHHC20 | 0.111847283 | 0.046938 |
| | | | | chr13\|21949093\|ZDHHC20 | 0.127284769 | 0.048284 |
| | | | | chr13\|21949165\|ZDHHC20 | 0.119750317 | 0.031824 |

In Table 4.7 it can be observed that the majority of the editing events favor a good prognosis (labeled in blue), and that more edited sites are reported in TCGA cohort, due to the larger sample of editing events available. High editing frequencies in chr15|65442098|CLPX and chr13|21948495|ZDHHC20 position correlate with better overall survival in both datasets.

### 4.3.2 The prognostic role of editing events in CLPX and ZDHHC20 gene

In order to categorize our patients based on editing frequencies at the reported editing sites in Table 4.7, a Log rank multiple cut–off (LRMC) analysis was performed to select a threshold editing level for each event. From the LRMC plots we can observe that ZDHHC20 editing frequencies plotted in x-axis span the same range in both cohorts. Multiple significant cut-offs can be identified, mainly in the 2nd quartile of editing levels which is between 30%-40% in both cohorts (Fig4.12). Also we notice that in all of the samples in Singapore Cohort and in most of the samples in TCGA cohort, editing events in ZDHHC20 gene are associated to good prognosis. The most significant threshold was selected as a cut-off to group patients into "High" and "Low" based on the editing levels.

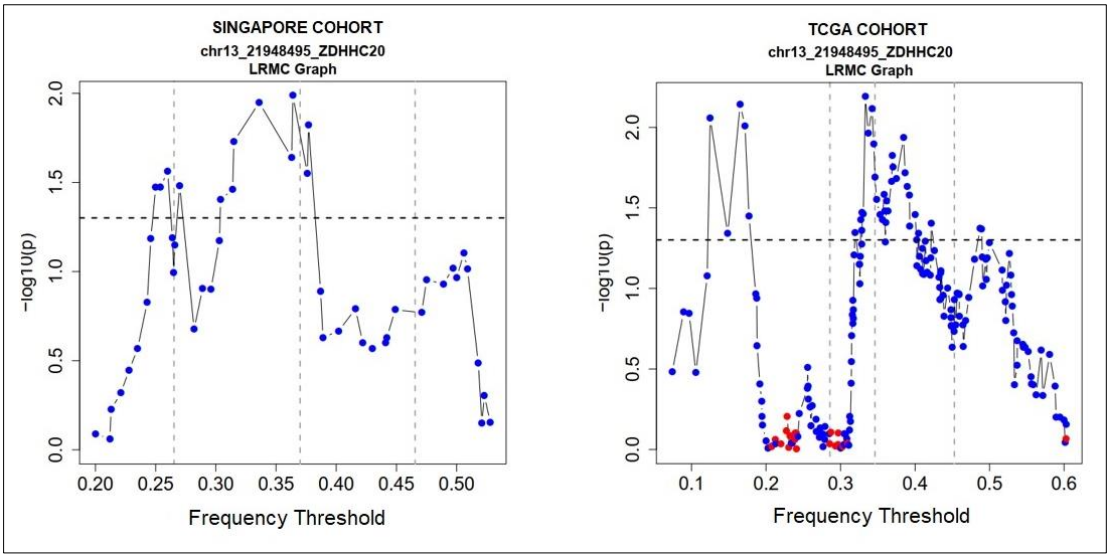



**Figure 4.12:** LRMC plots for editing event at 21948495 position in ZDDHC20 gene. Editing frequencies across patients are depicted in x-axis and –log10 of p-value on y-axis. The horizontal dashed line represents the significance cut off (0.05) and the vertical lines represent the 1st, 2nd and 3rd quartile of editing frequency. Each dot is a representation of patient editing values.

For editing frequencies in CPLX, we notice that the editing levels in patients are the same in both cohorts ranging from 10-50%. Overall editing events associate to good prognosis, but we notice a shift to bad prognosis at editing levels closer to 50%. Most of the significant cut-offs in TCGA cohort are found before the 1st quartile, while in the Singapore cohort a higher number of significant cut-offs at different frequency levels can be identified. In order to stratify the patients uniformly in both cohorts we decided to opt for a cut-off value in the first 25th percentage of editing frequencies, despite it being of a lower significance than other thresholds available (**Fig4.13**).

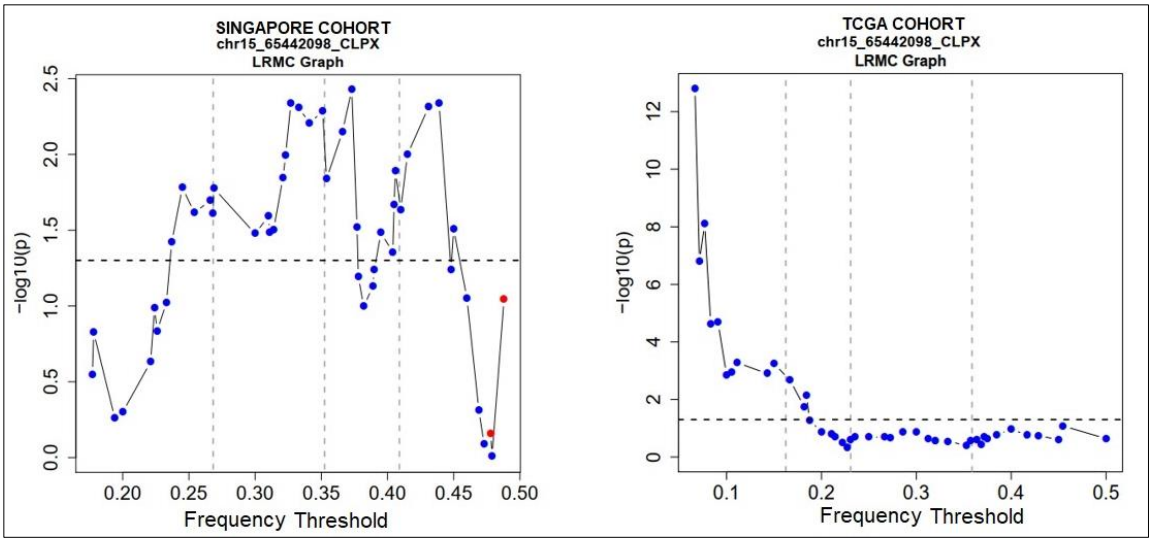



**Figure 4.13.** LRMC plots for editing events in 65442098 position in CLPX gene.

After categorizing the patients into "High" and "Low" based on the editing frequencies in ZDHHC20 and CLPX gene respectively, we analyzed their survival differences. For ZDHHC20, the high edited group shows a significant better OS than the low edited group in both datasets. In the Singapore Cohort the high edited group shows a better median survival time by more than 3 months compared to the low edited one.

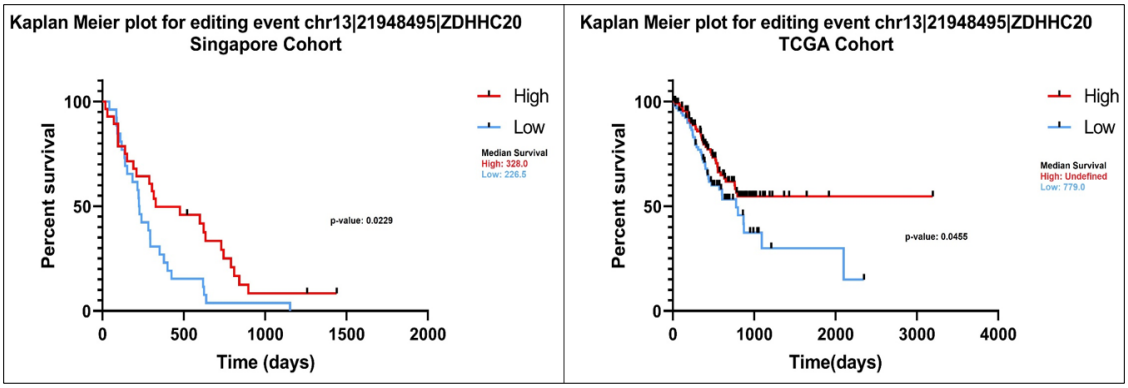



**Figure 4.14.** Kaplan-Meier plots for editing at 21948495 position in ZDHHC20 gene. Red line represents high edited group and blue line low edited group in this position.

The same trends are observed when patients are stratified based on CLPX editing frequencies. Patients in the "High" group demonstrate a five times higher median survival time than those in the "Low" group. Even though the follow-up periods are different, the survival trends observed are the same in both cohorts.

To test the robustness of the commonly identified editing events in CLPX and ZDHHC20 gene, and identify the covariates that independently affect OS, multiple logistics regression analysis (Backward Wald) was performed. In the TCGA cohort, baseline characteristics such as, gender, TNM staging and response to primary treatment were included in the analysis together with categorical (high and low) editing levels in CLPX. Editing in CLPX and responsiveness to primary treatment were identified as independent good prognostic factors (HR<1, p-value <0.05). In the first step of the Backward Wald analysis we notice that editing levels in CLPX is an independent prognostic factor even when other co-variates such as gender and tumor stage are included in the analysis. In the last step of the analysis gender and stage are discarded based on their insignificant association with patient survival. In the Singapore Cohort, clinic-pathological parameters such as gender, race, Lauren`s classification, treatment and overall response were included. In the Singapore Cohort, more covariates were included due to the clinical data being more informative than TCGA`s. Editing frequency in CLPX is an independent good prognostic factor, aside race and overall response (HR<1, p-value <0.05). Editing levels in CLPX accurately predict a

better OS even in the first analysis when all co-variates selected are included (95% CI: 0.195-0.983, p-value: 0.045).

Backwards Wald regression analysis was performed with categorical editing levels in ZDHHC20 gene and the above mentioned parameters as well. In both cohorts, even though categorical editing level in ZDHHC20 shows an association to better OS, it cannot be used as an independent prognostic factor (HR<1, p-value >0.05).

**Table 4.8:** Backward Wald analysis results for Singapore Cohort. In the first step all listed co-variates are included and in the last step only best fitting variables are shown. HR: Hazard Ratio, CI: Confidence Interval, PD: Progressive disease, PR: Partial Response, SD: Stable Disease.

| **Singapore** | | **First Step** | | | **Last Step** | | |
|---|---|---|---|---|---|---|---|
| **Characteristics** | | **HR** | **95% CI** | **p-value** | **HR** | **95% CI** | **p-value** |
| **chr15\|65442098\|CLPX** | | 0.403 | 0.163-0.995 | 0.049 | 0.438 | 0.195-0.983 | **0.045*** |
| Race | | 0.517 | 0.323-0.827 | 0.006 | 0.512 | 0.318-0.826 | **0.006**** |
| Gender | | 1.088 | 0.484-2.449 | 0.838 | -- | -- | -- |
| Treatment | | 0.679 | 0.324-1.424 | 0.306 | -- | -- | -- |
| Histology | Diffused | 1.729 | 0.421-7.094 | 0.447 | 1.648 | 0.420-6.475 | 0.474 |
| | Intestinal | 0.736 | 0.211-2.570 | 0.631 | 0.674 | 0.242-1.880 | 0.451 |
| Overall Response | PD | 0.302 | 0.070-0.1.313 | 0.110 | 0.327 | 0.089-1.208 | 0.094 |
| | PR | 0.025 | 0.005-0.113 | 0.000 | 0.032 | 0.008-0.129 | **0.000***** |
| | SD | 0.089 | 0.023-0.337 | 0.000 | 0.095 | 0.026-0.342 | **0.000***** |
| | | | | | | | |
| **chr13\|21948495\|ZDHHC20** | | 0.753 | 0.362-1.566 | 0.447 | -- | -- | -- |
| Race | | 0.582 | 0.359-0.943 | 0.028 | 0.625 | 0.421-0.929 | **0.020*** |
| Gender | | 1.217 | 0.524-2.829 | 0.648 | -- | -- | -- |
| Treatment | | 0.828 | 0.413-1.662 | 0.596 | -- | -- | -- |
| Histology | Diffused | 0.983 | 0.257-3.759 | 0.980 | -- | -- | -- |
| | Intestinal | 0.505 | 0.159-1.608 | 0.248 | -- | -- | -- |
| Overall Response | PD | 0.471 | 0.120-1.849 | 0.280 | 0.440 | 0.128-1.515 | 0.193 |
| | PR | 0.037 | 0.009-0.163 | 0.000 | 0.042 | 0.011-0.160 | **0.000***** |
| | SD | 0.128 | 0.037-0.447 | 0.001 | 0.131 | 0.040-0.431 | **0.001**** |

**Table 4.9:** Backward Wald analysis results for TCGA Cohort.

| **TCGA** | | **First Step** | | | **Last Step** | | |
|---|---|---|---|---|---|---|---|
| **Characteristics** | | **HR** | **95% CI** | **p-value** | **HR** | **95% CI** | **p-value** |
| **chr15\|65442098\|CLPX** | | 0.366 | 0.180-0.741 | 0.005 | 0.361 | 0.177-0.734 | 0.005 |
| Gender | | 0.724 | 0.339-1.547 | 0.404 | -- | -- | -- |
| Stage | | 1.294 | 0.718-2.332 | 0.392 | -- | -- | -- |
| Primary Response | Responsive | 0.336 | 0.119-0.950 | 0.040 | 0.290 | 0.117-0.716 | 0.007 |
| | Non-responsive | 1.339 | 0.372-4.827 | 0.655 | 1.125 | 0.338-3.745 | 0.848 |
| | Mixed | 0.000 | 0.000 | 0.980 | 0.000 | 0.000 | 0.980 |
| | | | | | | | |
| **chr13\|21948495\|ZDHHC20** | | 0.670 | 0.426-1.055 | 0.084 | 0.673 | 0.427-1.059 | 0.087 |
| Gender | | 0.831 | 0.525-1.313 | 0.427 | -- | -- | -- |
| Stage | | 1.432 | 1.020-2.010 | 0.038 | 1.414 | 1.009-1.980 | 0.044 |
| Primary Response | Responsive | 0.380 | 0.199-0.727 | 0.003 | 0.391 | 0.205-0.746 | 0.004 |
| | Non-responsive | 0.932 | 0.445-1.951 | 0.851 | 0.945 | 0.451-1.978 | 0.880 |
| | Mixed | 0.243 | 0.053-1.108 | 0.068 | 0.244 | 0.054-1.113 | 0.069 |

### 4.3.3 Prognostic potential of gene expression levels in GC patients

To ensure that the identified editing events associated with OS are not redundant prognostic factors, we decided to perform a univariate cox regression analysis to check if the gene expression levels of the 15 genes listed in Table 7 show a correlation with patient survival. We used gene expression data from TCGA cohort and identified that out of 15 genes only ZDHHC20 and GNPNAT1 expressions are associated to OS (**Appendix G**). Univariate cox regression analysis showed that CLPX gene expression is not associated to survival in TCGA cohort patients (**Table 4.10**). Our analysis indicates that ZDHHC20 gene expression but not its editing levels can serve as an independent prognostic factor for OS in GC patients. Gene expression level of CLPX gene was not associated to OS, indicating that the identified editing event in CLPX serves as an autonomous prognostic factor.

**Table 4.10:** Univariate Cox Regression analysis results for CLPX gene expression.

| **Univariate Cox Regression TCGA (n=224)** | | | |
|---|---|---|---|
| Variable | HR | 95% CI | P-value |
| CLPX | 1.000 | 0.999-1.001 | 0.83 |

### 4.3.4 Editing events associate to good and bad prognosis in GC

Editing events in CLPX and ZDHHC20 genes were selected based on their consistent association with OS in GC patients in two different cohorts. Nevertheless, we decided to explore the prognostic potential of two editing events, one associated with bad prognosis and one with good prognosis, from each dataset. The selected editing events are not reported in both cohorts, thus at the time being their validation is not possible. However this does not take away from their prognostic significance. From the LRMC analysis we noticed that some editing events showed promising prognostic potential based on different criteria such as; highly significant cut-off values close to the median editing frequency and a good range of editing frequency.

In Singapore Cohort, we selected the editing event at chr19|58372048 position in ZNF587 gene, as a good prognostic biomarker. From the LRMC plot we can observe that it has a good editing frequency ranging from 50-80%. The significant cut-off data points are distributed in between the 1st and 3rd quartile and the most significant threshold, close to the median editing frequency, was selected for patient stratification. At all frequencies editings in ZNF587 are associated to good prognosis (**Fig4.15A**). As a bad prognostic biomarker editing event at chr4|17803559 position in DCAF16 gene was selected.

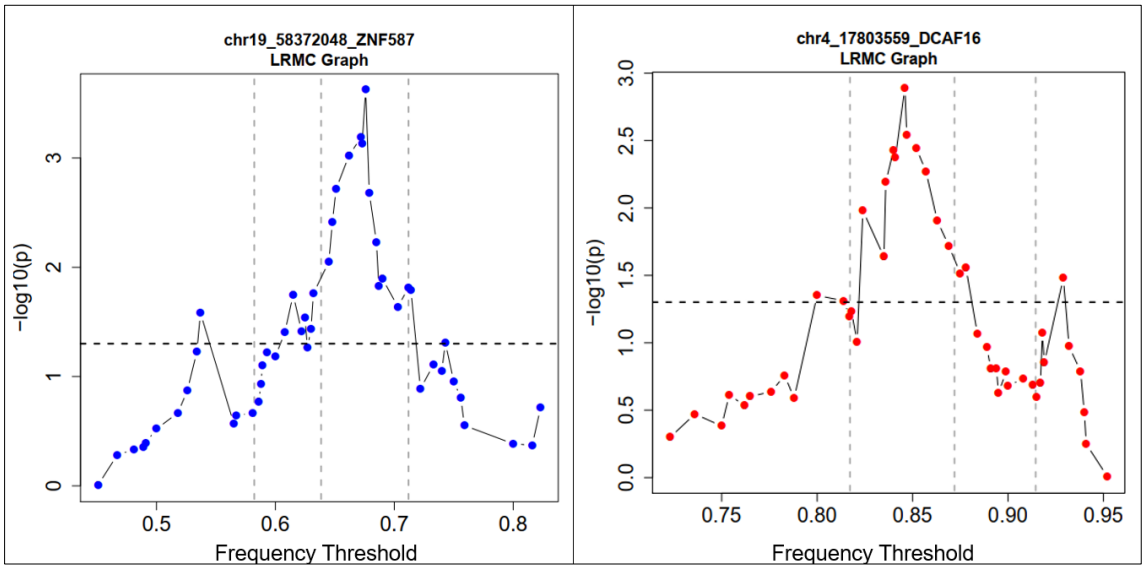



**Figure 4.15:** LRMC plot for editing events in ZNF587 (A) and DCAF16 (B) gene in Singapore Cohort.

The LRMC plot shows similar distribution trends as the editing event in ZNF587, however we notice a relatively high editing frequency as compared to other editing events reported in the dataset (Range 75%-95%). The most significant cut-off point was selected, since beside its significance, the proximity to the 2nd quartile allows for the division of patients into almost equal groups (**Fig 4.15B**).

The patients were classified into "High" and "Low" groups based on the described cut-offs. In the Kaplan Meier plot for ZNF587 we can observe that the high edited group shows a significant better OS and longer median survival time compared to the low edited group (**Fig 4.16A**). The opposite trend is observed in the DCAF16 KM plot, where the high edited group has a significant worse OS and a 64% lower median survival time compared to the low edited group (**Fig 4.16B**).

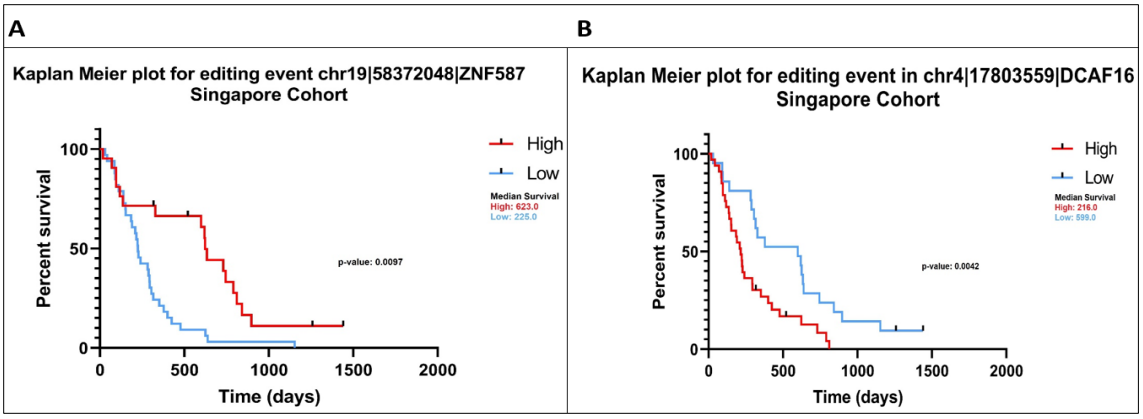



**Figure 4.16:** Kaplan Meier plot for editing event in ZNF587 (A) and DCAF16 gene (B).

To test their validity as independent prognostic factors, Backward Wald regression analysis was performed, including the covariates listed in Table 6. Editing frequency in ZNF587 is an independent good prognostic factor (HR<1, p-value< 0.05) and editing frequency in DCAF16 is an independent bad prognostic factor (HR>1, p-value< 0.05). Race and overall response to therapy are independent prognostic factors as well associated to better OS in Singapore Cohort (HR<1) (**Table 4.11**).

**Table 4.11:** Backward Wald analysis results for editing event in ZNF587 and DCAF16 gene in Singapore Cohort.

| **Singapore** | | **First Step** | | | **Last Step** | | |
|---|---|---|---|---|---|---|---|
| **Characteristics** | | **HR** | **95% CI** | **p-value** | **HR** | **95% CI** | **p-value** |
| **chr19\|58372048\|ZNF587** | | 0.319 | 0.123-0.828 | **0.019*** | 0.320 | 0.147-0.699 | **0.004**** |
| Race | | 0.519 | 0.318-0.846 | **0.009**** | 0.522 | 0.346-0.786 | **0.002**** |
| Gender | | 0.875 | 0.363-2.113 | 0.767 | -- | -- | -- |
| Treatment | | 0.815 | 0.420-1.581 | 0.545 | -- | -- | -- |
| Histology | Diffused | 0.632 | 0.155-2.571 | 0.521 | -- | -- | 0.474 |
| | Intestinal | 0.514 | 0.158-1.674 | 0.269 | -- | -- | 0.451 |
| Overall Response | PD | 0.250 | 0.054-1.160 | 0.077 | 0.247 | 0.066-0.922 | 0.094 |
| | PR | 0.033 | 0.007-0.152 | **0.000***** | 0.035 | 0.009-0.139 | **0.000***** |
| | SD | 0.085 | 0.021-0.343 | **0.001**** | 0.078 | 0.022-0.283 | **0.000***** |
| | | | | | | | |
| **chr4\|17803559\|DCAF16** | | 2.816 | 1.369-5.793 | **0.005**** | 2.805 | 1.412-5.571 | **0.003**** |
| Race | | 0.559 | 0.347-0.901 | **0.017*** | 0.561 | 0.351-0.898 | **0.016*** |
| Gender | | 1.346 | 0.643-2.817 | 0.431 | -- | -- | -- |
| Treatment | | 1.118 | 0.541-2.309 | 0.764 | -- | -- | -- |
| Histology | Diffused | 1.229 | 0.333-4.538 | 0.757 | 1.253 | 0.342-4.585 | -- |
| | Intestinal | 0.369 | 0.118-1.151 | 0.086 | 0.440 | 0.156-1.239 | -- |
| Overall Response | PD | 0.530 | 0.129-2.166 | 0.376 | 0.430 | 0.122-1.516 | 0.193 |
| | PR | 0.047 | 0.011-0.205 | **0.000***** | 0.043 | 0.011-0.167 | **0.000***** |
| | SD | 0.136 | 0.038-0.483 | **0.002**** | 0.124 | 0.036-0.424 | **0.001**** |

In TCGA cohort editing event in chr8|11700426 position in CTSB gene was selected as a good prognostic biomarker and editing at chr6|150046473 position in NUP43 gene was selected as a bad prognostic biomarker. Both editing events have a similar range of editing levels (0-20%). Due to the lower sequencing depth in TCGA cohort RNA-sequencing data, a lot of patients lack editing frequency information at specific sites. The relatively higher number of data for these editing events was considered in the selection process as well, in order to perform a better informed analysis. In the LRMC plot of CTSB editing we observe that it is associated to good prognosis in the majority of the frequency values and a significant number of cut-offs are available, from the 1$^{st}$ to the 3$^{rd}$ quartile. We selected the most significant cut-off point since the significant values are clustered very closely to each other, and although closer to the 3$^{rd}$ quartile in frequency, the selected threshold separates the patients into almost equal groups (**Fig 4.17A**). The NUP43 editing LRMC plot shows its association to a worse survival and indicates that almost all the points can be used as significant cut-offs. We opted for the most significant cut-off value due to it being in the 3$^{rd}$ quartile unlike other significant values which are dispersed closer to the minimum and maximum frequency values which would skew our analysis result due to unequal patient stratification (**Fig 4.17B**).

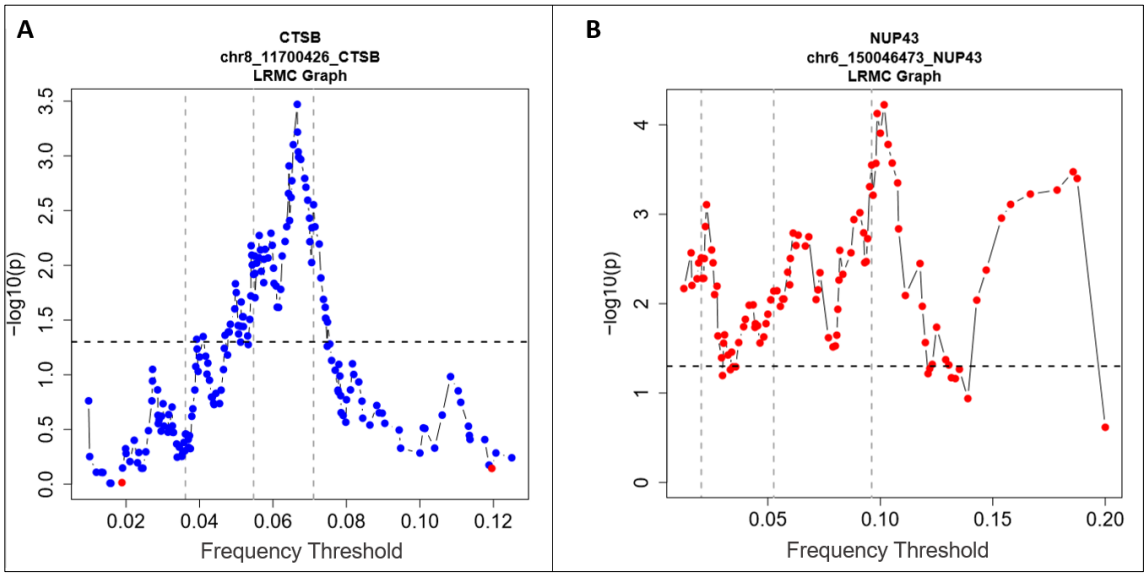



**Figure 4.17:** LRMC plot for editing events in CTSB (A) and NUP43 (B) gene in TCGA Cohort.

In the survival analysis we observe that NUP43 high edited group shows a significant worse survival and almost a 4-times lower median survival time compared to the low edited group (**Fig 4.18B**). The high edited group in CTSB shows a significant better OS than the low edited group, however due to the high amount of censored subjects the median survival time is undefined for the high edited group (**Fig 4.18A).**

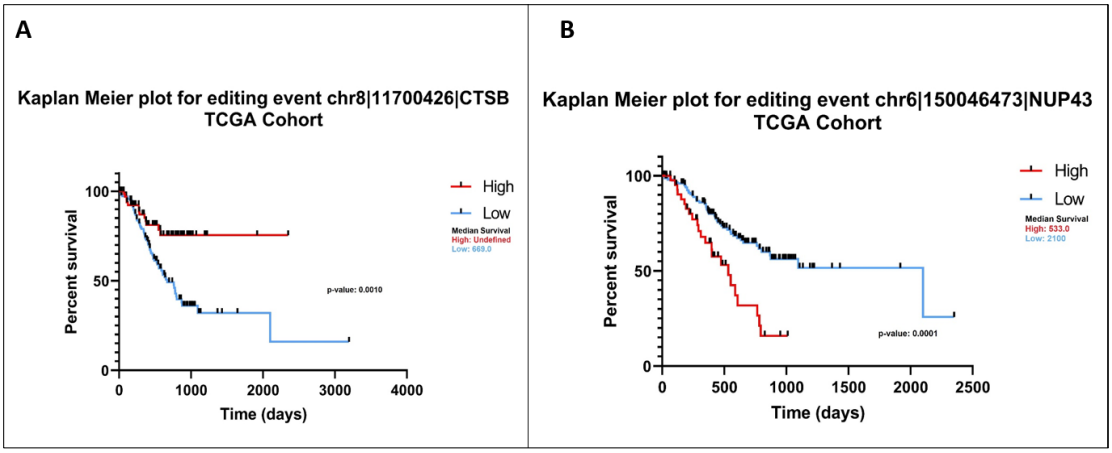



**Figure 4.18:** Kaplan Meier plot for editing event in CTSB (A) and NUP43 gene (B).

Backward Wald analysis indicates that editing frequency in NUP43 has a high significance but lower precision in independently predicting a worse OS in GC patients (p-value< 0.0001, 95% CI: 1.613-4.510).

**Table 4.12:** Backward Wald analysis results for editing event in ZNF587 and DCAF16 gene in Singapore Cohort.

| **TCGA** | | **First Step** | | | **Last Step** | | |
|---|---|---|---|---|---|---|---|
| **Characteristics** | | **HR** | **95% CI** | **p-value** | **HR** | **95% CI** | **p-value** |
| **chr8\|11700426\|CTSB** | | 0.439 | 0.240-0.803 | **0.008**** | 0.440 | 0.240-0.805 | **0.008**** |
| Gender | | 0.837 | 0.530-1.321 | 0.444 | -- | -- | -- |
| Stage | | 1.435 | 1.031-1.997 | **0.032*** | 1.419 | 1.021-1.972 | **0.037*** |
| Primary Response | Responsive | 0.391 | 0.208-0.736 | **0.004**** | 0.407 | 0.218-0.759 | **0.005**** |
| | Non-responsive | 0.857 | 0.417-1.759 | 0.674 | 0.873 | 0.425-1.791 | 0.711 |
| | Mixed | 0.232 | 0.052-1.041 | 0.056 | 0.232 | 0.052-1.040 | 0.056 |
| | | | | | | | |
| **chr6\|150046473\|NUP43** | | 2.642 | 1.574-4.436 | **0.000***** | 2.697 | 1.613-4.510 | **0.000***** |
| Gender | | 1.051 | 0.648-1.707 | 0.839 | -- | -- | -- |
| Stage | | 1.283 | 0.879-1.873 | 0.197 | -- | -- | -- |
| Primary Response | Responsive | 0.256 | 0.125-0.523 | **0.000***** | 0.232 | 0.116-0.462 | **0.000***** |
| | Non-responsive | 0.533 | 0.232-1.222 | 0.137 | 0.500 | 0.221-1.134 | 0.097 |
| | Mixed | 0.138 | 0.029-0.649 | **0.012*** | 0.121 | 0.026-0.559 | **0.007**** |

Editing in CTBS is an independent and accurate prognostic factor when considering significant and non-significant clinic-pathological parameters in the analysis (**Table 4.12).**

Overall 5 unique editing events show independent prognostic function (**Table 4.13**). Interestingly all the editing sites are located in the 3'UTR region of the respective transcripts.

**Table 4.13**. Summary of the identified prognostic Editing Events

| EDITING EVENT | TRANSCRIPT REGION | PROGNOSIS | COHORT |
|---|---|---|---|
| chr15\|65442098\|CLPX | 3'UTR | Good | Both (TCGA & Singapore) |
| chr19\|58372048\|ZNF587 | 3'UTR | Good | Singapore |
| chr4\|17803559\|DCAF16 | 3'UTR | Bad | Singapore |
| chr8\|11700426\|CTSB | 3'UTR | Good | TCGA |
| chr6\|150046473\|NUP43 | 3'UTR | Bad | TCGA |

# CHAPTER 5

# 5. Discussion and Conclusion

Early diagnosis of Gastric Cancer still remains a major obstacle, thus most of the cases present itself at advanced stages, exhibiting a poor prognosis and response to chemotherapy [28]. The heterogeneous nature of the disease causes limited application of the so far discovered molecular biomarkers, suggesting an existent need for novel ones [85]. Advances in next generation sequencing technology in the last decades, have opened new avenues for exploring post-transcriptional modifications such as RNA-editing events [86]. A-to-I editing events have been linked to pathogenesis and progression of different cancer types, however the role of these modifications in gastric cancer is poorly elucidated [68, 87].

## 5.1 Identified Editing Landscape in GC Cell Lines

In this study we analyzed RNA-sequencing data from gastric cancer cell lines, using a RNA editing event identification tool, REDItools. Using REDItools we were able to identify more than 500.000 editing sites, previously reported in literature as sites where editing events occur. However only about 100.000 of these sites were found to be edited in at least one of the samples, possibly due to the low sequencing depth and the lack of a significant amount of reads at the specific site. The positive correlation observed between the total number of reads and number of edited sites identified per sample also indicates that a higher sequencing depth would result in a higher number of editing events. Other studies also report similar correlations and suggest that higher sequencing depth, results in a higher percentage of detectable editing events [68, 72, 75].Editing frequency, the ratio of edited reads to the total number of reads at a specific site, was used to quantify the editing levels. We show that editing levels are not correlated to the number of mapped reads at a specific position, which indicates there was no bias in the quantification method towards editing events located in genomic positions with a higher sequencing coverage. Only editing sites edited in at least 50% of the samples were selected to ensure the statistical significance of our analysis. The selected editing events were located mainly in the 3’ UTR, intergenic and intronic regions, which is consistent with the distribution other studies have reported [68,

72]. Unsurprisingly most of the editing events are located in non-coding regions due to the high density of ALU-rich sequences, which have been shown to be preferred target sites for ADAR enzymes [88].

## 5.2 Selection of Drug Sensitivity Measures

Given their involvement in cancer development and progression, we decided to test the potential of the identified RNA editing event as chemosensitivity biomarkers. Drug sensitivity data from 3 independent datasets, GDSC (version 1 and version2), CTRP and CCLE were used in the analysis. Even though IC50 (50% inhibitory concentration) is the traditional used metric for drug sensitivity, studies suggest that its accuracy is low due to the assumption that the drug-dose response curve will fit a canonical shaped sigmoid curve [89]. However, many drug sensitivity curves are incomplete and never reach 50% inhibitory concentrations. Different datasets have taken different approaches to resolve this issue such as capping the IC50 value by assigning the maximal concentration tested, or inferring IC50 values by using the available curve, which leads to out of tested range values [90]. Both of these approaches do not reflect actual experimental outcomes making IC50 an inaccurate metric, especially when comparing drug response values from different datasets. On the other hand AUC (Area under the curve) or its inverse metric Activity Area, is a sensitivity metric that can be calculated for any drug-dose response curve , and is more apprehensive of the experimental set up, as it combines information about the potency and efficacy of a said drug. AUC has been shown to be a robust measure when comparing drug sensitivity across cell lines, however it is dependent on the drug concentration range, especially when there is a significant difference in the maximum concentration tested between datasets [91, 92]. We tried to address the differences in concentration range, by using PharmacoGx, an R package developed for integration of drug sensitivity measures [93]. Using this package we tried re-computing the AUC using only the common concentration range between datasets. However, we identified some labelling issues, in some of the data provided by the package which would skew our interpretation of the drug sensitivity. We reported the issue, to the developing team but due to the time limitations of our study we decided to proceed in our analysis with the published AUC/ Activity Area

values in each dataset, taking into consideration that the differences in methodologies might reflect in our results

## 5.3 Chemosensitivity Biomarkers in GC Cell Lines

To support the clinical translation of our findings, a pool of 29 FDA approved chemotherapeutics was selected, including drugs used on GC patients and other cancer types. We were able to identify significant correlations between editing events and drug sensitivity in all tested datasets, confirming our initial claim that editing events could be used as chemosensitivity biomarkers. In most of the cases high editing levels are associated with increased chemosensitivity. A recent study in GC, shows a similar trend as high average editing levels in 26 editing positions relate to high chemosensitivity toward Oxaliplatin in GC cell lines [72]. However based on our observations a generalization on the effect of the editing events on chemosensitivity cannot be made, as high editing levels could be associated to an increased responsiveness or resistance, depending on the site that is being edited. Due to the editing position specific behavior observed, we aimed to identify a drug sensitivity-editing event combination and validate their association across different datasets. In our initial analysis we were not able to identify common editing event-drug combinations in between datasets. We believe the reason for the lack of cross-validation is related to the different methodologies employed by each dataset to produce the drug response data and the fact that the applied statistical corrections might have be too stringent for the non-uniform nature of the data. Using the RANKSUM approach we were able to identify highly associated drug-editing event combination, validated in all tested datasets. Editing events seem to most significantly affect cell line response to tyrosine kinase inhibitors (Dasatinib, Erlotinib) and HER2 targeting drugs (Fulvestrant, Tamoxifen). Interestingly none of the listed drugs in Table 4 is currently used as a treatment option for GC and for each of them editing events whose high frequency is related to increased sensitivity are reported. Our findings suggest the identified editing events could be useful biomarkers to select patients for treatment with repurposed chemotherapeutics. We also confirmed that the observed associations are exclusively reliant on editing levels of the selected sites and not a mere reflection of a correlation at a transcript level. Even though, a small number of transcripts show an association to drug sensitivity in specific datasets,

the correlation observed is weaker than the one observed with the editing levels, which would not be the case if the association of editing events was dependent on the gene expression levels of the gene they are located in.

Despite the fact that ADAR enzymes are the executors of A-to-I editings in mammals, we show that the editing levels at only a few editing sites correlate with ADAR1 expression levels. The editome landscape in GC, is described as mis-edited when compared to normal tissue, and since deamination of Adenine is not a cancer specific event, overexpression of ADAR is not required for editing events to occur [69]. Additionally, the editing sites whose frequency correlates with ADAR1 expression have relatively high editing levels compared with the average levels identified in GC cell lines. Considering these facts, we would not expect to observe a high number of editing events correlating with ADAR expression, since the editing levels are not expected to be high for the majority of the sites. Other studies suggest that ADAR expression levels do not always correlate with editing frequencies, due to cis- and trans-acting factors or differences in cellular spatial localization [94, 95]. ADAR expression levels also are not proportionally correlated to the deaminase activity. As expected the editing frequencies positively correlate with ADAR1 expression. Considering that ADAR2 is mainly responsible for deamination of non-repetitive coding sequences, it is not surprising that editing frequencies at the identified sites did not correlate with its expression since the majority are located on non-coding, ALU repeats [95].

We were able to demonstrate the functional consequences of A-to-I editings in coding and non-coding regions. Editing levels in the 3'UTR of SOGA1 positively correlate with transcript levels. 3'UTR is involved in gene expression regulation by being a target site for miRNA or RNA binding proteins [96]. Previous studies have shown that miRNA target editing is one of the mechanisms by which editing events regulate gene expression [66, 97]. The identified editing event in SOGA1 gene, is located in the seed region of hsa-miR-9-5 target site, thus we propose that editing in this site disrupts the interaction miRNA-mRNA interaction. The annulment effect of the editing on the target site would result in increased transcript levels of SOGA1, as suggested by the positive correlation observed. We were able to confirm the expression of hsa-miR-9 in GC cell lines using expression data from CCLE, however we could not detect a correlation between SOGA1 and hsa-miR-

9 levels. SOGA1 is involved in the negative regulation of gluconeogenesis and autophagy. Even though the role of autophagy is debatable, most studies support the anti-tumor role specifically in hepatocytes [98]. Elevated levels of SOGA1 have been reported in hepatocellular carcinoma, which would could be explained by the inhibition of autophagy and its tumor-suppressive role [99]. SOGA1 is also an unfavorable prognostic marker for ovarian and renal cancer [100]. According to our findings, high editing levels in SOGA1 are associated to increased chemoresistance of GC cell lines to Vorinostat (SAHA), a HDAC inhibitor. A previous study shows that HDAC inhibitors such as SAHA are able to induce autophagy-mediated cell death in cancer cell lines [101]. We suspect that SOGA1 autophagy-inhibitory effect might be one of the mechanism involved in conveying chemoresistance to SAHA in GC cell lines (**Appendix H**). Unfortunately, this proposed mechanism is entirely hypothetical, as experimental validations are required to elucidate if the mechanism by which SOGA1 editing events convey chemoresistance is related to its autophagy inhibitory functions or not. Nevertheless our findings provide a good basis to consider exploring editings in SOGA1.

Despite the low percentage of editing events identified in coding regions in our analysis (<1%), we were able to identify a non-synonymous editing event in C11orf80 (TOP6BL) gene. The A-to-G editing translated into substitution of Serine at position 133 to Glycine, which based on prediction tools seems to have a neutral effect on protein function. Editing levels in C11orf80 CDS, are associated to decreased sensitivity to Lapatinib based on drug response data from 3 different datasets. C11orf80 is reported in a 6-gene prognostic signature for cervical cancer, where its expression levels are elevated in high-risk group [102]. To the best of our knowledge, the functional implication of this gene in gastric cancer or other cancers has not been investigated. In this study we report for the first time, the chemosensitivity biomarker potential of the editing event at position 66523904 in the coding sequence of C11orf80.

In an attempt to elucidate possible mechanisms involved in the chemosensitivity changes conveyed by editing events, we decided to perform a pathway enrichment analysis using the edited genes, for each drug separately. Only genes containing editings significantly correlated to Lapatinib sensitivity out of 17 drugs, were enriched significantly in certain

pathways. Editing events were located in genes, enriched in ribonucleoprotein, mitochondrion and ribosomal protein pathways (**Appendix I**). In breast cancer it has been shown that mitochondrion plays a role in metabolic adaption of cancer cells resistant to Lapatinib [103]. Ribonucleoproteins and ribosomal proteins being crucial factors in gene and protein expression regulations are involved in chemoresistance mechanisms in cancer, including Lapatinib resistance [104, 105]. The identified enriched pathways have a broad impact in tumorigenesis, thus a more detailed in depth analysis is required to identify the specific aspects affecting Lapatinib resistance in GC. As for the drugs where we could not identify enriched pathways in the edited genes, it could be a reflection of multiple pathways being deregulated by editing events at the same time, or it could be the case that these editing events are related by non-functional factors.

From our chemosensitivity analysis, we conclude that editing events in Gastric Cancer, could serve as autonomous biomarkers for chemosensitivity. Independently of gene expression levels or ADAR expression, editing events associate with drug sensitivity, consistently in different datasets. Our findings provide insights on the functional consequences these editing events relay and encourage a more thorough investigation to elucidate the mechanisms by which chemosensitivity is conveyed.

## 5.4 Prognostic Biomarkers in GC Patients

In the prognostic part of the study we aim to identify editing events with prognostic potential in gastric cancer patients using two cohorts, the Singapore and TCGA cohort. The low availability of patient cohorts providing RNA-sequencing or RNA editing data hindered our discovery and validation, so to ensure the robustness of our findings we aimed to identify editing events at the exact same position in both datasets. The univariate cox regression analysis revealed that high editing levels associate with better prognosis for the most part (HR<1), which is consistent with the findings of the Singapore Cohort study. Despite the discrepancies in sequencing techniques, sequencing data processing and selection of the editing events in between the cohorts, we were still able to identify 2 editing events in common. Editing events in CLPX and ZDHHC20 gene are both associated to better prognosis. However, we could only identify CLPX as an independent prognostic factor in both cohort, as ZDHHC20 editing levels could not associate with prognosis in the

multiple regression analysis. Based on univariate and multivariate analysis ZDHHC20 gene expression levels associate with better prognosis. This indicates that the association of the editing event in ZDHHC20 gene was a reflection of the correlation of transcript levels. CLPX gene expression was not a prognostic factor in both cohorts. Based on literature, CLPX gene expression or editing events have not been associated with prognosis or any other mechanism in gastric cancer, however CLPX is reported as favorable prognostic marker in renal cancer based on TCGA RNA sequencing data [100].

Even though these cohorts have been used in separate studies to identify the prognostic editing events in gastric cancer, their approaches quite differs from ours. In the Singapore Cohort, 780 editing events were used to categorize patients into high and low editing clusters, making the quantification process unfeasible for use in clinics due to the high cost [72]. In the TCGA cohort, focus was placed on editing events causing non-synonymous amino acid changes, which are rarely encountered and might not be consistently present in patients [68]. Our approach uses editing levels at specific sites to categorize patients into high and low edited, taking into account the range of editing frequency to select the most significant cut-off. In addition to the common editing events we were able to identify 13 common edited genes in both cohorts. We selected 2 editings events per cohort, one good and one bad prognostic factor that would best classify our patients into two groups of comparable size to avoid any discrepancies in our survival analysis. From the Singapore cohort editing event in 3'UTR of ZNF587 was selected as good prognostic biomarker and our analysis indicate that it could be a significant independent biomarker. Zinc finger proteins are involved in crucial processes such as transcription regulation so it is to no surprise that over the past few years research has indicated they play a role in cancer progression [106]. ZNF587 is highly expressed in stomach tissue and a recent study suggests that downregulation of ZNF587 by miR-4537, increases cell sensitivity to radiation in GC. Although the mechanism by which ZNF587 conveys resistance to radiation are not elucidated, the study suggest ZNF587 has more of an oncogenic function in GC [107]. Due to the location 3'UTR location of our editing event, it might be possible that new target sites are created, causing downregulation of ZNF587, resulting in better survival in high edited patients. DCAF16 was selected as the bad prognostic editing event in the Singapore cohort, and it was noted that the editing levels were unusually high ranging

from 75%-95%. These high editing levels were reported in the A-to-I editing atlas as well, in almost all tissue types, suggesting that although this editing event is an independent prognostic factor, it might not be specific to GC [77]. In TCGA cohort, interestingly the editing showing a significant correlation to good prognosis is located in Cathepsin B (CTSB) gene which codes for a crucial peptidase, and is involved in many different cancer mechanisms [108]. High serum levels of CTSB have been reported as bad prognostic markers, however our results indicate that the editing levels associate to good prognosis [109]. Such a dual role has been observed previously in gastric cancer as polymorphisms in CTSB have been associated with lower risk and better OS [110]. These contradictory associations put emphasis on the fact that editing events are independent biomarkers from gene expression. Lastly, editing at 3'UTR of NUP43 was associated with bad prognosis, as reported in Human Protein Atlas for breast and liver cancer [100]. NUP43 has been proposed as a putative target of miR-137 whose activity is blocked in Gastric Cancer, causing an increased expression of NUP43 which leads to a more aggressive cancer phenotype. It is propose that NCK1-AS1 serves as a cushion RNA to inhibit NUP43 targeting but in light of the discovered associations, it could be possible that the targeted sites are nullified via RNA editing [111]. The backwards Wald analysis indicates that the proposed editing events are robust prognostic markers.

All of the identified editing events with prognostic significance are located at the 3'UTR of the respective genes. RNA editing events have been reported to modify miRNA target sequences at the 3'UTR, regulating gene expression. To shed more light on the mechanisms by which the identified editing events convey their prognostic value, we wanted to investigate if the edited genes were targets of the same miRNAs. miRNET was used to identify the mRNA-miRNA interactions between the 5 genes in which the prognostic editing events are located, in gastric tissue (**Appendix J**). DCAF16, CTSB and NUP43 are common targets of 2 miRNAs, hsa-miR-449b-5p and hsa-miR-22-3p, which are significantly enriched in Leukemia, indicating an involvement in tumorigenesis. CTSB is targeted by hsa-miR-200c-3p, which has been reported to be involved in occurrence and progression of various cancers, including gastric cancer [112]. These preliminary findings from the miRNET network analysis, suggest further investigation of the role of editing

events in miRNA-mRNA interaction regulation is needed to understand the prognostic association.

In conclusion we demonstrate the prognostic biomarker potential of different editing events in different Gastric Cancer cohorts. Editing events are capable of independently predicting patient survival and are worth exploring not just as biomarkers but also at a functional level.

## 5.4 Future Perspectives

The findings of this study provide us with good indications of the potential of editing events as biomarkers for chemosensitivity. However, we hope to expand on these findings, starting with improvements in the editing identification pipeline. The use of higher depth sequencing data will increase the number of identified editing events, and allow us to apply more robust filters in the identification pipeline. Applying a minimum editing frequency or minimum edited read filter will provide us with more informative editing events. We hope to achieve this by sequencing our own RNA samples isolated from gastric cancer cell lines. To increase the coverage, enrichment of A-to-I edited transcript could be achieved by EndoVIPER-seq (Endonuclease V inosine precipitation enrichment sequencing) prior to sequencing analysis [113]. By performing drug screening in-vitro experiments on GC cell lines using some of the 17 selected drugs from our chemosensitivity analysis, we hope to confirm the role of the proposed biomarkers. Validation of our in-silico finding could open new avenues for the use of editing events as chemosensitivity biomarkers in clinics. This would improve chemotherapy treatment protocols for GC patients, as we propose several biomarkers for different drugs currently used in clinics, which would allow a wider application despite the heterogeneous nature of GC. Furthermore, we wish to elucidate the mechanisms by which the editing events convey chemosensitivity in GC cell lines by expanding the in-silico analysis and trying to validate some of the proposed mechanisms in-vitro.

For the prognostic part of the analysis, we wish to identify other datasets containing RNA-sequencing and clinical data of GC patients in order to validate the proposed prognostic biomarkers reported in one of the cohorts only. Also we wish to analyze the RNA-

sequencing data from TCGA cohort, to identify editing events using the pipeline established in the chemosensitivity part of the analysis. We hope that by re-analyzing the data we will be able to detect editing events more consistently present in GC patients. The end goal of the study is to validate the proposed editing events as prognostic biomarkers in patient tissue, as this would bring these findings closer to the clinical application phase. Despite the diversity of the utilized cohorts in our discovery, we were still able to identify robust prognostic markers, suggesting that a wider application of these biomarkers might be possible. The quantification of single editing events in patients would be more feasible than the so far proposed prognostic editing clusters, giving the identified editing events in this study an upper hand in clinical application. Lastly, we would like to expand on the mechanisms by which editing events convey the observed prognostic association, such as miRNA-mRNA interactions.

# CHAPTER 6

## 6. BIBLIOGRAPHY


1. Silverthorn, D.U., et al., *Human physiology : an integrated approach*. 6th ed. 2013, Boston: Pearson Education. xxii, 890, 69 p.
2. Hu, B., et al., *Gastric cancer: Classification, histology and application of molecular pathology*. J Gastrointest Oncol, 2012. **3**(3): p. 251-61.
3. Karimi, P., et al., *Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention.* Cancer Epidemiol Biomarkers Prev, 2014. **23**(5): p. 700-13.
4. Heidrun Rotterdam M.D, H.T.E.M.D., *Carcinoma of the Stomach. In: Pathology of the Stomach and Duodenum.* 1989, New York, NY: Springer.
5. Arnold, M., et al., *Global burden of oesophageal and gastric cancer by histology and subsite in 2018.* Gut, 2020. **69**(9): p. 1564-1571.
6. Crew, K.D. and A.I. Neugut, *Epidemiology of gastric cancer.* World J Gastroenterol, 2006. **12**(3): p. 354-62.
7. Lauren, P., *THE TWO HISTOLOGICAL MAIN TYPES OF GASTRIC CARCINOMA: DIFFUSE AND SO-CALLED INTESTINAL-TYPE CARCINOMA. AN ATTEMPT AT A HISTO-CLINICAL CLASSIFICATION.* Acta Pathol Microbiol Scand, 1965. **64**: p. 31-49.
8. Waldum, H.L. and R. Fossmark, *Types of Gastric Carcinomas.* Int J Mol Sci, 2018. **19**(12).
9. Munoz, N. *Gastric carcinogenesis*. in *Gastric carcinogenesis: proceedings of the 6th Annual Symposium of the European Organization for Cooperation in Cancer Prevention Studies (ECP). Amsterdam: Elsevier Science*. 1988.
10. Sobin, L.H., *The international histological classification of tumours.* Bull World Health Organ, 1981. **59**(6): p. 813-9.
11. *Comprehensive molecular characterization of gastric adenocarcinoma.* Nature, 2014. **513**(7517): p. 202-9.
12. Cristescu, R., et al., *Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes.* Nature Medicine, 2015. **21**(5): p. 449-456.
13. Niederhuber, J.E., et al., *Abeloff's Clinical Oncology E-Book*. 2019: Elsevier Health Sciences. 1197-1210.
14. Correa, P., *Gastric cancer: overview.* Gastroenterol Clin North Am, 2013. **42**(2): p. 211-7.
15. Cancer, I.A.f.R.o., *Schistosomes, liver flukes and Helicobacter pylori. IARC monographs on the evaluation of carcinogenic risks to humans.* 1994. **60**.
16. Correa, P., *Helicobacter pylori and gastric cancer: state of the art.* Cancer Epidemiology and Prevention Biomarkers, 1996. **5**(6): p. 477-481.
17. Helicobacter and C.C. Group, *Gastric cancer and Helicobacter pylori: a combined analysis of 12 case control studies nested within prospective cohorts.* Gut, 2001. **49**(3): p. 347-353.
18. Wiseman, M., *The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective.* Proc Nutr Soc, 2008. **67**(3): p. 253-6.
19. Ge, S., et al., *Association between Habitual Dietary Salt Intake and Risk of Gastric Cancer: A Systematic Review of Observational Studies.* Gastroenterol Res Pract, 2012. **2012**: p. 808120.
20. Yusefi, A.R., et al., *Risk Factors for Gastric Cancer: A Systematic Review.* Asian Pac J Cancer Prev, 2018. **19**(3): p. 591-603.
21. Ma, K., et al., *Alcohol Consumption and Gastric Cancer Risk: A Meta-Analysis.* Med Sci Monit, 2017. **23**: p. 238-246.

22. Fuchs, C.S. and R.J. Mayer, *Gastric carcinoma.* N Engl J Med, 1995. **333**(1): p. 32-41.
23. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: a cancer journal for clinicians, 2021. **71**(3): p. 209-249.
24. Morgan, E., et al., *The current and future incidence and mortality of gastric cancer in 185 countries, 2020–40: A population-based modelling study.* eClinicalMedicine, 2022. **47**: p. 101404.
25. Colquhoun, A., et al., *Global patterns of cardia and non-cardia gastric cancer incidence in 2012.* Gut, 2015. **64**(12): p. 1881-1888.
26. Thrift, A.P. and H.B. El-Serag, *Burden of Gastric Cancer.* Clin Gastroenterol Hepatol, 2020. **18**(3): p. 534-542.
27. Arnold, M., et al., *Is gastric cancer becoming a rare disease? A global assessment of predicted incidence trends to 2035.* Gut, 2020. **69**(5): p. 823-829.
28. Song, Z., et al., *Progress in the treatment of advanced gastric cancer.* Tumor Biology, 2017. **39**(7): p. 1010428317714626.
29. Oba, K., et al., *Role of chemotherapy for advanced/recurrent gastric cancer: an individual-patient-data meta-analysis.* Eur J Cancer, 2013. **49**(7): p. 1565-77.
30. Verdecchia, A., et al., *Explaining gastric cancer survival differences among European countries.* Int J Cancer, 2004. **109**(5): p. 737-41.
31. Sexton, R.E., et al., *Gastric cancer: a comprehensive review of current and future treatment strategies.* 2020. **39**(4): p. 1179-1203.
32. Sasako, M., et al., *Five-year outcomes of a randomized phase III trial comparing adjuvant chemotherapy with S-1 versus surgery alone in stage II or III gastric cancer.* J Clin Oncol, 2011. **29**(33): p. 4387-93.
33. Jayanathan, M., et al., *MAGIC versus MacDonald treatment regimens for gastric cancer: Trends and predictors of multimodal therapy for gastric cancer using the National Cancer Database.* Am J Surg, 2020. **219**(1): p. 129-135.
34. Digklia, A. and A.D. Wagner, *Advanced gastric cancer: Current treatment landscape and future perspectives.* World J Gastroenterol, 2016. **22**(8): p. 2403-14.
35. Matsuoka, T. and M. Yashiro, *Biomarkers of gastric cancer: Current topics and future perspective.* World J Gastroenterol, 2018. **24**(26): p. 2818-2832.
36. Shimada, H., et al., *Clinical significance of serum tumor markers for gastric cancer: a systematic review of literature by the Task Force of the Japanese Gastric Cancer Association.* Gastric Cancer, 2014. **17**(1): p. 26-33.
37. Deng, K., et al., *The prognostic significance of pretreatment serum CEA levels in gastric cancer: a meta-analysis including 14651 patients.* PLoS One, 2015. **10**(4): p. e0124151.
38. Nagatsuma, A.K., et al., *Expression profiles of HER2, EGFR, MET and FGFR2 in a large cohort of patients with gastric adenocarcinoma.* Gastric Cancer, 2015. **18**(2): p. 227-38.
39. Hosoda, K., et al., *Prognostic relevance of FGFR2 expression in stage II/III gastric cancer with curative resection and S-1 chemotherapy.* Oncol Lett, 2018. **15**(2): p. 1853-1860.
40. Huang, X., et al., *Clinical value of CagA, c-Met, PI3K and Beclin-1 expressed in gastric cancer and their association with prognosis.* Oncol Lett, 2018. **15**(1): p. 947-955.
41. Mendonsa, A.M., T.Y. Na, and B.M. Gumbiner, *E-cadherin in contact inhibition and cancer.* Oncogene, 2018. **37**(35): p. 4769-4780.
42. Pharoah, P.D., P. Guilford, and C. Caldas, *Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families.* Gastroenterology, 2001. **121**(6): p. 1348-53.
43. Corso, G., et al., *Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer.* J Clin Oncol, 2013. **31**(7): p. 868-75.
44. Wen, J., et al., *Promoter methylation of tumor-related genes as a potential biomarker using blood samples for gastric cancer detection.* Oncotarget, 2017. **8**(44): p. 77783-77793.

45. Hudson, K., et al., *The Extrinsic and Intrinsic Roles of PD-L1 and Its Receptor PD-1: Implications for Immunotherapy Treatment.* Front Immunol, 2020. **11**: p. 568931.
46. Gu, L., et al., *PD-L1 and gastric cancer prognosis: A systematic review and meta-analysis.* PLoS One, 2017. **12**(8): p. e0182692.
47. Muro, K., et al., *Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial.* Lancet Oncol, 2016. **17**(6): p. 717-726.
48. Curea, F.G., et al., *Current Targeted Therapies in HER2-Positive Gastric Adenocarcinoma.* Cancer Biother Radiopharm, 2017. **32**(10): p. 351-363.
49. Benne, R., et al., *Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA.* Cell, 1986. **46**(6): p. 819-26.
50. Christofi, T. and A. Zaravinos, *RNA editing in the forefront of epitranscriptomics and human health.* Journal of Translational Medicine, 2019. **17**(1): p. 319.
51. Gray, M.W., *Evolutionary origin of RNA editing.* Biochemistry, 2012. **51**(26): p. 5235-42.
52. Smith, H.C., et al., *Functions and regulation of the APOBEC family of proteins.* Semin Cell Dev Biol, 2012. **23**(3): p. 258-68.
53. Anant, S. and N.O. Davidson, *An AU-rich sequence element (UUUN[A/U]U) downstream of the edited C in apolipoprotein B mRNA is a high-affinity binding site for Apobec-1: binding of Apobec-1 to this motif in the 3' untranslated region of c-myc increases mRNA stability.* Mol Cell Biol, 2000. **20**(6): p. 1982-92.
54. Kinomoto, M., et al., *All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.* Nucleic Acids Res, 2007. **35**(9): p. 2955-64.
55. Bazak, L., et al., *A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes.* Genome Res, 2014. **24**(3): p. 365-76.
56. Daniel, C., J. Lagergren, and M. Öhman, *RNA editing of non-coding RNA and its role in gene regulation.* Biochimie, 2015. **117**: p. 22-7.
57. Gott, J.M. and R.B. Emeson, *Functions and mechanisms of RNA editing.* Annu Rev Genet, 2000. **34**: p. 499-531.
58. Nishikura, K., *A-to-I editing of coding and non-coding RNAs by ADARs.* Nature Reviews Molecular Cell Biology, 2016. **17**(2): p. 83-96.
59. Nishikura, K., et al., *Substrate specificity of the dsRNA unwinding/modifying activity.* EMBO J, 1991. **10**(11): p. 3523-32.
60. Lehmann, K.A. and B.L. Bass, *The importance of internal loops within RNA substrates of ADAR1.* J Mol Biol, 1999. **291**(1): p. 1-13.
61. Nishikura, K., *Functions and regulation of RNA editing by ADAR deaminases.* Annu Rev Biochem, 2010. **79**: p. 321-49.
62. Park, E., et al., *RNA editing in the human ENCODE RNA-seq data.* Genome Res, 2012. **22**(9): p. 1626-33.
63. Liu, Y. and C.E. Samuel, *Editing of glutamate receptor subunit B pre-mRNA by splice-site variants of interferon-inducible double-stranded RNA-specific adenosine deaminase ADAR1.* J Biol Chem, 1999. **274**(8): p. 5070-7.
64. Shen, H., et al., *ADARs act as potent regulators of circular transcriptome in cancer.* Nature Communications, 2022. **13**(1): p. 1508.
65. Wang, Isabel X., et al., *ADAR Regulates RNA Editing, Transcript Stability, and Gene Expression.* Cell Reports, 2013. **5**(3): p. 849-860.
66. Yang, W., et al., *Modulation of microRNA processing and expression through RNA editing by ADAR deaminases.* Nat Struct Mol Biol, 2006. **13**(1): p. 13-21.
67. Paz-Yaacov, N., et al., *Elevated RNA Editing Activity Is a Major Contributor to Transcriptomic Diversity in Tumors.* Cell Rep, 2015. **13**(2): p. 267-76.

68. Han, L., et al., *The Genomic Landscape and Clinical Relevance of A-to-I RNA Editing in Human Cancers.* Cancer Cell, 2015. **28**(4): p. 515-528.
69. Chan, T.H., et al., *ADAR-Mediated RNA Editing Predicts Progression and Prognosis of Gastric Cancer.* Gastroenterology, 2016. **151**(4): p. 637-650 e10.
70. Cho, C.J., et al., *Combinatory RNA-Sequencing Analyses Reveal a Dual Mode of Gene Regulation by ADAR1 in Gastric Cancer.* Digestive Diseases and Sciences, 2018. **63**(7): p. 1835-1850.
71. Okugawa, Y., et al., *Enhanced AZIN1 RNA editing and overexpression of its regulatory enzyme ADAR1 are important prognostic biomarkers in gastric cancer.* J Transl Med, 2018. **16**(1): p. 366.
72. An, O., et al., *"3G" Trial: An RNA Editing Signature to Guide Gastric Cancer Chemotherapy.* Cancer Res, 2021. **81**(10): p. 2788-2798.
73. Ghandi, M., et al., *Next-generation characterization of the Cancer Cell Line Encyclopedia.* Nature, 2019. **569**(7757): p. 503-508.
74. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* GigaScience, 2021. **10**(2).
75. Picardi, E. and G. Pesole, *REDItools: high-throughput RNA editing detection made easy.* Bioinformatics, 2013. **29**(14): p. 1813-1814.
76. Kiran, A.M., et al., *Darned in 2013: inclusion of model organisms and linking with Wikipedia.* Nucleic Acids Res, 2013. **41**(Database issue): p. D258-61.
77. Picardi, E., et al., *REDIportal: a comprehensive database of A-to-I RNA editing events in humans.* Nucleic Acids Res, 2017. **45**(D1): p. D750-D757.
78. Picardi, E., et al., *Using REDItools to Detect RNA Editing Events in NGS Datasets.* Current Protocols in Bioinformatics, 2015. **49**(1): p. 12.12.1-12.12.15.
79. Lo Giudice, C., et al., *Investigating RNA editing in deep transcriptome datasets with REDItools and REDIportal.* Nature Protocols, 2020. **15**(3): p. 1098-1131.
80. Bass, A.J., et al., *Comprehensive molecular characterization of gastric adenocarcinoma.* Nature, 2014. **513**(7517): p. 202-209.
81. Ramaswami, G., et al., *Identifying RNA editing sites using RNA sequencing data alone.* Nature Methods, 2013. **10**(2): p. 128-132.
82. Bendl, J., et al., *PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations.* PLOS Computational Biology, 2014. **10**(1): p. e1003440.
83. Szklarczyk, D., et al., *The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets.* Nucleic Acids Res, 2021. **49**(D1): p. D605-D612.
84. Chang, L., et al., *miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology.* Nucleic Acids Research, 2020. **48**(W1): p. W244-W251.
85. Baniak, N., et al., *Gastric biomarkers: a global review.* World Journal of Surgical Oncology, 2016. **14**(1): p. 212.
86. Ramaswami, G. and J.B. Li, *Identification of human RNA editing sites: A historical perspective.* Methods, 2016. **107**: p. 42-7.
87. Paz, N., et al., *Altered adenosine-to-inosine RNA editing in human cancer.* Genome Res, 2007. **17**(11): p. 1586-95.
88. Athanasiadis, A., A. Rich, and S. Maas, *Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome.* PLoS Biol, 2004. **2**(12): p. e391.
89. Jang, I.S., et al., *Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data.* Pac Symp Biocomput, 2014: p. 63-74.
90. Pozdeyev, N., et al., *Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies.* Oncotarget, 2016. **7**(32): p. 51619-51625.
91. Fallahi-Sichani, M., et al., *Metrics other than potency reveal systematic variation in responses to cancer drugs.* Nat Chem Biol, 2013. **9**(11): p. 708-14.

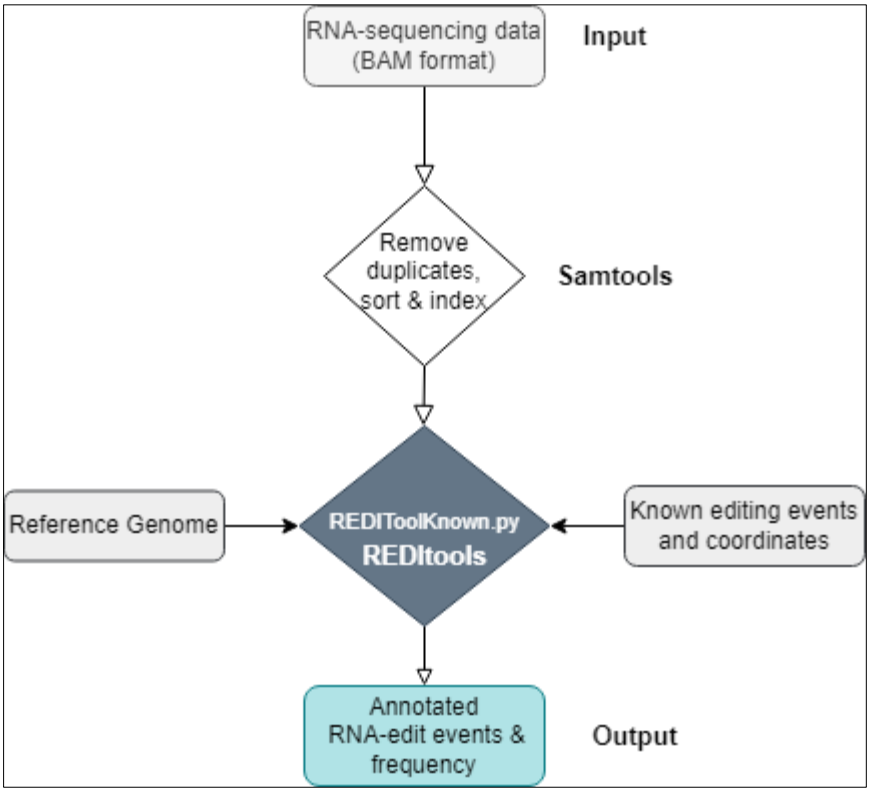
92. Shimada, K., et al., *Cell-Line Selectivity Improves the Predictive Power of Pharmacogenomic Analyses and Helps Identify NADPH as Biomarker for Ferroptosis Sensitivity.* Cell Chem Biol, 2016. **23**(2): p. 225-235.
93. Smirnov, P., et al., *PharmacoGx: an R package for analysis of large pharmacogenomic datasets.* Bioinformatics, 2015. **32**(8): p. 1244-1246.
94. Hong, H., et al., *Bidirectional regulation of adenosine-to-inosine (A-to-I) RNA editing by DEAH box helicase 9 (DHX9) in cancer.* Nucleic Acids Research, 2018. **46**(15): p. 7953-7969.
95. Tan, M.H., et al., *Dynamic landscape and regulation of RNA editing in mammals.* Nature, 2017. **550**(7675): p. 249-254.
96. Vislovukh, A., et al., *Role of 3'-untranslated region translational control in cancer development, diagnostics and treatment.* World J Biol Chem, 2014. **5**(1): p. 40-57.
97. Kawahara, Y., et al., *Redirection of silencing targets by adenosine-to-inosine editing of miRNAs.* Science, 2007. **315**(5815): p. 1137-40.
98. Yang, S., et al., *New insights into autophagy in hepatocellular carcinoma: mechanisms and therapeutic strategies.* Am J Cancer Res, 2019. **9**(7): p. 1329-1353.
99. Habieb, A., et al., *Potential role of lncRNA-TSIX, miR-548-a-3p, and SOGA1 mRNA in the diagnosis of hepatocellular carcinoma.* Molecular Biology Reports, 2019. **46**(4): p. 4581-4590.
100. Uhlen, M., et al., *A pathology atlas of the human cancer transcriptome.* Science, 2017. **357**(6352).
101. Shao, Y., et al., *Apoptotic and autophagic cell death induced by histone deacetylase inhibitors.* Proceedings of the National Academy of Sciences, 2004. **101**(52): p. 18030-18035.
102. Li, N., et al., *Identifying a cervical cancer survival signature based on mRNA expression and genome-wide copy number variations.* Exp Biol Med (Maywood), 2022. **247**(3): p. 207-220.
103. Li, X., et al., *ERRα activates SHMT2 transcription to enhance the resistance of breast cancer to lapatinib via modulating the mitochondrial metabolic adaption.* Biosci Rep, 2020. **40**(1).
104. Elhamamsy, A.R., et al., *Ribosome Biogenesis: A Central Player in Cancer Metastasis and Therapeutic Resistance.* Cancer Research, 2022. **82**(13): p. 2344-2353.
105. Kang, H., et al., *Post-Transcriptional Controls by Ribonucleoprotein Complexes in the Acquisition of Drug Resistance.* International Journal of Molecular Sciences, 2013. **14**(8): p. 17204-17220.
106. Cassandri, M., et al., *Zinc-finger proteins in health and disease.* Cell Death Discovery, 2017. **3**(1): p. 17071.
107. Liu, J., et al., *MiRNA-4537 functions as a tumor suppressor in gastric cancer and increases the radiosensitivity of gastric cancer cells.* Bioengineered, 2021. **12**(1): p. 8457-8467.
108. Olson, O.C. and J.A. Joyce, *Cysteine cathepsin proteases: regulators of cancer progression and therapeutic response.* Nat Rev Cancer, 2015. **15**(12): p. 712-29.
109. Yan, Y., et al., *Clinical significance of serum cathepsin B and cystatin C levels and their ratio in the prognosis of patients with esophageal cancer.* Onco Targets Ther, 2017. **10**: p. 1947-1954.
110. Ma, X., et al., *Genetic polymorphisms of Cathepsin B are associated with gastric cancer risk and prognosis in a Chinese population.* Cancer Biomark, 2021. **32**(2): p. 189-198.
111. Li, W., et al., *Long Non-Coding RNA NCK1-AS1 Serves an Oncogenic Role in Gastric Cancer by Regulating miR-137/NUP43 Axis.* Onco Targets Ther, 2020. **13**: p. 9929-9939.
112. Wang, Y., et al., *MiR-200c-3p aggravates gastric cell carcinoma via KLF6.* Genes Genomics, 2021. **43**(11): p. 1307-1316.

113. Knutson, S.D., et al., *Selective Enrichment of A-to-I Edited Transcripts from Cellular RNA Using Endonuclease V*. Journal of the American Chemical Society, 2020. **142**(11): p. 5241-5251.

# 7. Appendix

## 7.1 Appendix A

Schematic representation of REDItoolknown.py editing event identification pipeline




Summary table of the applied parameters in REDItools

| *Parameter* | *Applied setting* |
|---|---|
| *Base interval to explore* | 100000 |
| *Number of cores used for computation* | 4 |
| *Minimum read coverage* | 10 |
| *Minimum base quality* | 25 |
| *Minimum mapping quality* | 255 |
| *Minimum homopolymeric sequence length* | 5 |
| *Strand orientation* | 0-1 (both) |
| *Exclude multiple hits* | TRUE |
| *Exclude duplicated* | TRUE |
| *Consider mapping quality* | TRUE |
| *Number of bases to trim up and down the read* | 5-5 |
| *Exclude positions with multiple substitutions* | TRUE |
| *Number of positions near splice site to exclude* | 4 |

# 7.2 Appendix B

Gastric Cancer Cell Lines with available RNA-sequencing data

| Cell Line | Sex | Age at extraction | Source of cells | Morphology | Tissue of Extraction | Growth properties | Cancer Type | Ethnicity | Prior Treatment |
|---|---|---|---|---|---|---|---|---|---|
| 23132-87 | M | 72 | Primary | Intestinal | Unknown | Adherent | Gastric Adenocarcinoma | Unknown | Unknown |
| AGS | F | 54 | Primary | Diffuse | Stomach | Adherent | Gastric Adenocarcinoma | White | None |
| ECC10 | M | 73 | Primary | | Stomach | Suspension | Gastric Small Cell Neuroendocrine Carcinoma | East Asian | Unknown |
| ECC12 | M | 63 | | Epithelial-like | Stomach | Adherent | Gastric Small Cell Neuroendocrine Carcinoma | East Asian | Unknown |
| GCIY | F | 39 | Metastasis | Epithelial-like | Stomach | Adherent | Gastric Adenocarcinoma | East Asian | Unknown |
| HGC-27 | U | U | Metastasis | Epithelial-like | Lymph Node | Adherent | Gastric Carcinoma | East Asian | Unknown |
| KATOIII | M | 57 | Metastasis | Diffuse | Pleural effusion | Semi-Adherent | Signet Ring Cell Gastric Adenocarcinoma | Unknown | Unknown |
| MKN1 | M | 72 | Metastasis | Epithelial-like | Lymph Node | Adherent | Gastric Adenosquamous Carcinoma | East Asian | Unknown |
| MKN45 | F | 62 | Metastasis | Diffuse | Liver | Adherent | Gastric Adenocarcinoma | East Asian | Unknown |
| MKN7 | M | 39 | Metastasis | Intestinal | Stomach | Adherent | Gastric Tubular Adenocarcinoma | Unknown | Unknown |
| NCI-N87 | M | U | Metastasis | Tubular | Liver | Adherent | Gastric Tubular Adenocarcinoma | Unknown | Unknown |
| NCI-SNU-1 | M | 44 | Metastasis | Epithelial-like | Stomach | Suspension | Gastric Carcinoma | East Asian | None |
| NCI-SNU-16 | F | 33 | Metastasis | Epithelial-like | Ascites | Suspension | Gastric Carcinoma | East Asian | Unknown |
| NCI-SNU-5 | F | 33 | Metastasis | Epithelial-like | Ascites | Suspension | Gastric Carcinoma | East Asian | fluorouracil, doxorubicin and mitomycin C |
| NUGC-3 | M | 72 | Metastasis | Diffuse | Branchial Muscle | Adherent | Stomach | Unknown | Unknown |
| TGBC11TKB | F | 72 | Metastasis | Diffuse | Lymph Node | Adherent | Stomach | East Asian | Unknown |
| FU97 | F | 60 | Primary | Diffuse | Stomach | Adherent | Stomach | East Asian | Unknown |
| GSS | M | 75 | Metastasis | Indeterminate | Liver | Adherent | Stomach | East Asian | Unknown |
| GSU | M | 37 | Metastasis | Indeterminate | Liver | Adherent | Stomach | East Asian | Unknown |
| Hs 746T | M | 74 | Metastasis | Epithelial like | Left leg | Adherent | Gastric Adenocarcinoma | White | Unknown |
| HuG1-N | M | 55 | Metastasis | Lymphocyte-like | Ascites | Suspension | Gastric Tubular Adenocarcinoma | East Asian | Unknown |
| IM-95 | M | 63 | Tumour | Epithelial-like | Stomach | Unknown | Gastric Adenocarcinoma | East Asian | Unknown |
| KE-39 | M | 77 | Metastasis | Epithelial like | Lymph Node | Adherent | Gastric Adenocarcinoma | East Asian | Unknown |
| LMSU | F | 27 | Metastasis | Epithelial-like | Lymph Node | Adherent | Gastric Adenocarcinoma | East Asian | Unknown |
| MKN74 | M | 37 | Metastasis | Epithelial like | Liver | Adherent | Gastric Tubular Adenocarcinoma | East Asian | Unknown |
| NCC-StC-K140 | F | 36 | Tumour | Lymphoblast-like in | Stomach | Suspension | Gastric Adenocarcinoma | East Asian | Unknown |
| NUGC-2 | F | 56 | Metastasis | - | Lymph Node | Adherent | Gastric Adenocarcinoma | Unknown | Unknown |
| NUGC-4 | F | 35 | Metastasis | Spherical | Paragastric Lymph N | Unknown | Signet Ring Cell Gastric Adenocarcinoma | Unknown | Unknown |
| OCUM-1 | F | 38 | Tumour | Lymphocyte-like w | Unknown | Unknown | Signet Ring Cell Gastric Adenocarcinoma | East Asian | Unknown |
| RERF-GC-1B | F | 41 | Metastasis | Epithelial-like | Pyloric lymph nodes | Unknown | Gastric Adenocarcinoma | East Asian | Unknown |
| SH-10-TC | U | U | Unknown | Epithelial like | Unknown | Adherent | Gastric Adenocarcinoma | Asian | Unknown |
| SNU-216 | F | 46 | Metastasis | Epithelial like | Lymph Node | Adherent-monolayer | Gastric Adenocarcinoma | East Asian | Unknown |
| SNU-520 | F | 60 | Tumour | Floating aggregates | Stomach | Suspension | Gastric Adenocarcinoma | East Asian | None |
| SNU-601 | M | 34 | Metastasis | Epithelial | Ascites | Adherent-monolayer | Signet Ring Cell Gastric Adenocarcinoma | East Asian | Unknown |
| SNU-620 | F | 59 | Metastasis | Scattered floating | Ascites | Suspension | Gastric Adenocarcinoma | East Asian | None |
| SNU-668 | M | 63 | Metastasis | Epithelial | Ascites | Adherent-monolayer | Signet Ring Cell Gastric Adenocarcinoma | East Asian | Unknown |
| SNU-719 | M | 53 | Tumour | Epithelial | Stomach | Adherent-monolayer | Gastric Adenocarcinoma | East Asian | Unknown |

## 7.3 Appendix C

Table of editing events correlating with drug sensitivity identified in Ranksum analysis

| Drug | Gene | Genomic coordinates | Region | GDSC1 | GDSC2 | CTRP | CCLE |
|---|---|---|---|---|---|---|---|
| Axitinib | MAD2L1 | chr4:120979770 | 3UTR | 0.81 | | 0.64 | |
| | METTL2A | ch17:60527756 | downstream | -0.83 | | -0.57 | |
| | FPGS | chr9:130568717 | intronic | -0.70 | | -0.60 | |
| | DNAL1 | chr14:74165276 | 3UTR | 0.81* | | 0.49 | |
| | SYNRG | chr17:35877278 | 3UTR | 0.63 | | 0.62* | |
| | CENPU | chr4:185620850 | intronic | -0.69 | | -0.53 | |
| | VHL | chr3:10192617 | 3UTR | 0.63 | | 0.53 | |
| | TMEM44-AS1 | chr3:194309829 | ncRNA_exonic | -0.61 | | -0.56 | |
| | XIAP | chrX:123045142 | 3UTR | -0.52 | | -0.66** | |
| | NIF3L1 | chr2:201755743 | intronic | 0.63 | | 0.53 | |
| | TRIM56 | ch7:100736991 | 3UTR | 0.59 | | 0.47 | |
| | POLR1A | ch2:86249930 | 3UTR | 0.60 | | 0.42 | |
| | EIF2AK2 | chr2:37329238 | 3UTR | 0.54 | | 0.45 | |
| Cabozatinib | HP1BP3 | chr1:2108334 | intronic | -0.80 | | -0.71* | |
| | DCAF16 | chr4:17803681 | 3UTR | 0.84** | | 0.67** | |
| | LOC100129917 | chr4:774138 | ncRNA_exonic | 0.69 | | 0.73** | |
| | MRPL20 | chr1:1338166 | intronic | 0.79** | | 0.61* | |
| | GNPTG | chr16:1409326 | intronic | 0.79 | | 0.59 | |
| | MLEC | chr12:121128049 | intronic | 0.75* | | 0.60 | |
| | NUP43 | chr6:150046673 | 3UTR | 0.72 | | 0.64* | |
| Crizotinib | SCG2;AP1S3 | chr2:224618846 | intergenic | | 0.75 | 0.55 | -0.67 |
| | SH3BP2 | chr4:2841691 | 3UTR | | -0.80 | -0.59* | |
| | KAT8 | chr16:31134704 | intronic | | 0.71* | 0.59* | |
| | NDUFS1 | chr2:206986246 | 3UTR | | 0.88* | 0.52 | |
| | PTPN18 | chr2:131119805 | intronic | | -0.72 | -0.55 | |
| | GNPTG | chr16:1409326 | intronic | | 0.74 | 0.50 | |
| | ZNF320 | chr19:53383600 | 3UTR | | 0.85* | 0.44 | |
| | IPO5P1 | chr19:23441380 | ncRNA_exonic | | | 0.69 | -0.88* |
| | XPNPEP3 | chr22:41325329 | 3UTR | | | 0.75* | -0.79 |
| | NMNAT1 | chr1:10044367 | 3UTR | | | 0.62 | -0.79* |
| | USP1 | chr1:62915436 | intronic | | | 0.62* | -0.77 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | TRMT10B | chr9:37775695 | intronic | | | 0.71 | -0.73 |
| | SLC20A2 | chr8: 42388502 | intronic | | | 0.55 | -0.84 |
| | GOLGA3 | chr12:133347372 | 3UTR | | | 0.56 | -0.79 |
| | PGPEP1 | chr19:18478858 | 3UTR | | | -0.72** | 0.69 |
| | ZNF69 | chr19:12024893 | 3UTR | | | 0.54 | -0.84 |
| | MAPKAPK5 | chr12:112334978 | 3UTR | | 0.78* | | -0.83* |
| | PLEKHA2 | chr8:38828267 | 3UTR | | 0.88** | | -0.73 |
| | SLC12A9 | chr7:100452736 | intronic | | 0.79 | | -0.77 |
| | EIF3K | chr19:39121713 | intronic | | 0.80* | | -0.72* |
| | MPST | chr22:37424169 | intronic | | 0.77* | | -0.77* |
| | CZ1P-ASNS&CCZ1P-OR7E38P | Chr7:97598690 | ncRNA_intronic | | -0.87* | | 0.68 |
| | PAX-AS1 | Chr2:114010421 | ncRNA_intronic | | 0.67 | | -0.79 |
| | METTL16 | Chr17:2320645 | 3UTR | | 0.80* | | -0.68 |
| | APOL6 | Chr22:36056598 | 3UTR | | 0.66 | | -0.79* |
| | SLC25A25-AS1 | Chr9:130875311 | ncRNA_intronic | | 0.67 | | -0.77* |
| **Dasatanib** | DMTF1 | Ch7:86791607 | intronic | | 0.93*** | 0.83** | |
| | ZNF440 | Ch19:11945216 | 3UTR | | -0.78* | -0.63* | |
| | RAB11FIP4 | Chr17:29861470 | 3UTR | | -0.79* | -0.60 | |
| | CSAD | Chr12:53572673 | intronic | | 0.77* | 0.60 | |
| | DNAJC11 | Chr1:6707280 | intronic | | -0.90*** | -0.55 | |
| | LYRM7 | Chr5:130538072 | 3UTR | | 0.78* | 0.55 | |
| **Docetaxel** | LOC100132781;PTMAP11 | Chr9:99653927 | intergenic | 0.68 | 0.70 | 0.78 | |
| | ZNF432 | Chr19:52535209 | 3UTR | -0.86 | -0.96** | | |
| | FPGS | Chr9:130567644 | intronic | -0.81* | -0.70 | | |
| | ZNF688;ZNF785 | Chr16:30590009 | intergenic | 0.68 | 0.75* | | |
| | PHAX | chr5:125961447 | 3UTR | -0.77* | -0.67* | | |
| | LONP1 | Chr19:5710985 | intronic | -0.68* | -0.74* | | |
| | ZNF28 | Chr19:53301315 | 3UTR | -0.65 | -0.78* | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | PCSK7 | Chr11:117080491 | intronic | -0.67 | -0.74 | | |
| | C16orf72 | Chr16:9204883 | intronic | 0.73* | 0.65 | | |
| | FMC1-LUC7L2&LUC7L2 | Chr7:139055227 | intronic | 0.68 | | 0.83 | |
| | CCAR1 | Chr10:70518191 | intronic | 0.68* | | 0.76 | |
| | UGDH-AS1 | Chr4:39551131 | ncRNA_intronic | 0.81* | | 0.71 | |
| | LOC100129917 | Chr4:774074 | ncRNA_exonic | 0.62 | | 0.74 | |
| | FLX | chrX:73421663 | ncRNA_intronic | -0.60 | | -0.75 | |
| | FDPS | Chr1:155285386 | intronic | -0.61 | | -0.73 | |
| | EIF6 | Chr20:33871527 | intronic | -0.64 | | -0.66 | |
| | H2AZ2 | Chr7:44870547 | intronic | | 0.65* | 0.92** | |
| | COQ4 | Chr9:131093377 | intronic | | 0.72* | 0.72 | |
| | RILPL1 | Chr12:124007203 | intronic | | 0.65* | 0.75 | |
| | MYO1C | Chr17:1368283 | 3UTR | | -0.64* | -0.76 | |
| | CRCP;TPST1 | Chr7:65622507 | intergenic | | -0.59 | -0.83 | |
| **Erlotinib** | ESRP1 | Chr8:95688911 | intronic | | -0.68 | -0.66* | 0.73 |
| | EMC1-AS1 | Chr1:19543465 | ncRNA_intronic | | -0.85** | -0.60 | 0.58 |
| | SOGA1 | Chr20:35410515 | 3UTR | | -0.71 | -0.55 | 0.73 |
| | SLC25A29 | Chr14:100763074 | intronic | | -0.71* | -0.62* | |
| | MZF1-AS1 | Chr19:59078751 | ncRNA_intronic | | -0.71* | -0.59* | |
| | CCNYL1 | Chr2:208619377 | 3UTR | | -0.73* | -0.50 | |
| | TNFRSF10D;TNFRSF10A | Chr8:23045834 | intergenic | | 0.70 | | -0.73 |
| | KNL1;RAD51-AS1 | Chr15:40980776 | intergenic | | 0.63 | | -0.93** |
| | WDR12 | Chr2:203743850 | 3UTR | | 0.64 | | -0.72 |
| | TFDP2 | Chr3:141667800 | 3UTR | | -0.69* | | 0.66 |
| | FAM20B | Chr1:179044180 | 3UTR | | 0.68* | | -0.65 |
| | COPE | Chr19:19019763 | intronic | | 0.62 | | -0.72* |
| | GINS4&LOC102723729 | Chr8:41401304 | 3UTR ncRNA_intronic | | -0.56 | | 0.76* |
| | SMARCA4 | Chr19:11147075 | intronic | | | 0.79** | -0.87* |
| | B3GNT3;INSL3 | Chr19:17925387 | intergenic | | | -0.70* | 0.89 |
| | AAK1 | Chr2:69688991 | 3UTR | | | -0.72* | 0.82 |
| | GPN2 | Chr1:27205044 | 3UTR | | | -0.76** | 0.75 |
| | NUP43 | Chr6:150047362 | 3UTR | | | 0.58* | -0.81** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | PCGF3;LOC100129917 | Chr4:771590 | intergenic | | | -0.63* | 0.75 |
| | ADCY6 | Chr12:49180421 | intronic | | | -0.62 | 0.73 |
| | CENPN | Chr16:81063363 | 3UTR | | | 0.55 | -0.86** |
| | DMTF1 | Chr7:86791622 | intronic | | | 0.57 | -0.77 |
| **Fulvestrant** | RHNO1 | Chr12:2997876 | 3UTR | -0.73* | -0.82** | -0.74* | |
| | LANCL2;VOPP1 | Chr7:55522972 | intergenic | 0.75* | 0.80* | 0.72 | |
| | PTPN18 | Chr2:131119681 | intronic | 0.71 | 0.83* | 0.72 | |
| | CENPBD1P1 | Chr19:59093878 | ncRNA_exonic | -0.72* | -0.66* | -0.80* | |
| | PROSER3 | Chr19:36260415 | 3UTR | -0.80* | -0.70* | -0.65 | |
| | DDX52 | Chr17:35972246 | 3UTR | 0.67 | 0.92** | 0.67 | |
| | UBXN2A | Chr2:24223382 | 3UTR | -0.79* | -0.74* | 0.60 | |
| | RNF115 | Chr1:145689701 | 3UTR | -0.64 | -0.59 | -0.79* | |
| | DFFA | Chr1:10520715 | 3UTR | 0.71* | 0.53 | 0.71 | |
| | PRR11 | Chr17:57279077 | 3UTR | -0.74* | -0.53 | -0.69 | |
| | CCDC84 | Chr11:118875356 | intronic | -0.56 | -0.74* | -0.68 | |
| | SMYD4 | Chr17:1683647 | 3UTR | 0.67 | 0.61 | 0.70 | |
| | SLC12A9 | Chr7:100454274 | intronic | -0.54 | -0.67* | -0.64 | |
| | C19orf71 | Chr19:3542080 | intronic | 0.84* | 0.87** | | |
| | MED21;C12orf71 | Chr12:27188059 | intergenic | -0.81* | -0.92** | | |
| | HERPUD1 | Chr16:56974318 | intronic | 0.94* | 0.77 | | |
| | UGDH-AS1 | Chr4:39550558 | ncRNA_intronic | -0.81* | -0.83* | | |
| | DAP3 | Chr1:155705556 | intronic | -0.86* | -0.78* | | |
| **Fulvestrant** | RBSN | Chr3:15114400 | 3UTR | -0.87** | -0.74* | | |
| | FBXW8 | Chr12:117469471 | downstream | 0.82 | 0.75 | | |
| | RPUSD4 | Chr11:126079051 | intronic | -0.92** | -0.72 | | |
| | PVR | Chr19:5165576 | 3UTR | -0.77* | -0.78** | | |
| | LIMD-AS1 | Chr3:45721990 | ncRNA_intronic | -0.82** | | -0.86** | |
| | MIS18BP1 | Chr14:45676794 | intronic | -0.75 | | -0.82 | |
| | TRUB2 | Chr9:131069762 | 3UTR | 0.75 | | 0.82 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | NPLOC4 | Chr17:79530761 | intronic | -0.83** | | -0.72 | |
| | LOC150776 | Chr2:132271513 | ncRNA_intronic | -0.80* | | -0.75 | |
| | OGA | Chr10:103547635 | intronic | -0.80* | | -0.73 | |
| | FAM157C | Chr16:90235906 | ncRNA_intronic | -0.71 | | -0.81 | |
| | ANKRD10 | Chr13:111547773 | intronic | -0.69* | | -0.84* | |
| | ANKFY1 | Chr17:4071711 | intronic | | -0.80* | -0.85 | |
| | GOLGA2 | Chr9:131031509 | intronic | | -0.67 | -0.84 | |
| | MAVS | Chr20:3849026 | 3UTR | | 0.60 | 0.85* | |
| | ATAD3B | Chr1:1419700 | intronic | | -0.65 | -0.83 | |
| | RILPL1 | Chr12:124007367 | intronic | | 0.69 | 0.74 | |
| | SYAP1 | chrX:16781615 | 3UTR | | 0.67 | 0.77 | |
| **Gemcitabine** | UFD1 | Chr22:19438535 | intronic | -0.58 | -0.83** | -0.56 | |
| | POLR3E | Chr16:22341640 | intronic | 0.56 | 0.58 | 0.47 | |
| | TPD52L2 | Chr20:62505436 | intronic | 0.93* | 0.93* | | |
| | LOC101927668 MACC1 | Chr7:20177603 | ncRNA_intronic 3UTR | -0.81 | -0.92* | | |
| | BSDC1 | Chr1:32833539 | 3UTR | -0.74* | -0.80* | | |
| | C11orf54 | Chr11:93476376 | 5UTR | -0.82* | -0.75* | | |
| | TFDP2 | Chr3:141665627 | 3UTR | -0.81 | -0.74 | | |
| | ANKRD10 | Chr13:111551371 | intronic | 0.78* | 0.76* | | |
| | KLC1 | Chr14:104148970 | intronic | -0.72 | -0.77 | | |
| | HP1BP3 | Chr1:21083340 | intronic | 0.71 | 0.78 | | |
| | SOGA1 | Chr20:35409856 | 3UTR | | 0.86 | 0.72* | |
| | UBB;TRPV2 | Chr17:16287196 | intergenic | | -0.92* | -0.63 | |
| | PPIE | Chr1:40214148 | intronic | | 0.86 | 0.60 | |
| | ZNF789 | Ch7:99081043 | intronic | | -0.70 | -0.63 | |
| | ATAD3B | Chr1:1419625 | intronic | | -0.84* | -0.55 | |
| | PDK1 | Chr2:173461888 | 3UTR | | -0.84* | -0.54 | |
| | MATR3 | Chr5:138620183 | intronic | | -0.71* | -0.62* | |
| | MAFK | Chr7:1575029 | intronic | | -0.72 | -0.59 | |
| **Lapatinib** | RBBP9 | Chr20:18468511 | 3UTR | | 0.60 | 0.56* | -0.55 |
| | C11orf80 | Chr11:66523904 | CDS – exonic S>>G | | 0.59 | 0.49 | -0.62 |
| | YIPF4 | Chr2:32535746 | 3UTR | | 0.52 | 0.43 | -0.61 |
| | SNRPD3 | Chr22:24969061 | 3UTR | | 0.66 | 0.70** | |
| | DARS2 | Chr:173828457 | downstream | | 0.70* | 0.63* | |
| | NDUFB7 | Chr19:14680184 | intronic | | -0.72 | -0.56 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | GP6;RDH13 | Chr19:55551722 | intergenic | | 0.74 | 0.56 | |
| | POLG2 | Chr17:62485002 | intronic | | -0.61 | -0.63 | |
| | EXOSC2 | Chr9:133575155 | intronic | | 0.62 | 0.57* | |
| | PCGF3;LOC100129917 | Chr4:772527 | intergenic | | -0.82** | -0.48 | |
| | ZNF440 | Ch19:11945216 | 3UTR | | -0.71* | -0.49 | |
| | CCDC84 | Chr11:118871252 | intronic | | -0.82 | | 0.83 |
| | ZNF417 | Chr19:58416290 | downstream | | -0.86 | | 0.81 |
| | TEP1 | Chr14:20835186 | 3UTR | | -0.69 | | 0.86 |
| | MRPL48;COA4 | Chr11:73582606 | intergenic | | 0.78 | | -0.74 |
| | MREG | Chr2:216808974 | 3UTR | | 0.78 | | -0.72 |
| | RPL36A&RPL36A-HNRNPH2 | ChrX:100648057 | intronic | | 0.61 | | -0.81* |
| | **SPAG9** | Chr17:49042252 | 3UTR | | -0.68 | | 0.72 |
| | MRPL30 | Chr2:99812464 | 3UTR | | -0.69* | | 0.60 |
| | XPNPEP3 | Ch22:41327597 | 3UTR | | -0.57 | | 0.70 |
| | SMARCA4 | Chr19:11147075 | intronic | | | 0.78* | -0.85* |
| | CCDC84 | Chr11:118876710 | intronic | | | 0.63 | -0.88* |
| | ICOSLG | Chr21:45644164 | 3UTR | | | 0.64* | -0.82* |
| | SPATA5 | Chr4:124237187 | 3UTR | | | 0.65 | -0.77 |
| | NMRAL1 | Chr16:4512245 | intronic | | | -0.70 | 0.76 |
| | METTL2B;LINC01000 | Chr7:128145040 | intergenic | | | 0.73 | -0.75 |
| | CYCS | Chr7:25160735 | 3UTR | | | 0.58 | -0.89* |
| | DNAJC24 | Chr11:31452750 | 3UTR | | | 0.69* | -0.74 |
| | UGDH-AS1 | Chr4:39552062 | ncRNA_intronic | | | 0.58* | -0.80* |
| | PLEKHA2 | Chr8:38829754 | 3UTR | | | 0.66* | -0.74* |
| **Methotrexate** | TARDBP | Chr1:11081296 | intronic | 0.81* | | 0.59 | |
| | H2AFV | Chr7:44873177 | intronic | 0.74* | | 0.64* | |
| | CCDC84 | Chr11:118870684 | intronic | 0.71 | | 0.66* | |
| | SAMD5 | Ch6:147886688 | 3UTR | -0.74 | | -0.64 | |
| | EEF2K | Chr16:22297236 | 3UTR | 0.76* | | 0.54* | |
| | CDK9;FPGS | Chr9:130554115 | intergenic | 0.68 | | 0.62* | |
| | MDM4 | Chr1:204521634 | 3UTR | 0.87* | | 0.52 | |
| | HSPB11 | Ch1:54387902 | intronic | 0.75 | | 0.54 | |
| | EBP | chrX:48381084 | intronic | 0.63 | | 0.69* | |
| **Nilotinib** | DNAJC24 | Chr11:31452750 | 3UTR | 0.73 | | 0.56 | -0.90* |
| | FDPS | Chr1: 155282727 | intronic | 0.71* | | 0.63* | -0.65 |
| | LOC100129917 | Chr4: 774115 | ncRNA_exonic | 0.67* | | 0.50 | -0.72 |
| | GATD1 | Chr11: 768289 | 3UTR | 0.59 | | 0.79** | -0.65 |
| | BPNT1 | Chr1: 220231263 | 3UTR | 0.60 | | 0.56* | -0.63 |
| | PHACTR4 | Chr1: 28825505 | 3UTR | 0.55 | | 0.47 | -0.80* |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | H2AFV | Chr7: 44873098 | intronic | 0.79** | | 0.45 | -0.64 |
| | H2AFV | Chr7: 44872732 | intronic | 0.76** | | 0.46 | -0.62 |
| | GSR | Chr8: 30536544 | 3UTR | 0.56 | | 0.47 | -0.60 |
| | RNASEH1 | Chr2: 3589521 | 3UTR | 0.70* | 0.70* | 0.70** | |
| | THEM4 | Chr1: 151847110 | 3UTR | 0.72* | 0.58 | 0.73** | |
| | DNAJC22 | Chr12: 49745536 | 3UTR | -0.66 | -0.71* | -0.60 | |
| | MBD3 | Chr19: 1589247 | intronic | 0.68* | 0.60 | 0.65* | |
| | DFFA | Chr1: 10520725 | 3UTR | 0.53 | 0.66* | 0.70** | |
| | VHL | Chr3: 10192753 | 3UTR | 0.65 | 0.58 | 0.68* | |
| | ANAPC16 | Chr10: 73979859 | 5UTR | 0.67* | 0.72* | 0.45 | |
| | ARPIN | Chr15: 90442606 | 3UTR | 0.69* | 0.53 | 0.54 | |
| | TCP11L1;LINC00294 | Chr11: 33096285 | intergenic | 0.63 | 0.57 | 0.52 | |
| | VHL | Chr3: 10194651 | 3UTR | 0.72* | 0.53 | 0.45 | |
| | SNAP23 | Chr15: 42824369 | 3UTR | 0.59 | 0.64 | 0.45 | |
| | CRCP;TPST1 | Chr7: 65620798 | intergenic | 0.54 | 0.55 | 0.51 | |
| **Olaparib** | KXD1 | chr19:18669846 | Intronic | -0.81* | -0.81* | -0.63 | |
| | METTL2A | chr17:60527756 | Downstream | -0.79 | -0.90* | -0.60 | |
| | NMRAL1 | chr16:4512810 | Intronic | 0.71 | 0.74 | 0.71 | |
| | ZNF655 | chr7:99164430 | Intronic | 0.87* | 0.68 | 0.63 | |
| | HOOK3 | chr8:42878293 | 3UTR | 0.72 | 0.72 | 0.61 | |
| | XIAP | chrX:123045142 | 3UTR | -0.61 | -0.67* | -0.68** | |
| | GP6;RDH13 | chr19:55551755 | Intergenic | 0.71* | 0.70 | 0.54 | |
| | KAT8 | chr16:31133346 | Intronic | 0.59 | 0.73* | 0.58* | |
| | GTF2IRD2;STAG3L2 | chr7:74294767 | Intergenic | 0.74* | 0.69* | 0.49 | |
| | ZNHIT6 | chr1:86119219 | 3UTR | 0.68* | 0.56 | 0.64* | |
| | GSEC | chr11:126216595 | ncRNA_intronic | 0.66 | 0.66 | 0.57 | |
| | FKBP3 | chr14:45600669 | Intronic | -0.59 | -0.70 | -0.56 | |
| | METTL2A | chr17:60528004 | Downstream | 0.61 | 0.56 | 0.64 | |
| | ARPIN | chr15:90440660 | 3UTR | 0.69* | 0.61 | 0.55* | |
| | PHACTR4 | chr1:28825530 | 3UTR | 0.57 | 0.61 | 0.63* | |
| | NDUFS1 | chr2:206987096 | 3UTR | 0.58 | 0.59 | 0.58* | |
| | FCF1 | chr14:75202669 | 3UTR | 0.63 | 0.62 | 0.45 | |
| | NFATC2IP | chr16:28969640 | Intronic | -0.54 | -0.58 | -0.60* | |
| | ZNF431 | chr19:21366997 | 3UTR | 0.58 | 0.67 | 0.45 | |
| | SKA1 | chr18:47918813 | 3UTR | 0.60 | 0.59 | 0.48 | |
| | TAF1B;GRHL1 | chr2:10076828 | Intergenic | 0.57 | 0.54 | 0.49 | |
| | H2AFV | chr7:44870759 | Intronic | 0.56 | 0.54 | 0.44 | |
| **Paclitaxel** | LYRM7 | chr5:130539816 | 3UTR | | 0.71 | | -0.83* |
| | C11orf58 | chr11:16778078 | 3UTR | | 0.94** | | -0.70 |
| | EIF6 | chr20:33871527 | Intronic | | -0.65 | | 0.72 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | SYNRG | chr17:35877278 | 3UTR | | 0.63 | | -0.71* |
| | LRRC37A4P | chr17:43583373 | ncRNA_exonic | | 0.63 | | -0.71 |
| | TRAP1 | chr16:3729134 | Intronic | | | 0.72* | -0.84* |
| | GTF3C2 | chr2:27580312 | Upstream | | | 0.73** | -0.75* |
| | ICOSLG | chr21:45644562 | 3UTR | | | 0.61 | -0.92* |
| | AHR | chr7:17384437 | 3UTR | | | 0.62* | -0.82* |
| | MRPL20 | chr1:1338720 | Intronic | | | 0.71* | -0.74 |
| | SLC20A2 | chr8:42388502 | Intronic | | | 0.59 | -0.90 |
| | PPIA | chr7:44841388 | 3UTR | | 0.53 | 0.47 | -0.72* |
| **Sorafenib** | GRSF1 | chr4:71683024 | 3UTR | | 0.87* | 0.74* | -0.83 |
| | METTL2B | chr7:128143932 | Downstream | | 0.56 | 0.62* | -0.85** |
| | PSMA5 | chr1:109952010 | Intronic | | 0.57 | 0.65* | -0.67 |
| | PGPEP1 | chr19:18476294 | 3UTR | | 0.71 | 0.47 | -0.75 |
| | CBFA2T2 | chr20:32234068 | 3UTR | | 0.61 | 0.48 | -0.76* |
| | SPCS3 | chr4:177252327 | 3UTR | | 0.57 | 0.61* | -0.69* |
| | MTDH | chr8:98739187 | 3UTR | | 0.53 | 0.57* | -0.73* |
| | TIMM29 | chr19:11041295 | Downstream | | 0.71* | 0.55 | -0.64 |
| | POLR1E;FBXO10 | chr9:37505692 | Intergenic | | 0.62 | 0.50 | -0.69* |
| | GINS1 | chr20:25428294 | 3UTR | | 0.63 | 0.50 | -0.64 |
| | PGPEP1 | chr19:18479129 | 3UTR | | 0.59 | 0.50 | -0.68 |
| | MRI1 | chr19:13884176 | 3UTR | | 0.55 | 0.49 | -0.70 |
| | ANAPC16 | chr10:73995138 | 3UTR | | 0.72* | 0.48 | -0.63 |
| | GNPNAT1 | chr14:53243438 | 3UTR | | 0.52 | 0.48 | -0.75* |
| | GSR | chr8:30536152 | 3UTR | | 0.54 | 0.65* | -0.56 |
| | SLC35A3 | chr1:100489835 | 3UTR | | 0.52 | 0.64* | -0.64 |
| | DNAJC9-AS1 | chr10:75009591 | ncRNA_intronic | | 0.53 | 0.55* | -0.60 |
| | SPCS3 | chr4:177252238 | 3UTR | | 0.58 | 0.46 | -0.57 |
| | SERBP1 | chr1:67874849 | 3UTR | | 0.52 | 0.46 | -0.58 |
| **Tamoxifen** | CRCP | Chr7: 65619708 | downstream | -0.76* | -0.63 | -0.66* | |
| | RPP38 | Chr10: 15144305 | 5UTR | -0.73* | -0.67* | -0.44 | |
| | RBX1 | Chr22: 41351338 | intronic | -0.66 | -0.58 | -0.64* | |
| | MAGT1 | chrX: 77083116 | 3UTR | 0.67 | 0.54 | 0.60* | |
| | TMEM170A | Chr16: 75479831 | 3UTR | -0.61 | -0.55 | -0.49 | |
| | PGPEP1 | Chr19: 18477446 | 3UTR | -0.88* | -0.90** | | |
| | NMRAL1 | Chr16: 4512262 | intronic | -0.90 | -0.79 | | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **P3H4** | Chr17: 39964920 | intronic | -0.89* | -0.81 | | |
| | SRP9 | Chr1:225976243 | intronic | -0.83 | -0.83 | | |
| | ACOT9 | chrX:23721568 | 3UTR | -0.81 | -0.88** | | |
| | CENPN | Chr16:81063389 | 3UTR | -0.79* | -0.96** | | |
| | CARD8-AS1 | Chr19:48759084 | ncRNA_intronic | -0.82* | -0.76* | | |
| | FUT11;CHCHD1 | Chr10:75537390 | intergenic | 0.82* | 0.74* | | |
| | RSBN1L | Chr7:77411181 | 3UTR | -0.82 | -0.74 | | |
| | DAP3 | Chr1:155705556 | intronic | -0.80 | -0.75 | | |
| | PLCXD1 | chrX:218554 | 3UTR | | -0.68 | -0.61* | |
| | AJUBA | Chr14:23442149 | 3UTR | | -0.71* | -0.50 | |
| | PGAM5 | Chr12:133297956 | 3UTR | | -0.68* | -0.50 | |
| | PLEKHA2 | chr8:38828267 | 3UTR | | -0.67 | -0.52 | |
| | PRKAR2A | Chr3:48787640 | 3UTR | | -0.65* | -0.58* | |
| | PUS1 | Chr12:132422733 | intronic | | 0.58 | 0.72** | |
| | ANAPC5 | Chr12:121772075 | intronic | -0.87* | | -0.63 | |
| | MREG | Chr2:216808321 | 3UTR | 0.70 | | 0.70* | |
| | FPGS | chr9:130568717 | intronic | -0.81 | | -0.60 | |
| | ZNF839 | Chr14:102789051 | intronic | 0.79 | | 0.55 | |
| | NDUFB7 | Chr19:14680184 | intronic | -0.76 | | -0.57 | |
| **Temsirolimus** | ANKFY1 | Chr17:4071711 | intronic | -0.87 | | -0.79 | |
| | LRP5 | Chr11:68208748 | intronic | -0.93* | | -0.75 | |
| | ZNF526 | Chr19:42732933 | downstream | -0.82 | | -0.80* | |
| | SPAG9 | Chr17:49041827 | 3UTR | -0.76 | | -0.83 | |
| | SLC20A1 | Chr2:113412798 | intronic | -0.84* | | -0.75 | |
| | CCDC84 | Chr11:118876748 | intronic | -0.85 | | -0.73 | |
| | HERPUD1 | Chr16:56974318 | intronic | 0.84 | | 0.72 | |
| **Vorinostat** | ABHD14B | chr3:52007637 | Intronic | 0.73 | 0.64 | 0.71* | |
| | ZNF431 | chr19:21367102 | 3UTR | -0.60 | -0.70* | -0.55 | |
| | LOC441242 | chr7:65205946 | ncRNA_intronic | -0.59 | -0.63 | -0.50 | |
| | AMZ1 | chr7:2762018 | Intronic | 0.86 | | 0.85** | |
| | RAD51-AS1 | chr15:40985066 | Downstream | 0.86 | | 0.67* | |
| | SOGA1 | chr20:35411013 | 3UTR | 0.78* | | 0.73* | |
| | GSEC | chr11:126219036 | ncRNA_intronic | 0.79 | | 0.66 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | TCTA | chr3:49452695 | 3UTR | -0.76** | | -0.66* | |
| | MIS18BP1 | chr14:45676771 | Intronic | -0.70 | | -0.62 | |
| | ARHGAP27P1-BPTFP1-KPNA2P3;PLEKHM1P1 | chr17:62779275 | Intergenic | 0.73 | | 0.60 | |
| | HELLS;CYP2C18 | chr10:96367532 | Intergenic | -0.68* | | -0.58 | |
| | HEATR3 | chr16:50139556 | 3UTR | -0.61 | | -0.62* | |
| | GNL3L | chrX:54587219 | 3UTR | | -0.77 | -0.75** | |
| | ZNF875 | chr19:37817994 | Intronic | | 0.78 | 0.62 | |
| | CXorf56 | chrX:118672525 | 3UTR | | 0.85 | 0.59 | |
| | RASSF1 | chr3:50370026 | Intronic | | -0.83* | -0.60 | |
| | SVIL-AS1 | chr10:29748923 | ncRNA_intronic | | -0.65 | -0.82** | |
| | TFDP2 | chr3:141665627 | 3UTR | | 0.71 | 0.60 | |
| | LRP5 | chr11:68208679 | Intronic | | -0.73 | -0.58 | |
| | MVP17 | Ch2: 27538566 | intronic | 0.82 | 0.79 | | |
| | TRUB2 | Chr9: 131069762 | 3UTR | 0.72 | 0.70 | | |
| | MAVS | Chr20: 3849160 | 3UTR | 0.80* | 0.67 | | |
| | WASHC2C | Chr10: 46223730 | intronic | -0.62 | -0.71 | | |
| | NMNAT1 | Chr1: 10043079 | 3UTR | -0.64 | -0.69 | | |
| | GSEC | Chr11: 126216838 | ncRNA_intronic | -0.71 | -0.64 | | |
| | LINC01578 | Chr15: 93433986 | ncRNA_intronic | -0.67 | -0.66 | | |
| | EXOSC2;ABL1 | Chr9:133581385 | intergenic | -0.74* | -0.63 | | |

## 7.4 Appendix D

Correlation of editing events with ADARB1 (ADAR2) expression levels

| **Editing** | **Spearman Rho** | **p-value** | **Adjusted p.val** |
|---|---|---|---|
| chr20\|3849026\|transcript,exon,3UTR\|MAVS | 0.29 | 0.16 | 0.77 |
| chr20\|3849160\|transcript,exon,3UTR\|MAVS | 0.02 | 0.93 | 0.98 |
| chr20\|18468511\|3UTR,exon,transcript\|RBBP9 | 0.00 | 0.99 | 1.00 |
| chr20\|25428294\|transcript,exon,3UTR\|GINS1 | -0.11 | 0.60 | 0.98 |
| chr20\|32234068\|transcript,exon,3UTR\|CBFA2T2 | 0.02 | 0.93 | 0.98 |
| chr20\|33871527\|transcript\|EIF6 | -0.35 | 0.13 | 0.76 |
| chr20\|35409856\|transcript,3UTR,exon\|SOGA1 | -0.17 | 0.56 | 0.98 |
| chr20\|35410515\|transcript,3UTR,exon\|SOGA1 | 0.16 | 0.55 | 0.98 |
| chr20\|35411013\|transcript,3UTR,exon\|SOGA1 | 0.40 | 0.09 | 0.76 |
| chr22\|19438535\|transcript\|UFD1 | -0.11 | 0.62 | 0.98 |
| chr22\|24969061\|transcript,exon,3UTR\|SNRPD3 | 0.17 | 0.42 | 0.97 |
| chr22\|36056598\|transcript,exon,3UTR\|APOL6 | 0.48 | 0.02 | 0.63 |
| chr22\|37424169\|transcript\|MPST | 0.27 | 0.20 | 0.82 |
| chr22\|41325329\|transcript,exon,3UTR\|XPNPEP3 | 0.07 | 0.81 | 0.98 |
| chr22\|41351338\|transcript\|RBX1 | -0.01 | 0.95 | 0.98 |
| chr1\|1338166\|transcript\|MRPL20 | 0.49 | 0.02 | 0.63 |
| chr1\|1338720\|transcript\|MRPL20 | 0.35 | 0.13 | 0.76 |
| chr1\|1419625\|transcript\|ATAD3B | -0.12 | 0.59 | 0.98 |
| chr1\|1419700\|transcript\|ATAD3B | 0.14 | 0.63 | 0.98 |
| chr1\|10043079\|transcript,exon,3UTR\|NMNAT1 | 0.09 | 0.71 | 0.98 |
| chr1\|10044367\|transcript,exon,3UTR\|NMNAT1 | 0.37 | 0.12 | 0.76 |
| chr1\|10520715\|3UTR,exon,transcript\|DFFA | 0.05 | 0.81 | 0.98 |
| chr1\|10520725\|3UTR,exon,transcript\|DFFA | 0.46 | 0.02 | 0.63 |
| chr1\|11081296\|transcript\|TARDBP | -0.14 | 0.58 | 0.98 |
| chr1\|19543465\|transcript&3UTR,exon,transcript\|RP1-43E13.2&EMC1 | 0.06 | 0.78 | 0.98 |
| chr1\|21083340\|transcript\|HP1BP3 | 0.10 | 0.71 | 0.98 |
| chr1\|21083340\|transcript\|HP1BP3.1 | 0.10 | 0.71 | 0.98 |
| chr1\|28825505\|transcript,exon,3UTR\|PHACTR4 | 0.29 | 0.16 | 0.77 |
| chr1\|28825530\|transcript,exon,3UTR\|PHACTR4 | 0.05 | 0.82 | 0.98 |
| chr1\|32833539\|3UTR,exon,transcript\|BSDC1 | 0.52 | 0.02 | 0.63 |

| | | | |
|---|---|---|---|
| chr1\|40214148\|transcript\|PPIE | 0.09 | 0.76 | 0.98 |
| chr1\|54387902\|transcript\|HSPB11 | -0.16 | 0.51 | 0.98 |
| chr1\|62915436\|transcript\|USP1 | 0.36 | 0.13 | 0.76 |
| chr1\|67874849\|3UTR,exon,transcript\|SERBP1 | 0.11 | 0.58 | 0.98 |
| chr1\|86119219\|3UTR,exon,transcript\|ZNHIT6 | 0.02 | 0.92 | 0.98 |
| chr1\|100489835\|transcript,exon,3UTR\|SLC35A3 | 0.06 | 0.78 | 0.98 |
| chr1\|109952010\|transcript\|PSMA5 | 0.25 | 0.28 | 0.89 |
| chr1\|145689701\|transcript,exon,3UTR$transcript$transcript$transcript\|RNF115$NBPF19$NBPF20$NBPF10 | -0.12 | 0.57 | 0.98 |
| chr1\|151847110\|3UTR,exon,transcript\|THEM4 | 0.47 | 0.02 | 0.63 |
| chr1\|155282727\|transcript\|FDPS | 0.15 | 0.52 | 0.98 |
| chr1\|155285386\|transcript\|FDPS | -0.01 | 0.95 | 0.98 |
| chr1\|179044180\|transcript,exon,3UTR\|FAM20B | -0.06 | 0.76 | 0.98 |
| chr1\|220231263\|3UTR,exon,transcript\|BPNT1 | 0.38 | 0.05 | 0.70 |
| chr3\|10192617\|transcript,3UTR,exon\|VHL | 0.11 | 0.61 | 0.98 |
| chr3\|10192753\|transcript,3UTR,exon\|VHL | 0.00 | 0.99 | 1.00 |
| chr3\|10194651\|transcript,3UTR,exon\|VHL | -0.11 | 0.58 | 0.98 |
| chr3\|15114400\|3UTR,exon,transcript\|RBSN | 0.08 | 0.71 | 0.98 |
| chr3\|45721990\|transcript,exon,3UTR&transcript\|LIMD1&LIMD1-AS1 | 0.11 | 0.60 | 0.98 |
| chr3\|48787640\|3UTR,exon,transcript\|PRKAR2A | -0.17 | 0.41 | 0.97 |
| chr3\|49452695\|transcript,exon,3UTR\|TCTA | -0.22 | 0.28 | 0.89 |
| chr3\|50370026\|transcript\|RASSF1 | -0.30 | 0.27 | 0.89 |
| chr3\|141667800\|3UTR,exon,transcript\|TFDP2 | 0.56 | 0.01 | 0.63 |
| chr2\|3589521\|transcript,3UTR,exon\|RNASEH1 | 0.34 | 0.09 | 0.76 |
| chr2\|10076828\|-\|- | 0.40 | 0.07 | 0.74 |
| chr2\|24223382\|transcript,exon,3UTR\|UBXN2A | 0.04 | 0.83 | 0.98 |
| chr2\|27538566\|transcript\|MPV17 | -0.06 | 0.84 | 0.98 |
| chr2\|27580312\|-\|- | 0.40 | 0.04 | 0.66 |
| chr2\|32535746\|transcript,exon,3UTR\|YIPF4 | 0.29 | 0.15 | 0.76 |
| chr2\|37329238\|3UTR,exon,transcript\|EIF2AK2 | 0.14 | 0.49 | 0.98 |
| chr2\|69688991\|3UTR,exon,transcript\|AAK1 | 0.04 | 0.89 | 0.98 |
| chr2\|86249930\|3UTR,exon,transcript\|POLR1A | -0.13 | 0.54 | 0.98 |
| chr2\|99812464\|transcript,exon,3UTR\|MRPL30 | -0.16 | 0.44 | 0.97 |
| chr2\|113412798\|transcript\|SLC20A1 | -0.09 | 0.71 | 0.98 |
| chr2\|114010421\|transcript&transcript\|PAX8-AS1&PAX8 | 0.21 | 0.43 | 0.97 |

| | | | |
|---|---|---|---|
| chr2\|132271513\|transcript\|LOC150776 | -0.23 | 0.35 | 0.92 |
| chr2\|201755743\|transcript\|NIF3L1 | -0.13 | 0.55 | 0.98 |
| chr2\|208619377\|transcript,exon,3UTR\|CCNYL1 | 0.22 | 0.28 | 0.89 |
| chr5\|125961447\|transcript,exon,3UTR\|PHAX | 0.20 | 0.36 | 0.93 |
| chr5\|130538072\|transcript,exon,3UTR\|LYRM7 | -0.09 | 0.71 | 0.98 |
| chr5\|130539816\|transcript,exon,3UTR\|LYRM7 | 0.38 | 0.12 | 0.76 |
| chr5\|138620183\|transcript\|MATR3 | 0.32 | 0.14 | 0.76 |
| chr4\|771590\|-\|- | -0.32 | 0.15 | 0.76 |
| chr4\|772527\|-\|- | -0.18 | 0.40 | 0.97 |
| chr4\|774074\|exon,transcript\|RP11-440L14.1 | -0.09 | 0.67 | 0.98 |
| chr4\|774115\|exon,transcript\|LOC100129917 | 0.08 | 0.70 | 0.98 |
| chr4\|774138\|exon,transcript\|LOC100129917 | 0.31 | 0.13 | 0.76 |
| chr4\|17803681\|transcript,3UTR,exon\|DCAF16 | 0.24 | 0.23 | 0.85 |
| chr4\|39550558\|transcript&3UTR,exon,transcript\|UGDH-AS1&SMIM14 | 0.04 | 0.87 | 0.98 |
| chr4\|39551131\|transcript&3UTR,exon,transcript\|UGDH-AS1&SMIM14 | -0.03 | 0.91 | 0.98 |
| chr4\|39552062\|transcript&3UTR,exon,transcript\|UGDH-AS1&SMIM14 | 0.31 | 0.14 | 0.76 |
| chr4\|124237187\|transcript,exon,3UTR\|SPATA5 | -0.17 | 0.55 | 0.98 |
| chr4\|177252238\|transcript,exon,3UTR\|SPCS3 | 0.07 | 0.73 | 0.98 |
| chr4\|177252327\|transcript,exon,3UTR\|SPCS3 | 0.09 | 0.67 | 0.98 |
| chr6\|150046673\|3UTR,exon,transcript\|NUP43 | -0.13 | 0.56 | 0.98 |
| chr6\|150047362\|3UTR,exon,transcript\|NUP43 | -0.11 | 0.58 | 0.98 |
| chr8\|23045834\|-\|- | 0.16 | 0.50 | 0.98 |
| chr8\|30536152\|3UTR,exon,transcript\|GSR | 0.27 | 0.19 | 0.80 |
| chr8\|30536544\|3UTR,exon,transcript\|GSR | 0.41 | 0.04 | 0.66 |
| chr8\|38828267\|transcript,exon,3UTR\|PLEKHA2 | 0.08 | 0.72 | 0.98 |
| chr8\|38829754\|transcript,exon,3UTR\|PLEKHA2 | 0.51 | 0.02 | 0.63 |
| chr8\|41401304\|transcript,exon,3UTR&transcript\|GINS4&LOC102723729 | 0.26 | 0.24 | 0.85 |
| chr8\|42388502\|transcript\|SLC20A2 | 0.43 | 0.09 | 0.76 |
| chr8\|42878293\|transcript,exon,3UTR\|HOOK3 | 0.24 | 0.37 | 0.95 |
| chr8\|98739187\|transcript,exon,3UTR\|MTDH | 0.12 | 0.56 | 0.98 |
| chr9\|37505692\|-\|- | -0.22 | 0.29 | 0.89 |
| chr9\|99653927\|-\|- | -0.20 | 0.42 | 0.97 |

| | | | |
|---|---|---|---|
| chr9\|130554115\|-\|- | 0.09 | 0.70 | 0.98 |
| chr9\|130567644\|transcript\|FPGS | 0.53 | 0.06 | 0.74 |
| chr9\|130568717\|transcript\|FPGS | 0.03 | 0.91 | 0.98 |
| chr9\|131031509\|transcript\|GOLGA2 | -0.14 | 0.60 | 0.98 |
| chr9\|131069762\|3UTR,exon,transcript\|TRUB2 | 0.02 | 0.94 | 0.98 |
| chr9\|131093377\|transcript\|COQ4 | -0.16 | 0.44 | 0.97 |
| chr9\|133575155\|transcript\|EXOSC2 | 0.07 | 0.74 | 0.98 |
| chr9\|133581385\|-\|- | 0.00 | 1.00 | 1.00 |
| chr7\|1575029\|transcript\|MAFK | 0.03 | 0.93 | 0.98 |
| chr7\|44841388\|transcript,exon,3UTR\|PPIA | 0.11 | 0.59 | 0.98 |
| chr7\|44870547\|transcript\|H2AFV | -0.15 | 0.45 | 0.97 |
| chr7\|44870759\|transcript\|H2AFV | 0.12 | 0.56 | 0.98 |
| chr7\|44872732\|transcript\|H2AFV | 0.27 | 0.18 | 0.80 |
| chr7\|44873098\|transcript\|H2AFV | 0.27 | 0.18 | 0.80 |
| chr7\|44873177\|transcript\|H2AFV | 0.08 | 0.69 | 0.98 |
| chr7\|55522972\|-\|- | 0.01 | 0.98 | 1.00 |
| chr7\|65205946\|transcript\|LOC441242 | -0.10 | 0.67 | 0.98 |
| chr7\|65619708\|-\|- | -0.13 | 0.58 | 0.98 |
| chr7\|65620798\|-\|- | -0.25 | 0.24 | 0.85 |
| chr7\|65622507\|-\|- | 0.11 | 0.64 | 0.98 |
| chr7\|74294767\|-\|- | 0.04 | 0.84 | 0.98 |
| chr7\|86791607\|transcript\|DMTF1 | -0.35 | 0.16 | 0.76 |
| chr7\|86791622\|transcript\|DMTF1 | 0.03 | 0.92 | 0.98 |
| chr7\|97598690\|transcript$transcript\|CZ1P-ASNS$CCZ1P-OR7E38P | -0.19 | 0.45 | 0.97 |
| chr7\|99081043\|transcript\|ZNF789 | -0.08 | 0.76 | 0.98 |
| chr7\|99164430\|transcript\|ZNF655 | -0.02 | 0.94 | 0.98 |
| chr7\|100454274\|transcript\|SLC12A9 | -0.12 | 0.55 | 0.98 |
| chr7\|100736991\|transcript,exon,3UTR\|TRIM56 | 0.02 | 0.93 | 0.98 |
| chr7\|128143932\|-\|- | 0.03 | 0.88 | 0.98 |
| chr7\|139055227\|transcript$transcript\|FMC1-LUC7L2$LUC7L2 | -0.13 | 0.61 | 0.98 |
| chrX\|218554\|transcript,exon,3UTR\|PLCXD1 | 0.21 | 0.33 | 0.92 |
| chrX\|16781615\|transcript,exon,3UTR\|SYAP1 | -0.22 | 0.42 | 0.97 |
| chrX\|23721568\|3UTR,exon,transcript\|ACOT9 | 0.04 | 0.87 | 0.98 |
| chrX\|73421663\|transcript\|FTX | 0.15 | 0.49 | 0.98 |

| | | | |
|---|---|---|---|
| chrX\|77083116\|3UTR,exon,transcript\|MAGT1 | -0.14 | 0.49 | 0.98 |
| chrX\|100648057\|transcript$transcript\|RPL36A$RPL36A-HNRNPH2 | 0.13 | 0.57 | 0.98 |
| chrX\|123045142\|transcript,exon,3UTR\|XIAP | -0.17 | 0.40 | 0.97 |
| chr11\|768289\|3UTR,exon,transcript\|GATD1 | 0.36 | 0.09 | 0.76 |
| chr11\|31452750\|transcript,exon,3UTR\|DNAJC24 | 0.13 | 0.62 | 0.98 |
| chr11\|33096285\|-\|- | 0.34 | 0.11 | 0.76 |
| chr11\|66523904\|transcript,CDS,exon\|C11orf80 | -0.04 | 0.87 | 0.98 |
| chr11\|73582606\|-\|- | -0.08 | 0.78 | 0.98 |
| chr11\|93476376\|transcript,5UTR,exon\|C11orf54 | -0.19 | 0.40 | 0.97 |
| chr11\|117080491\|transcript\|PCSK7 | 0.54 | 0.02 | 0.63 |
| chr11\|118870684\|transcript\|CCDC84 | -0.06 | 0.82 | 0.98 |
| chr11\|118871252\|transcript\|CCDC84 | 0.37 | 0.20 | 0.81 |
| chr11\|118875356\|transcript\|CCDC84 | 0.10 | 0.64 | 0.98 |
| chr11\|126079051\|transcript\|RPUSD4 | 0.26 | 0.29 | 0.89 |
| chr11\|126216595\|transcript,exon,3UTR&transcript\|DCPS&GSEC | -0.09 | 0.72 | 0.98 |
| chr10\|15144305\|transcript,5UTR,exon\|RPP38 | 0.14 | 0.50 | 0.98 |
| chr10\|29748923\|transcript&transcript\|SVIL-AS1&SVIL | 0.16 | 0.52 | 0.98 |
| chr10\|46223730\|transcript\|WASHC2C | -0.35 | 0.14 | 0.76 |
| chr10\|70518191\|transcript\|CCAR1 | -0.07 | 0.73 | 0.98 |
| chr10\|73979859\|transcript,5UTR,exon\|ANAPC16 | 0.20 | 0.32 | 0.92 |
| chr10\|73995138\|transcript,exon,3UTR\|ANAPC16 | -0.14 | 0.52 | 0.98 |
| chr10\|75009591\|transcript&3UTR,exon,transcript\|DNAJC9-AS1&MRPS16 | 0.04 | 0.85 | 0.98 |
| chr10\|75537390\|-\|- | -0.05 | 0.83 | 0.98 |
| chr10\|96367532\|-\|- | -0.18 | 0.42 | 0.97 |
| chr10\|103547635\|transcript\|OGA | 0.12 | 0.57 | 0.98 |
| chr13\|111547773\|transcript\|ANKRD10 | 0.01 | 0.94 | 0.98 |
| chr12\|2997876\|transcript,exon,3UTR\|RHNO1 | 0.03 | 0.90 | 0.98 |
| chr12\|27188059\|-\|MED21&C12orf71 | 0.02 | 0.92 | 0.98 |
| chr12\|49745536\|transcript,exon,3UTR\|DNAJC22 | -0.08 | 0.74 | 0.98 |
| chr12\|53572673\|transcript\|CSAD | -0.33 | 0.20 | 0.81 |
| chr12\|121128049\|transcript\|MLEC | -0.01 | 0.96 | 1.00 |
| chr12\|121772075\|transcript\|ANAPC5 | -0.42 | 0.12 | 0.76 |
| chr12\|124007203\|transcript\|RILPL1 | 0.17 | 0.44 | 0.97 |

| | | | |
|---|---|---|---|
| chr12\|124007367\|transcript\|RILPL1 | -0.15 | 0.56 | 0.98 |
| chr12\|132422733\|transcript\|PUS1 | -0.36 | 0.10 | 0.76 |
| chr12\|133297956\|transcript,exon,3UTR\|PGAM5 | -0.05 | 0.79 | 0.98 |
| chr12\|133347372\|3UTR,exon,transcript\|GOLGA3 | 0.09 | 0.75 | 0.98 |
| chr15\|40980776\|-\|CASC5&RAD51-AS1 | -0.20 | 0.34 | 0.92 |
| chr15\|42824369\|transcript,exon,3UTR\|SNAP23 | 0.23 | 0.27 | 0.89 |
| chr15\|90440660\|3UTR,exon,transcript$transcript\|ARPIN$ARPIN-AP3S2 | -0.25 | 0.24 | 0.85 |
| chr15\|90442606\|3UTR,exon,transcript$transcript\|ARPIN$ARPIN-AP3S2 | -0.02 | 0.93 | 0.98 |
| chr15\|93433986\|transcript\|LINC01578 | -0.22 | 0.34 | 0.92 |
| chr14\|20835186\|3UTR,exon,transcript\|TEP1 | -0.19 | 0.45 | 0.97 |
| chr14\|23442149\|3UTR,exon,transcript\|AJUBA | 0.21 | 0.31 | 0.91 |
| chr14\|45600669\|transcript\|FKBP3 | 0.00 | 1.00 | 1.00 |
| chr14\|45676771\|transcript\|MIS18BP1 | 0.11 | 0.66 | 0.98 |
| chr14\|45676794\|transcript\|MIS18BP1 | -0.07 | 0.80 | 0.98 |
| chr14\|53243438\|3UTR,exon,transcript\|GNPNAT1 | 0.24 | 0.24 | 0.85 |
| chr14\|74165276\|transcript,exon,3UTR\|DNAL1 | 0.02 | 0.93 | 0.98 |
| chr14\|75202669\|transcript,exon,3UTR\|FCF1 | 0.01 | 0.94 | 0.98 |
| chr14\|100763074\|transcript\|SLC25A29 | -0.19 | 0.44 | 0.97 |
| chr14\|102789051\|transcript\|ZNF839 | -0.01 | 0.98 | 1.00 |
| chr14\|104148970\|transcript\|KLC1 | 0.05 | 0.83 | 0.98 |
| chr19\|1589247\|transcript\|MBD3 | 0.05 | 0.82 | 0.98 |
| chr19\|5710985\|transcript\|LONP1 | 0.28 | 0.19 | 0.80 |
| chr19\|11041295\|-\|- | -0.17 | 0.44 | 0.97 |
| chr19\|11147075\|transcript\|SMARCA4 | 0.40 | 0.14 | 0.76 |
| chr19\|11945216\|transcript,exon,3UTR\|ZNF440 | -0.17 | 0.46 | 0.98 |
| chr19\|12024893\|transcript,exon,3UTR\|ZNF69 | 0.00 | 0.99 | 1.00 |
| chr19\|13884176\|transcript,exon,3UTR\|MRI1 | 0.10 | 0.68 | 0.98 |
| chr19\|14680184\|transcript\|NDUFB7 | -0.30 | 0.22 | 0.85 |
| chr19\|18476294\|transcript,exon,3UTR\|PGPEP1 | -0.41 | 0.07 | 0.74 |
| chr19\|18477446\|transcript,exon,3UTR\|PGPEP1 | -0.29 | 0.23 | 0.85 |
| chr19\|18478858\|transcript,exon,3UTR\|PGPEP1 | -0.29 | 0.21 | 0.83 |
| chr19\|18479129\|transcript,exon,3UTR\|PGPEP1 | 0.00 | 1.00 | 1.00 |
| chr19\|18669846\|transcript\|KXD1 | 0.06 | 0.81 | 0.98 |
| chr19\|19019763\|transcript\|COPE | -0.08 | 0.72 | 0.98 |

| | | | |
|---|---|---|---|
| chr19\|21366997\|transcript,exon,3UTR\|ZNF431 | -0.08 | 0.73 | 0.98 |
| chr19\|21367102\|transcript,exon,3UTR\|ZNF431 | -0.26 | 0.22 | 0.84 |
| chr19\|23441380\|exon,transcript\|IPO5P1 | -0.01 | 0.97 | 1.00 |
| chr19\|36260415\|transcript,exon,3UTR\|PROSER3 | -0.10 | 0.65 | 0.98 |
| chr19\|39121713\|transcript\|EIF3K | 0.28 | 0.27 | 0.89 |
| chr19\|45165576\|transcript,exon,3UTR\|PVR | 0.08 | 0.71 | 0.98 |
| chr19\|48759084\|transcript&transcript,5UTR,exon\|CTC-241F20.3&CARD8 | 0.22 | 0.30 | 0.89 |
| chr19\|53301315\|3UTR,exon,transcript\|ZNF28 | 0.23 | 0.27 | 0.89 |
| chr19\|59078751\|transcript&transcript\|MZF1-AS1&MZF1 | 0.03 | 0.87 | 0.98 |
| chr19\|59093878\|transcript,exon\|CENPBD1P1 | 0.05 | 0.82 | 0.98 |
| chr18\|47918813\|transcript,exon,3UTR\|SKA1 | -0.41 | 0.04 | 0.66 |
| chr16\|1409326\|transcript\|GNPTG | -0.18 | 0.45 | 0.97 |
| chr16\|9204883\|transcript\|C16orf72 | 0.11 | 0.61 | 0.98 |
| chr16\|22297236\|transcript,exon,3UTR\|EEF2K | -0.05 | 0.80 | 0.98 |
| chr16\|22341640\|transcript\|POLR3E | 0.03 | 0.90 | 0.98 |
| chr16\|28969640\|transcript\|NFATC2IP | 0.05 | 0.82 | 0.98 |
| chr16\|30590009\|-\|- | -0.11 | 0.67 | 0.98 |
| chr16\|31133346\|transcript\|KAT8 | 0.11 | 0.59 | 0.98 |
| chr16\|31134704\|transcript\|KAT8 | 0.12 | 0.58 | 0.98 |
| chr16\|50139556\|transcript,exon,3UTR\|HEATR3 | -0.59 | 0.00 | 0.43 |
| chr16\|75479831\|3UTR,exon,transcript\|TMEM170A | -0.06 | 0.77 | 0.98 |
| chr16\|81063363\|transcript,exon,3UTR\|CENPN | 0.04 | 0.87 | 0.98 |
| chr16\|81063389\|transcript,exon,3UTR\|CENPN | 0.31 | 0.15 | 0.76 |
| chr16\|90235906\|transcript\|FAM157C | -0.41 | 0.11 | 0.76 |
| chr17\|1368283\|3UTR,exon,transcript\|MYO1C | -0.35 | 0.08 | 0.76 |
| chr17\|1683647\|3UTR,exon,transcript\|SMYD4 | -0.09 | 0.65 | 0.98 |
| chr17\|2320645\|3UTR,exon,transcript\|METTL16 | 0.31 | 0.13 | 0.76 |
| chr17\|4071711\|transcript\|ANKFY1 | -0.18 | 0.52 | 0.98 |
| chr17\|16287196\|-\|- | 0.38 | 0.18 | 0.80 |
| chr17\|35877278\|3UTR,exon,transcript\|SYNRG | 0.23 | 0.30 | 0.89 |
| chr17\|39964920\|transcript\|P3H4 | -0.14 | 0.66 | 0.98 |
| chr17\|49041827\|3UTR,exon,transcript\|SPAG9 | 0.29 | 0.27 | 0.89 |
| chr17\|49042252\|3UTR,exon,transcript\|SPAG9 | 0.21 | 0.38 | 0.95 |
| chr17\|57279077\|transcript,exon,3UTR\|PRR11 | 0.12 | 0.57 | 0.98 |
| chr17\|60528004\|-\|- | 0.31 | 0.16 | 0.76 |

| | | | |
|---|---|---|---|
| chr17\|62485002\|transcript\|POLG2 | -0.23 | 0.35 | 0.92 |
| chr17\|62779275\|-\|- | 0.52 | 0.03 | 0.66 |
| chr17\|79530761\|transcript\|NPLOC4 | 0.23 | 0.27 | 0.89 |
| chr21\|45644164\|3UTR,exon,transcript\|ICOSLG | 0.52 | 0.01 | 0.63 |
| chr21\|45644562\|3UTR,exon,transcript\|ICOSLG | 0.52 | 0.04 | 0.66 |
| chr22\|41327597\|transcript,exon,3UTR\|XPNPEP3 | -0.28 | 0.22 | 0.85 |
| chr1\|6707280\|transcript\|DNAJC11 | 0.31 | 0.19 | 0.80 |
| chr1\|27205044\|3UTR,exon,transcript\|GPN2 | 0.00 | 0.99 | 1.00 |
| chr1\|155705556\|transcript\|DAP3 | 0.51 | 0.06 | 0.74 |
| chr1\|173828457\|-\|- | 0.58 | 0.00 | 0.61 |
| chr1\|225976243\|transcript\|SRP9 | 0.27 | 0.31 | 0.91 |
| chr3\|194309829\|transcript,exon&transcript\|TMEM44-AS1&TMEM44 | 0.44 | 0.04 | 0.66 |
| chr2\|131119805\|transcript\|PTPN18 | -0.05 | 0.83 | 0.98 |
| chr2\|173461888\|transcript,exon,3UTR\|PDK1 | 0.07 | 0.76 | 0.98 |
| chr2\|203743850\|3UTR,exon,transcript\|WDR12 | -0.20 | 0.40 | 0.97 |
| chr2\|216808974\|3UTR,exon,transcript\|MREG | 0.04 | 0.89 | 0.98 |
| chr2\|224618846\|-\|- | 0.22 | 0.35 | 0.92 |
| chr4\|2841691\|transcript,exon,3UTR\|SH3BP2 | -0.20 | 0.38 | 0.95 |
| chr4\|71683024\|3UTR,exon,transcript\|GRSF1 | -0.54 | 0.05 | 0.66 |
| chr4\|185620850\|transcript\|CENPU | 0.02 | 0.93 | 0.98 |
| chr7\|20177603\|transcript$3UTR,exon,transcript\|LOC101927668$MACC1 | 0.38 | 0.14 | 0.76 |
| chr7\|100452736\|transcript\|SLC12A9 | 0.30 | 0.28 | 0.89 |
| chr6\|147886688\|transcript,exon,3UTR\|SAMD5 | -0.06 | 0.84 | 0.98 |
| chr8\|95688911\|transcript\|ESRP1 | -0.33 | 0.19 | 0.80 |
| chr17\|29861470\|transcript,exon,3UTR\|RAB11FIP4 | -0.02 | 0.95 | 0.98 |
| chr17\|35972246\|3UTR,exon,transcript\|DDX52 | 0.16 | 0.51 | 0.98 |
| chr17\|43583373\|exon,transcript\|LRRC37A4P | 0.32 | 0.17 | 0.77 |
| chrX\|48381084\|transcript\|EBP | -0.21 | 0.41 | 0.97 |
| chrX\|54587219\|transcript,exon,3UTR\|GNL3L | -0.25 | 0.32 | 0.91 |
| chr11\|68208679\|transcript\|LRP5 | -0.02 | 0.93 | 0.98 |
| chr11\|118876710\|transcript\|CCDC84 | -0.08 | 0.77 | 0.98 |
| chr11\|118876748\|transcript\|CCDC84 | 0.19 | 0.46 | 0.98 |
| chr13\|111551371\|transcript\|ANKRD10 | -0.15 | 0.48 | 0.98 |
| chr19\|3542080\|transcript\|C19orf71 | -0.06 | 0.80 | 0.98 |

| | | | |
|---|---|---|---|
| chr19\|17925387\|-\|- | -0.05 | 0.86 | 0.98 |
| chr19\|37817994\|transcript\|ZNF875 | 0.40 | 0.14 | 0.76 |
| chr19\|42732933\|-\|- | 0.14 | 0.62 | 0.98 |
| chr19\|52535209\|3UTR,exon,transcript\|ZNF432 | -0.05 | 0.86 | 0.98 |
| chr19\|53383600\|transcript,3UTR,exon\|ZNF320 | 0.23 | 0.29 | 0.89 |
| chr19\|55551722\|-\|- | 0.51 | 0.03 | 0.66 |
| chr19\|55551755\|-\|- | 0.36 | 0.10 | 0.76 |
| chr19\|58416290\|-\|- | -0.28 | 0.36 | 0.93 |
| chr16\|3729134\|transcript\|TRAP1 | 0.48 | 0.03 | 0.66 |
| chr16\|4512245\|transcript\|NMRAL1 | 0.40 | 0.14 | 0.76 |
| chr16\|4512262\|transcript\|NMRAL1 | 0.45 | 0.09 | 0.76 |
| chr16\|4512810\|transcript\|NMRAL1 | 0.56 | 0.05 | 0.66 |
| chr12\|112334978\|transcript,exon,3UTR\|MAPKAPK5 | 0.37 | 0.12 | 0.76 |
| chr20\|62505436\|transcript\|TPD52L2 | -0.20 | 0.50 | 0.98 |
| chr3\|52007637\|transcript\|ABHD14B | -0.07 | 0.80 | 0.98 |
| chr3\|141665627\|3UTR,exon,transcript\|TFDP2 | 0.08 | 0.75 | 0.98 |
| chr2\|206987096\|3UTR,exon,transcript\|NDUFS1 | 0.21 | 0.35 | 0.92 |
| chr12\|49180421\|transcript\|ADCY6 | -0.39 | 0.16 | 0.76 |
| chr11\|126216838\|transcript,exon,3UTR&transcript\|DCPS&GSEC | -0.25 | 0.29 | 0.89 |
| chr11\|126219036\|transcript,exon,3UTR&transcript\|DCPS&GSEC | 0.06 | 0.86 | 0.98 |
| chr16\|56974318\|transcript\|HERPUD1 | -0.40 | 0.12 | 0.76 |
| chr4\|120979770\|3UTR,exon,transcript\|MAD2L1 | 0.22 | 0.46 | 0.98 |
| chr9\|130875311\|transcript\|SLC25A25-AS1 | 0.44 | 0.07 | 0.74 |
| chr7\|17384437\|transcript,exon,3UTR\|AHR | 0.43 | 0.05 | 0.66 |
| chr7\|128145040\|-\|- | -0.10 | 0.71 | 0.98 |
| chrX\|118672525\|3UTR,exon,transcript\|CXorf56 | 0.26 | 0.37 | 0.95 |
| chr11\|68208748\|transcript\|LRP5 | 0.25 | 0.34 | 0.92 |
| chr1\|204521634\|transcript,3UTR,exon\|MDM4 | -0.16 | 0.51 | 0.98 |
| chr2\|131119681\|transcript\|PTPN18 | 0.31 | 0.22 | 0.85 |
| chr2\|206986246\|3UTR,exon,transcript\|NDUFS1 | -0.03 | 0.91 | 0.98 |
| chr2\|216808321\|3UTR,exon,transcript\|MREG | 0.17 | 0.48 | 0.98 |
| chr9\|37775695\|transcript\|TRMT10B | 0.22 | 0.48 | 0.98 |
| chr7\|2762018\|transcript\|AMZ1 | 0.49 | 0.07 | 0.74 |
| chr11\|16778078\|transcript,exon,3UTR\|C11orf58 | -0.03 | 0.91 | 0.98 |

| | | | |
|---|---|---|---|
| chr15\|40985066\|-\|- | 0.51 | 0.04 | 0.66 |
| chr12\|117469471\|-\|- | 0.04 | 0.88 | 0.98 |
| chr7\|25160735\|3UTR,exon,transcript\|CYCS | 0.30 | 0.32 | 0.91 |
| chr7\|77411181\|transcript,exon,3UTR\|RSBN1L | -0.06 | 0.82 | 0.98 |
| chr17\|60527756\|-\|- | 0.06 | 0.83 | 0.98 |

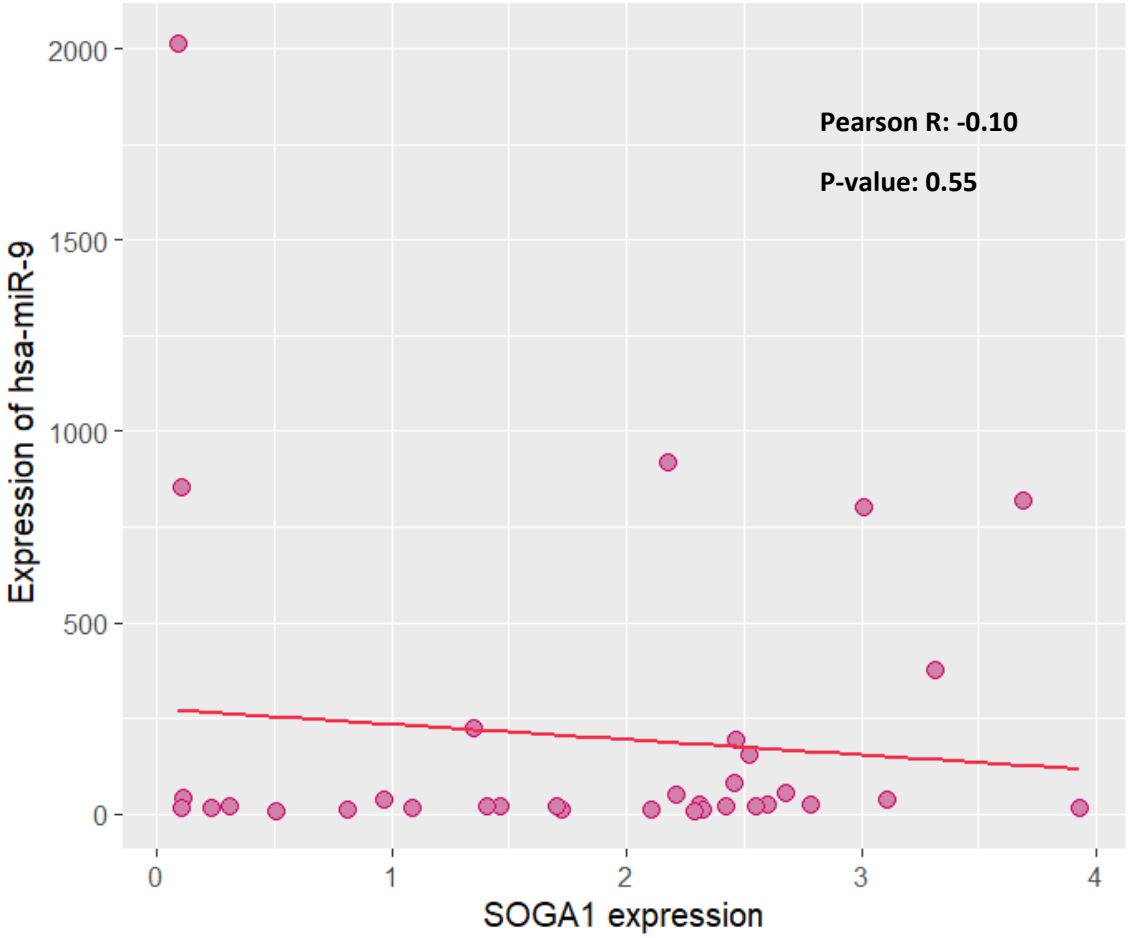
## 7.5 Appendix E

Table of editing events correlated to gene expression

| Editing | Gene | Spearman Rho | Raw p-value | BH correction |
|---|---|---|---|---|
| chr20\|35411013\|transcript,3UTR,exon\|SOGA1 | SOGA1 | 0.66 | 0.00 | 0.49 |
| chr6\|147886688\|transcript,exon,3UTR\|SAMD5 | SAMD5 | -0.71 | 0.01 | 0.49 |
| chr3\|10192617\|transcript,3UTR,exon\|VHL | VHL | 0.53 | 0.01 | 0.49 |
| chr11\|126216595\|transcript,exon,3UTR&transcript\|DCPS&GSEC | DCPS | -0.57 | 0.01 | 0.49 |
| chr19\|1589247\|transcript\|MBD3 | MBD3 | -0.49 | 0.01 | 0.49 |
| chr1\|67874849\|3UTR,exon,transcript\|SERBP1 | SERBP1 | -0.47 | 0.02 | 0.49 |
| chr12\|53572673\|transcript\|CSAD | CSAD | -0.56 | 0.02 | 0.49 |
| chr19\|48759084\|transcript&transcript,5UTR,exon\|CTC-241F20.3&CARD8 | CTC-241F20.3 | 0.46 | 0.02 | 0.49 |
| chr1\|100489835\|transcript,exon,3UTR\|SLC35A3 | SLC35A3 | 0.45 | 0.02 | 0.49 |
| chr17\|2320645\|3UTR,exon,transcript\|METTL16 | METTL16 | 0.45 | 0.02 | 0.49 |
| chr17\|35877278\|3UTR,exon,transcript\|SYNRG | SYNRG | 0.49 | 0.02 | 0.49 |
| chr3\|10194651\|transcript,3UTR,exon\|VHL | VHL | 0.44 | 0.02 | 0.51 |
| chr1\|1338166\|transcript\|MRPL20 | MRPL20 | -0.45 | 0.03 | 0.51 |
| chr7\|44873177\|transcript\|H2AFV | H2AFV | -0.42 | 0.03 | 0.53 |
| chr11\|118876710\|transcript\|CCDC84 | CCDC84 | 0.52 | 0.03 | 0.53 |
| chr4\|774074\|exon,transcript\|RP11-440L14.1 | RP11-440L14.1 | 0.45 | 0.03 | 0.53 |
| chr21\|45644562\|3UTR,exon,transcript\|ICOSLG | ICOSLG | 0.51 | 0.04 | 0.59 |
| chr18\|47918813\|transcript,exon,3UTR\|SKA1 | SKA1 | -0.41 | 0.04 | 0.59 |
| chr20\|62505436\|transcript\|TPD52L2 | TPD52L2 | 0.55 | 0.04 | 0.59 |
| chr19\|39121713\|transcript\|EIF3K | EIF3K | -0.47 | 0.05 | 0.60 |

## 7.6 Appendix F

Correlation graph of SOGA1 and hsa-miR-9 expression in Gastric Cancer Cell Lines



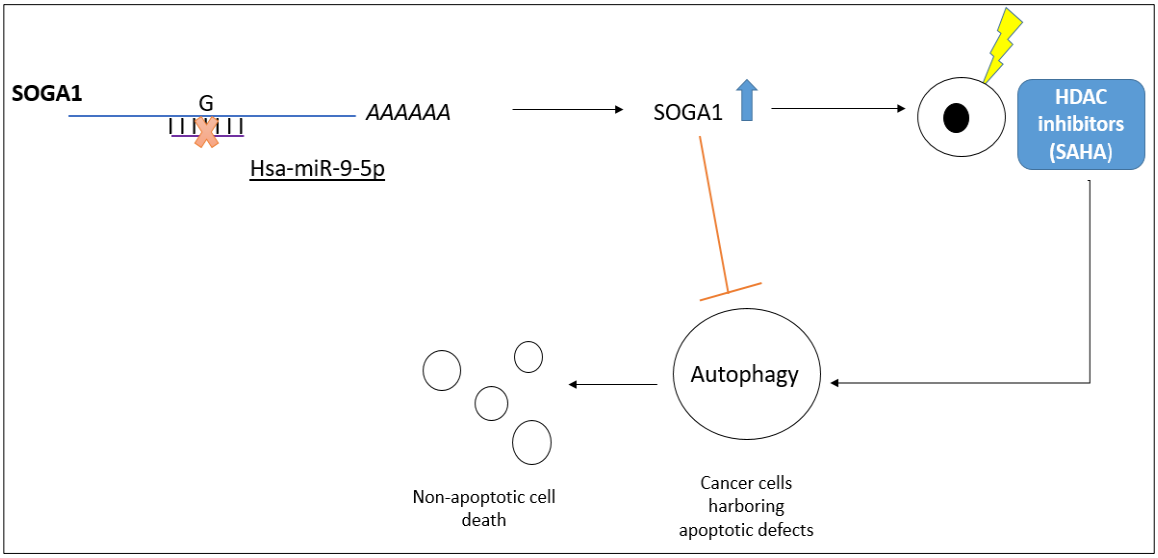
## 7.7 Appendix G

Univariate Cox Regression Analysis of gene expression in TCGA cohort

| Gene | Hazard Ratio (HR) | p-value |
|---|---|---|
| TEP1 | 0.999904 | 0.612528 |
| NUP43 | 0.999736 | 0.546181 |
| ZDHHC20 | 0.999173 | 0.02425 |
| DCAF16 | 0.999585 | 0.320054 |
| CTSB | 1.000003 | 0.382319 |
| ACOX1 | 1.000117 | 0.097744 |
| CPT1A | 1.000046 | 0.399466 |
| CLPX | 1.000103 | 0.830455 |
| NDUFS1 | 1.000061 | 0.751828 |
| CCDC125 | 0.998564 | 0.161628 |
| ARSD | 0.999933 | 0.636048 |
| CTSS | 0.999928 | 0.117274 |
| BPNT1 | 0.999748 | 0.489457 |
| GNPNAT1 | 0.99964 | 0.024581 |
| NUP155 | 0.999864 | 0.563637 |

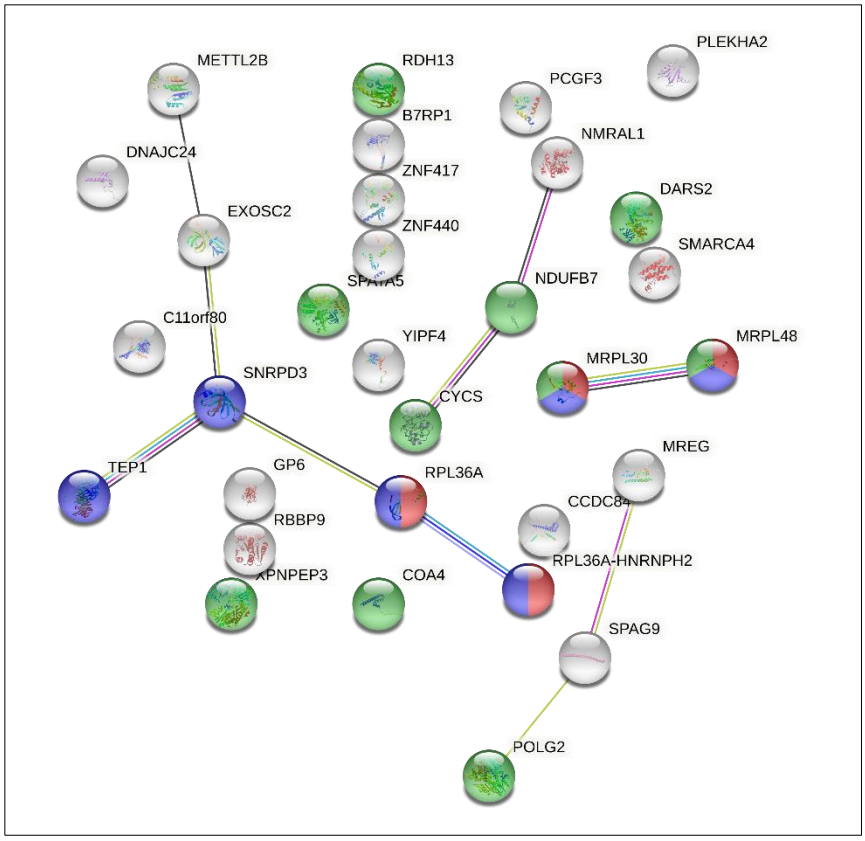
## 7.8 Appendix H

Schematic representation of possible mechanism by which editing event in SOGA1 3'UTR induce Vorinostat resistance



## 7.9 Appendix I

STRING enrichment analysis of genes edited correlating with Lapatinib sensitivity




| Pathway | observed gene count | background gene count | strength | FDR | matching proteins in your network |
|---|---|---|---|---|---|
| Mitochondrion | 10 | 1176 | 0.73 | 0.00032 | NDUFB7,SPATA5,CYCS,MRPL48, MRPL30,COA4,XPNPEP3,DARS2, RDH13,POLG2 |
| Ribonucleoprotein | 6 | 278 | 1.13 | 0.00032 | SNRPD3,TEP1,MRPL48,MRPL30,RPL36A-HNRNPH2,RPL36A |
| Ribosomal protein | 4 | 173 | 1.16 | 0.0377 | MRPL48,MRPL30,RPL36A-HNRNPH2,RPL36A |

## 7.10 Appendix J

miRNET network analysis results