Group Project 2 – Group 5

Introduction

Our group project focuses on an article written by Li and colleagues,[1] who examined 6 single nucleotide polymorphisms (SNPs) associated with lung cancer: rs2736100, rs4488809, rs753955, rs12296850, rs2853677, and rs2741354.[2-4] The authors genotyped these SNPs in Chinese patients with lung cancer (n=391) as well as control subjects (n=337). Although no associations were detected that reached genome-wide significance, subgroup analysis did uncover one significant association: rs2853677 was associated with adenocarcinoma when using a dominant genotypic model (odds ratio [OR]: 2.33, p-value=0.0024).

Next, Li and colleagues focused on this particular SNP, which is located in the second intron of a gene called telomerase reverse transcriptase (*TERT*). They demonstrated that the region harboring rs2853677 increased the expression of *TERT*, suggesting that it might be located in an area that represents a transcriptional enhancer. Additional analyses indicated that this SNP serves as a binding site for the Snail family of proteins. Moreover, they were able to demonstrate that binding of Snail1 is affected by rs2853677 and that this influences the activity of the *TERT* enhancer. Although Snail1 is known to act as a transcriptional repressor that can regulate epithelial-mesenchymal transition (EMT),[5] the authors did not obtain any evidence that rs2853677 stimulates metastasis through EMT. They did, however, provide data supporting the hypothesis that a long-range physical interaction exists between the *TERT* promoter and its enhancer. They conclude, therefore, that rs2853677 may hamper the binding of Snail1, which subsequently affects the function of the *TERT* enhancer, and consequently, the transcription of *TERT* itself. Given the pivotal role of TERT in telomere maintenance and the regulation of self-renewal, it is not surprising that aberrant expression of *TERT* is associated with cancer.[6] As such, this study provides an excellent example of a non-coding SNP that promotes limitless self-renewal by impacting the binding of a co-factor.[1]

Part 1 – Regulation

TERT and rs2853677

The *TERT* gene, at band 5p15.33, encodes a ribonucleoprotein polymerase that synthesizes telomere ends by adding hexanucleotide repeats (TTAGGG).[7] This addition allows complete replication at the ends of chromosomes for most eukaryotes. The TERT enzyme contains both a protein and an RNA component.[8] Its protein component carries out the reverse transcriptase activity, while its RNA component serves as a template for the extension of chromosomal termini.[7]

TERT is primarily active in progenitor cells and is repressed in postnatal somatic cells, explaining its involvement in cellular senescence.[7-8] Deregulation or increased activity of this enzyme has been observed in cancer cells, contributing to oncogenesis through enzyme-dependent telomere lengthening and enzyme-independent intermolecular interactions with p53, poly (ADP-ribose) polymerase (PARP) and, as discussed in the study conducted by Li and colleagues, binding of co-factor Snail1.[1,9]

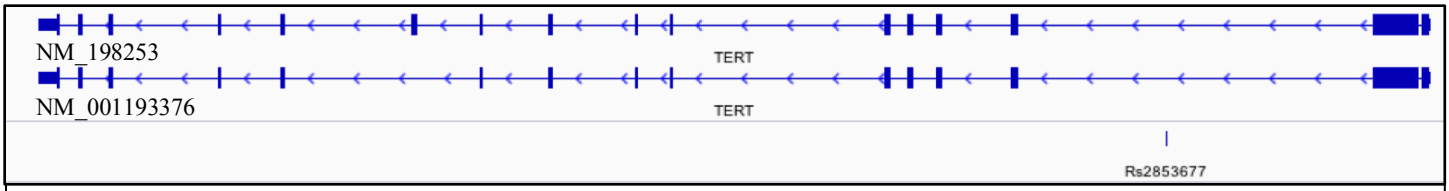

Figure 1 | Graphical overview of region containing TERT. A region is displayed on chromosome 5 between 1,253,167 and 1,295,047. This region contains a gene called *TERT*. The direction of the **arrow** specifies that this gene is located on the reverse strand (**right-to-left**). The 2 curated RefSeq genes are shown: NM_198253 and NM_001193376. The SNP of interest, rs2853677, is located in the second intron of both these transcripts.

The rs2853677 variant is an intronic variant where the ancestral allele (thymine [T]) is replaced by a cytosine (C). It is located on chromosome 5 (chr5:1,287,079-1,287,079) in the second intron of the two curated *TERT* transcripts that are transcribed from the reverse strand (NM_198253 and NM_001193376; **Figure 1**).[10-12]

Telomerase and cancer

Normally, cells have a cell-autonomous program that controls multiplication.[13] This is achieved by telomeres at the ends of chromosomes, which contain a repetitive sequence that protects chromosomal DNA. Since DNA polymerase cannot replicate the 3' end, telomeres are shortened during each cell division. Beyond a certain point, chromosomal DNA is no longer safeguarded, leading to cell death. Importantly, overexpression of the telomerase enzyme results in the addition of extra hexanucleotide repeats at the end of the telomeres.

Consequently, this produces cells with unlimited replicative potential, basically immortalizing them. In fact, limitless replicative potential is one of the hallmarks of cancer, just like self-sufficiency in growth signals, sustained angiogenesis, tissue invasion & metastasis, insensitivity to anti-growth signals, and evading apoptosis. Together, these hallmarks result in uncontrolled cell proliferation, laying the foundation for cancerous cell growth (**Figure 2**).[13]

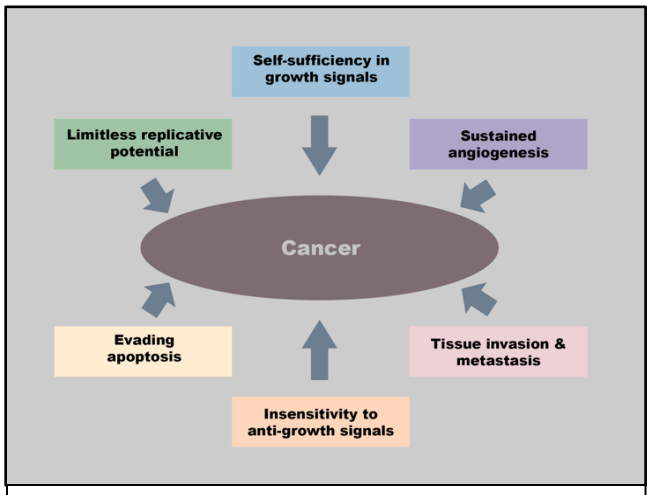

Figure 2* | *Hallmarks of cancer. Six hallmarks of cancer are displayed: limitless replicative potential, self-sufficiency in growth signals, sustained angiogenesis, tissue invasion & metastasis, insensitivity to anti-growth signals, and evading apoptosis. Adapted from Hanahan & Weinberg.[13]

Clinical consequences

As mentioned above, Li and colleagues revealed that rs2853677 might impact binding of Snail1 to an enhancer that affects the expression of *TERT*, which has the ability to contribute to limitless replication and might explain the association with lung adenocarcinoma.[1] Lung cancer is the leading cause of death worldwide, resulting in 1.6 million cancer-related deaths.[14,15] Two categories exist: non-small cell lung cancer (NSCLC; 85%) and small cell lung cancer (SCLC; 15%); the latter has a worse prognosis. Adenocarcinoma (gland-forming) is one of the histologic subtypes of NSCLC and is the most common subtype detected in people who have never smoked. Unfortunately, clinical symptoms are frequently absent, and patients may present at a fairly late stage of the disease with chest pain, cough, weight loss, dyspnea, and hemoptysis. Such an advanced-stage disease generally results in a poor prognosis with a median overall survival of 10-12 months.[15]

SNP frequency

According to dbSNP, the minor allele frequency (MAF) of rs2853677 is approximately 39%.[12] Based on 1000 Genomes and HapMap,[12] however, the frequency of the risk allele (C) varies considerably, from 22% to 61% (**Table 1**). The lowest frequency is detected in Yoruba from Nigeria, whereas the highest frequency is observed in Gujarati Indians from Texas.

Table 1 | Allele frequencies for rs2853677

Study	Cohort	Description	Count	T Allele	C Allele

1000 Genomes	EAS	East Asia	1008	64%	36%
	EUR	Europe	1006	59%	41%
	AFR	Africa	1322	74%	26%
	AMR	America	694	65%	35%
	SAS	South Asia	978	42%	58%
HapMap	CEU	Utah residents with Northern and Western European ancestry	266	58%	42%
	HCB	Unrelated Han Chinese in Beijing, China	86	60%	40%
	JPT	Japanese in Tokyo, Japan	172	74%	26%
	YRI	Yoruba in Ibadan, Nigeria	226	78%	22%
	AWS	African ancestry in Southwest USA	98	64%	36%
	CHB	Han Chinese in Beijing, China	82	62%	38%
	CHD	Chinese in Metropolitan Denver, Colorado	170	66%	34%
	GIH	Gujarati Indians in Houston, Texas	176	39%	61%
	LWK	Luhya in Webuye, Kenya	180	69%	31%
	MEX	Mexican ancestry in Los Angeles, California	100	71%	29%
	MKK	Maasai in Kinyawa, Kenya	286	52%	48%
	TSI	Toscans in Italy	176	60%	40%

In 1000 Genomes, the frequencies most relevant to Li's study are probably a risk-allele frequency of 41% in Europeans, 35% in Americans, 36% in East Asians, and 58% in South Asians. For the HapMap project, we focused on individuals of European (MAF: 42%), Japanese (MAF: 26%), and Han Chinese (MAF: 38-40%) ancestry; in **Table 2**, we have specified the genotypes for each of those populations. Importantly, the frequencies observed by Li and colleagues (in patients and controls) seem comparable to those reported in the Han Chinese population.[1,12]

Table 2 | Genotype frequencies for rs2853677

Cohort	**T/T**	**C/T**	**C/C**
CEU	33%	50%	17%
JPT	55%	38%	7%
HCB	35%	51%	14%
CHB	37%	51%	12%
Li et al. Controls	42%	47%	11%
Li et al. Patients	35%	50%	15%

Encyclopedia of DNA Elements (ENCODE)

Histone modifications represent signals that can be used to find regulatory regions and to examine the transcriptional activity of genes (**Table 3**).[16] Histone monomethylation can identify active genes as well as cell-type-specific enhancer sites.[17] Importantly, histone trimethylation data obtained from lymphoblastoid cells is known to show a strong signal at the transcription start site (TSS). An H3K27ac signal, on the other hand, can be used to distinguish active enhancers from inactive/poised enhancers.[17] ENCODE data for our region of interest, which surrounds rs2853677, does not demonstrate a strong signal for H3K27ac (**Figure 3**), suggesting that it may actually represent an inactive enhancer that retains the ability to become activated. As an example, many inactive developmental genes in hematopoietic stem cells express H3K4me1-enriched enhancers.[17] These patterns, therefore, appear to indicate that the enhancer studied by Li and colleagues may, in fact, represent a poised enhancer.

Table 3 | Effects of histone modifications

Modification Type	**H3K4**	**H3K9**	**H3K27**	**H3K79**
Monomethylation (me1)	Activation	Activation	Activation	Activation
Dimethylation (me2)	--	Repression	Repression	Activation
Trimethylation (me3)	Activation	Repression	Repression	Activation/Repression
Acetylation (ac)	--	Activation	Activation	--

DNase I hypersensitivity sites (DHSs) mark regions of chromatin that have lost its condensed structure. These exposed regions are vulnerable, and therefore, they demonstrate elevated DNase I activity. DHSs are considered hallmarks of regulatory DNA, because they are often found near a TSS, enhancer, or silencer. *TERT* contains 2 DHS clusters, one close to the TSS, p137235, and one in intron 6, p137234 (**Figure 3**). The *TERT* gene also contains 9 GpG islands (**Figure 3**). The largest Cpg island (CpG:_485) is located close to our SNP (chr5:1,289,161-1,295,855); this feature marks the promoter region and the TSS of *TERT*. Interestingly, it has been suggested that there is a relationship between CpG island length and expression complexity.[18] Long CpG islands or promoters (>2,000 bases), for instance, are associated with genes that have an intermediate level of tissue specificity and that are involved in development and regulation.[18] These types of promoters have increased polymerase binding sites and TSS regions when comparing them to shorter promoters. Overall, these observations suggest that the expression of the *TERT* gene might be complicated, and potentially, challenging to untangle.

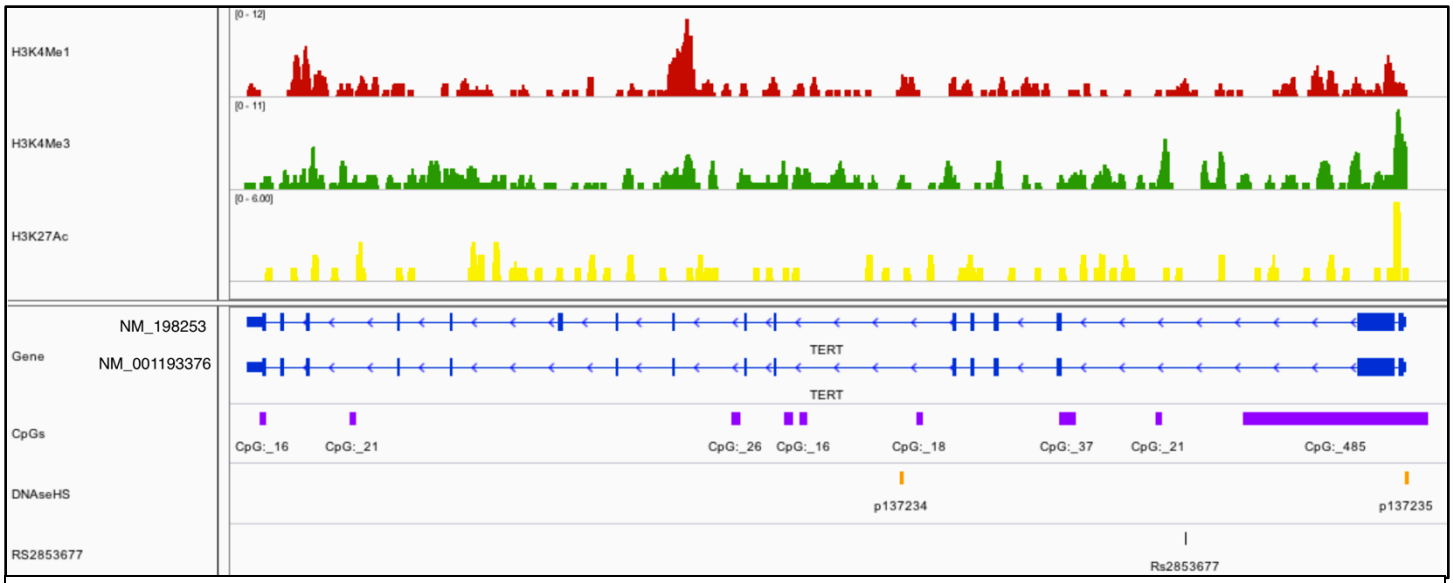

Figure 3 | Visualization of TERT's regulatory regions in G12817 cells based on data obtained from Encyclopedia of DNA Elements. In addition to the *TERT* transcripts (**blue**), histone marks are visualized. Mono- and trimethylation of H3K4 (**red** and **green**, respectively) and acetylation of H3K27 (**yellow**) are shown, which are an indication of active transcription. Additionally, locations of 9 CpG islands (**purple**) and 2 DNase hypersensitivity regions (**orange**) are outlined. The SNP of interest, rs2853677, is also displayed (**black**).

Binding of transcription factors

In the promoter, SNPs can interfere with normal transcription factor binding or create a new binding site for transcription factors. This can lead to either up- or down-regulation of a given gene. However, when located in other non-coding areas of a gene, including different regulatory regions where transcription factors bind, the effect of SNPs has not been well studied.[1] Li and colleagues focused on rs2853677,

which was found to be located in a *TERT* enhancer, where the protein Snail1 binds. If positioned near the promoter of *TERT*, this enhancer is able to increase the transcription of *TERT*. When transcription factor Snail1 is bound to the enhancer, this causes a conformational change in the chromatin and prevents co-localization of the *TERT* promoter and enhancer, resulting in down-regulation of *TERT*.[1]

Li and colleagues, therefore, hypothesized that rs2853677 may block Snail1 from binding, enhancing transcription and causing overexpression of *TERT*. They tested this hypothesis *in vitro* using an electrophoretic mobility shift assay (EMSA) with the Snail1 binding consensus sequence and the sequence containing the SNP. To measure Snail1 binding to the enhancer containing the polymorphism *in vivo*, they used chromatin immunoprecipitation (ChIP) analysis to confirm that the SNP prevents Snail1 from binding.[1]

In addition to regulating expression of *TERT*, Snail1 is known to regulate the expression of telomeric repeat-containing RNA (*TERRA*), a long non-coding RNA.[19] TERRA has been found to be expressed near the ends of most chromosomes in the subtelomeric region and is regulated in response to signals based on the length of telomeres. Interestingly, TERRA is involved telomere maintenance and structuring of chromatin during development and cell differentiation.[20] Additionally, Snail1 is involved in the regulation of fibronectin (FN1), which binds to the promoter and enhances transcription during EMT. It is also known to repress transcription of several epithelial genes, such as E-cadherin (CDH1). Moreover, Snail1 recruits lipoxygenase 2 (LOX2), which plays a part in reorganization of heterochromatin regions, increasing mesenchymal traits in cells.[5,21]

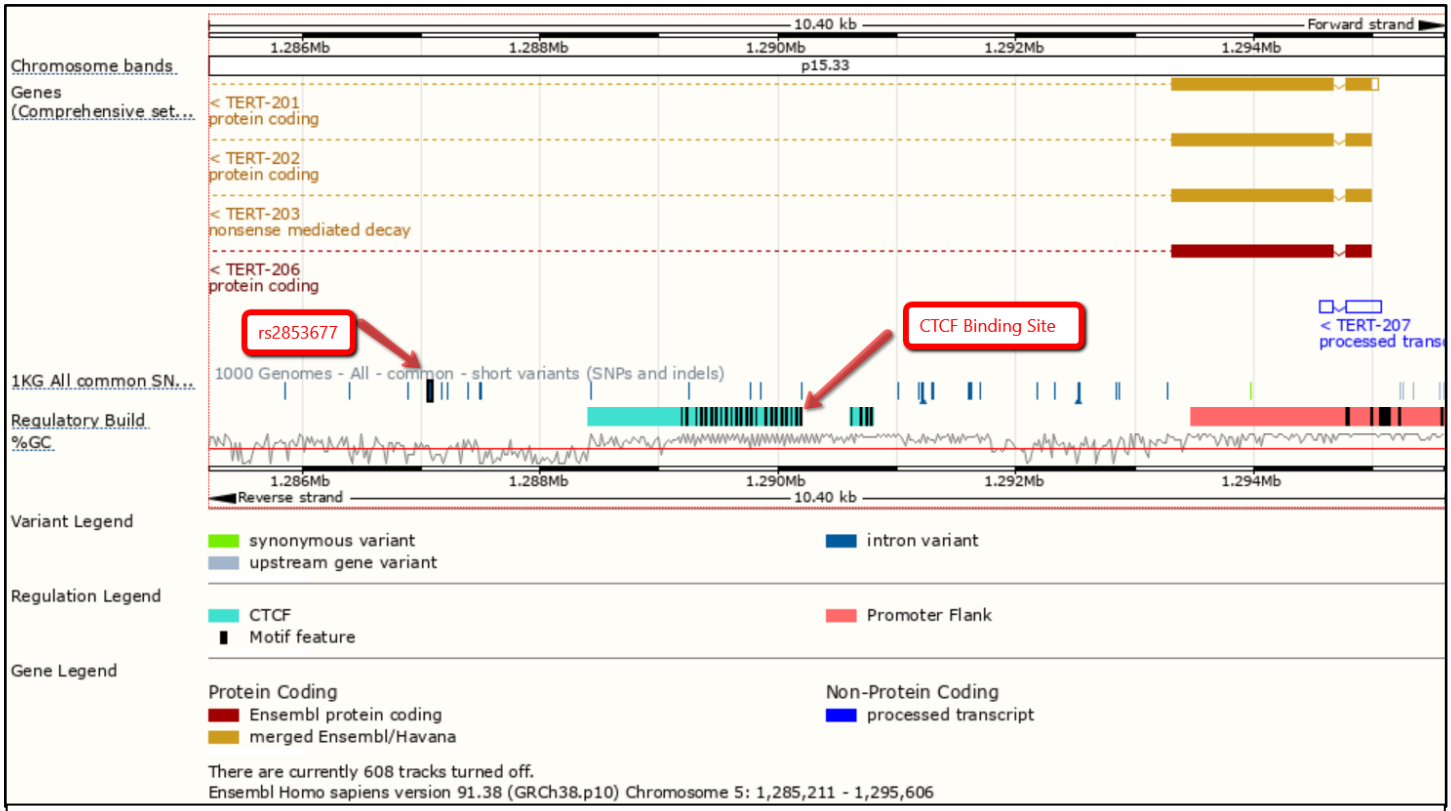

Figure 4 | TERT transcription factor binding sites. This figure shows the region in which the second intron of the *TERT* gene is located. Our SNP of interest, rs2853777, is highlighted in the track with common SNPs. A CTCF binding site (**turquoise**) is displayed in the regulatory track. Interestingly, a CTCF binding site is located near our SNP and approximately 4 kb downstream of *TERT*'s transcription start site (TSS). Binding of CTCF at this site is thought to repress transcription of the *TERT* gene.

Other transcription factors

Several other transcription factors are capable of regulating *TERT* transcription, and many of those are affected by SNPs in the promoter or other regulatory regions. Mutations in the *TERT* promoter region, for example, have been shown to cause recruitment of GA-binding protein (GABP), increasing *TERT* expression in various types of cancer.[22] Other mutations in the promoter region create a binding site for E26 transformation-specific (ETS) transcription factors and ternary complex factors (TCFs), which can up to double transcription of *TERT* in human melanoma.[23] Most significantly, chromatin remodeling protein CCCTC-binding factor (CTCF) is able to repress transcription of *TERT* by binding to an enhancer element approximately 4 kb downstream of the *TERT* TSS (**Figure 4**).[24]

Future experiments

In this document, we have provided an overview of the article written by Li and colleagues, integrating information about the area surrounding rs2853677. Overall, this data supports the hypothesis that rs2853677 might influence the expression of a gene involved in telomere maintenance (*TERT*) through interference with the binding of a co-factor to an enhancer. It should be noted, however, that the genetic association between rs2853677 and lung adenocarcinoma needs to be replicated in an independent cohort. Given the fairly small sample size of the cohort examined by Li and colleagues, and the fact that their association was only present in a specific histologic subtype, future studies should investigate a larger number of patients belonging to each of the subtypes. Moreover, since the authors solely investigated Chinese patients, it remains to be determined whether this association can be encountered in other populations. Ideally, clinical studies should also be performed to elucidate whether the presence of the risk allele (C) affects the severity of the disease. One could postulate, for instance, that it might be associated with a more aggressive form of lung adenocarcinoma, possibly resulting in a shorter survival. Additionally, it would be important to investigate a potential association with other types of cancer to establish whether the association is specific for lung adenocarcinoma or whether it is broadly associated with cancer. We are confident that these studies will help to assess the clinical relevance of the findings reported by Li and colleagues.

Part 2 – Next-generation sequencing (NGS) processing

Methodology

For our analysis, we used the following paired-end NGS data:

- ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data/HG04020/sequence_read/SRR791475_1.filt.fastq.gz
- ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data/HG04020/sequence_read/SRR791475_2.filt.fastq.gz

We uploaded this data using the ‘Paste/Fetch data’ tool in Galaxy,[25] selecting ‘fastqsanger’ as format and ‘hg38’ as reference genome. Next, we obtained box plots of the quality scores using the ‘FASTQC’ tool from ‘NGS: QC and manipulation’ (**Figure 5**). We noticed that the read length was 76 bases and that the total number of reads equaled 2,693,534. Although the quality score of the reads decreased slightly when the read length increased, on average, the quality was good (within green area). Because the encoding was ‘Sanger/Illumina 1.9’, we did not have to change the file format.

Next, we trimmed our files using the ‘Trimmomatic’ tool from ‘NGS: QC and manipulation’, ensuring that we selected paired-end, a sliding window of 4, and a quality score of 20. We then used the ‘FASTQC’ tool again to determine how the trimming had affected our quality scores (**Figure 5**). After trimming, the total number of reads was 2,564,724 and their length ranged from 1 to 76. All reads were located within the green area, and consequently, their quality was good.

To align the trimmed reads, we selected ‘NGS: Mapping’, ‘Map with BWA’, ‘Human (Homo sapiens) (b38): hg38’, and ‘Paired fastq’. Subsequently, we used the ‘FreeBayes’ tool from ‘NGS: Variant Analysis’, selecting ‘Human (Homo sapiens): hg38’. Additionally, we limited our analysis to a region on chromosome 5, containing our gene of interest (*TERT*) as well as up- and downstream sequences: chr5:1200000-1300000. Our VCF contained 10 variants, before we filtered it using the ‘VCFfilter’ tool from ‘NGS: VCF Manipulation’. We kept variants with a read depth above 10 (DP > 10). After application of this filter, only 3 variants remained.

In order to annotate our 2 VCF files (unfiltered and filtered), we acquired a BED file using the ‘UCSC Main’ tool from ‘Get Data’. Our BED file contained information from NCBI RefSeq about 5 transcript variants, namely: NM_001003841.2 (solute carrier family 6 member 19 [*SLC6A19*]), NM_182632.2 (solute carrier family 6 member 18 [*SLC6A18*]), NM_198253.2 (*TERT*), NM_001193376.1 (*TERT*), and XM_017009796.1 (*TERT*). Of note, the latter transcript has been removed from NCBI; it is currently deemed obsolete. We then utilized the ‘VCFannotate’ tool from ‘NGS: VCF Manipulation’.

Figure 5 | FASTQC analysis before and after trimming. Before trimming, a total of 2,693,534 reads is obtained. Their read length is 76 nucleotides. Most of the box plots, including their interquartile range (IQR), fall within the **green** area. After trimming, the total number of reads is 2,564,724 with a read length varying between 1 and 76 nucleotides. Again, most of the box plots are located within the **green** area. It should be stressed that trimming of the reads further improved the quality, which is demonstrated by the tightness of the box plots.

Table 4 | Variants present in unfiltered VCF

Number	Chromosome	Position	dbSNP	Reference Allele	Alternative Allele	Read Depth	Transcript(s)	Gene
1	chr5	1212532	rs13180809	C	T	3	NM_001003841.2	*SLC6A19*
2	chr5	1214149	rs6554663	A	G	2	NM_001003841.2	*SLC6A19*
3	chr5	1235675	rs10042399	A	G	11	NM_182632.2	*SLC6A18*
4	chr5	1240642	rs7447815	C	G	10	NM_182632.2	*SLC6A18*
5	chr5	1243618	rs574913359	G	A	19	NM_182632.2	*SLC6A18*
6	chr5	1244310	rs4073918	C	T	12	NM_182632.2	*SLC6A18*
7	chr5	1244749	NA	G	A	9	NM_182632.2	*SLC6A18*
8	chr5	1244839	rs73034557	C	A	2	NM_182632.2	*SLC6A18*
9	chr5	1268530	NA	G	T	3	NM_198253.2:NM_001193376.1:XM_017009796.1	*TERT*
10	chr5	1280362	rs13167280	G	A	4	NM_198253.2:NM_001193376.1:XM_017009796.1	*TERT*

Results for TERT

The 10 variants we identified were present in *SLC6A19*, *SLC6A18*, and *TERT* (**Table 4**; **Galaxy19-[VCFannotate__on_data_18_and_data_16]-2.vcf**). Importantly, out of those 10 variants, only 2 were located in the *TERT* gene. One *TERT* variant was novel and had not been reported in dbSNP. This variant was located at position 1,268,530 (exon 9). It is thought to change a CGG codon to an AGG codon (on the reverse strand), which makes it a synonymous variant that encodes an arginine (Arg; both the wild-type allele and the mutant allele).[26] The second SNP was located in intron 3 and had been reported as rs13167280; no predictions of its potential effects were available.[26,27] Furthermore, it should be emphasized that these *TERT* SNPs demonstrated a low coverage (**Figure 6**). Not surprisingly, therefore, we lost both after filtering (**Table 5**; **Galaxy20-[VCFannotate__on_data_18_and_data_17]-2.vcf**). Given their low coverage, we cannot determine whether these variants are real and additional studies are needed to validate them. Moreover, we would like to emphasize that our variant of interest from Part 1, rs2853677, was not covered at all, and consequently, its presence could not be examined in this NGS data.

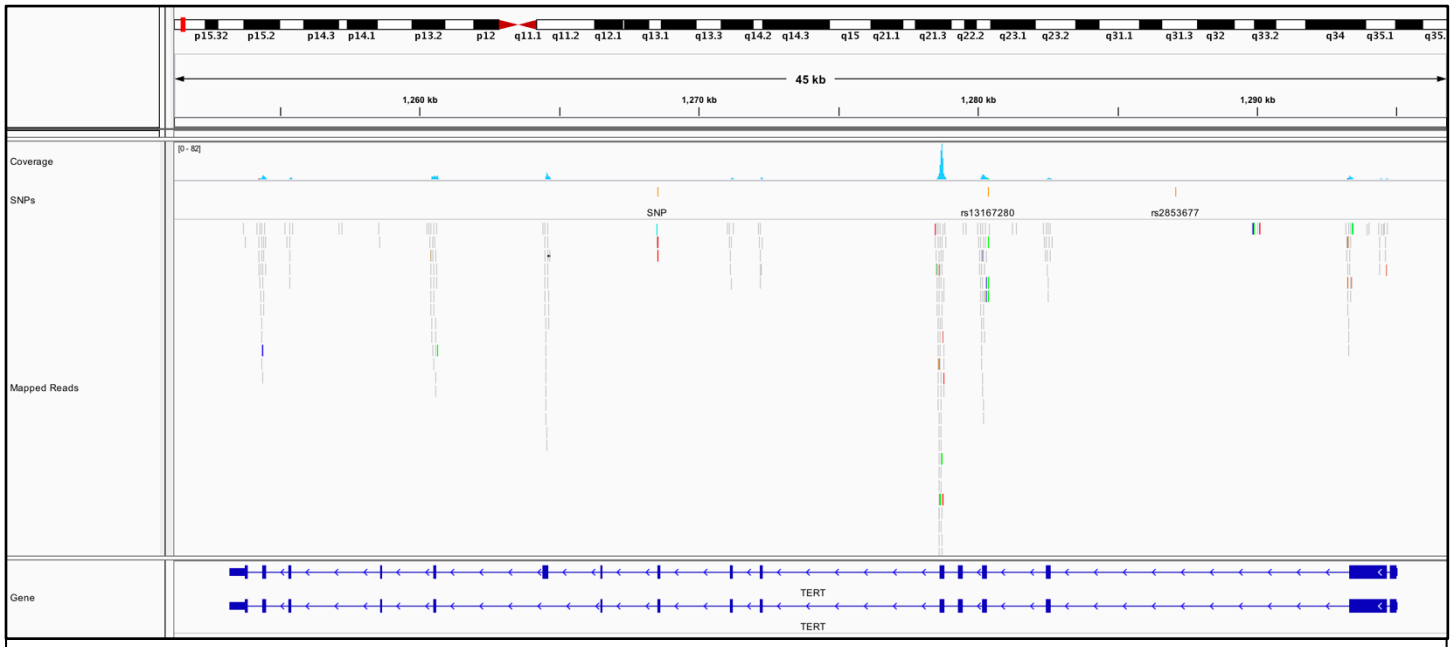

Figure 6 | Coverage of TERT gene. The coverage of the *TERT* gene is variable (**light blue**). Some areas show a high coverage (up to 82 reads), but the coverage in other areas is low. Three variants are displayed (**orange**), including the variant from Part 1 of this assignment (rs2853677), which is not covered at all. Two other variants that are present in our unfiltered VCF at position 1,268,530 and 1,280,362 demonstrate a low coverage of 3 and 4 reads, respectively.

Results for SLC6A18

While no variants were present in the *TERT* gene after filtering, the *SLC6A18* gene revealed 3 variants with a read depth greater than 10 that were worth exploring further. Two of those variants were found to be non-synonymous SNPs (rs4073918 and rs574913359), whereas the third variant (rs10042399) was intronic. The rs4073918 variant results in a proline (Pro; CCG) to leucine (Leu; CTG) change at amino acid position 478 and is predicted to be benign by both SIFT and Polyphen.[26,27] The other non-synonymous SNP we identified, rs574913359, causes a change from an alanine (Ala; GCT) to a threonine (Thr; ACT) at position 399, which is predicted to be deleterious.[26,27] This variant alters the SLC6A18 protein, which may have detrimental effects, since this protein is a sodium-dependent neutral amino acid transporter. Nevertheless, no phenotypic consequences have been recorded for this particular SNP and additional studies are warranted to elucidate its potential effects.[28]

Table 5 | Variants present in filtered VCF

Number	Chromosome	Position	dbSNP	Reference Allele	Alternative Allele	Read Depth	Transcript(s)	Gene
1	chr5	1235675	rs10042399	A	G	11	NM_182632.2	*SLC6A18*
2	chr5	1243618	rs574913359	G	A	19	NM_182632.2	*SLC6A18*
3	chr5	1244310	rs4073918	C	T	12	NM_182632.2	*SLC6A18*

References

1. Li X, Xu X, Fang J, Wang L, Mu Y, Zhang P, Yao Z, Ma Z, Liu Z. Rs2853677 modulates Snail1 binding to the TERT enhancer and affects lung adenocarcinoma susceptibility. Oncotarget. 2016 Jun 21;7(25):37825-37838. doi:10.18632/oncotarget.9339. PubMed PMID: 27191258; PubMed Central PMCID: PMC5122352.
2. Hu Z, Wu C, Shi Y, Guo H, Zhao X, Yin Z, Yang L, Dai J, Hu L, Tan W, Li Z, Deng Q, Wang J, Wu W, Jin G, Jiang Y, Yu D, Zhou G, Chen H, Guan P, Chen Y, Shu Y, Xu L, Liu X, Liu L, Xu P, Han B, Bai C, Zhao Y, Zhang H, Yan Y, Ma H, Chen J, Chu M, Lu F, Zhang Z, Chen F, Wang X, Jin L, Lu J, Zhou B, Lu D, Wu T, Lin D, Shen H. A genome-wide association study identifies two new lung cancer susceptibility loci at 13q12.12 and 22q12.2 in Han Chinese. Nat Genet. 2011 Jul 3;43(8):792-6. doi: 10.1038/ng.875. PubMed PMID: 21725308.
3. Shiraishi K, Kunitoh H, Daigo Y, Takahashi A, Goto K, Sakamoto H, Ohnami S, Shimada Y, Ashikawa K, Saito A, Watanabe S, Tsuta K, Kamatani N, Yoshida T, Nakamura Y, Yokota J, Kubo M, Kohno T. A genome-wide association study identifies two new susceptibility loci for lung adenocarcinoma in the Japanese population. Nat Genet. 2012 Jul 15;44(8):900-3. doi: 10.1038/ng.2353. PubMed PMID: 22797724.
4. Brenner DR, Brennan P, Boffetta P, Amos CI, Spitz MR, Chen C, Goodman G, Heinrich J, Bickeböller H, Rosenberger A, Risch A, Muley T, McLaughlin JR, Benhamou S, Bouchardy C, Lewinger JP, Witte JS, Chen G, Bull S, Hung RJ. Hierarchical modeling identifies novel lung cancer susceptibility variants in inflammation pathways among 10,140 cases and 11,012 controls. Hum Genet. 2013 May;132(5):579-89. doi: 10.1007/s00439-013-1270-y. Epub 2013 Feb 1. Erratum in: Hum Genet. 2016 Aug;135(8):963. PubMed PMID: 23370545; PubMed Central PMCID: PMC3628758.
5. Kaufhold S, Bonavida B. Central role of Snail1 in the regulation of EMT and resistance in cancer: a target for therapeutic intervention. J Exp Clin Cancer Res. 2014 Aug 2;33:62. doi: 10.1186/s13046-014-0062-0. Review. PubMed PMID: 25084828; PubMed Central PMCID: PMC4237825.
6. Leão R, Apolónio JD, Lee D, Figueiredo A, Tabori U, Castelo-Branco P. Mechanisms of human telomerase reverse transcriptase (hTERT) regulation: clinical impacts in cancer. J Biomed Sci. 2018 Mar 12;25(1):22. doi:10.1186/s12929-018-0422-8. Review. PubMed PMID: 29526163.
7. UniProtKB - O14746 (TERT_HUMAN), UniProt (n.d.), retrieved from: http://www.uniprot.org/uniprot/O14746#function, accessed on March 14, 2018.
8. TERT telomerase reverse transcriptase [Homo sapiens (human)], NCBI (n.d.), retrieved from; https://www.ncbi.nlm.nih.gov/gene/7015, accessed on March 14, 2018.

9. Cao Y, Li H, Deb S, Liu JP. TERT regulates cell survival independent of telomerase enzymatic activity. Oncogene. 2002 May 9;21(20):3130-8. PubMed PMID:12082628.
10. rs2853677 SNP, Ensembl (n.d.), retrieved from: https://useast.ensembl.org/Homo_sapiens/Variation/Mappings?db=core;r=5:1286579-1287579;v=rs2853677;vdb=variation;vf=2171664, accessed on March 14, 2018.
11. SNPedia all SNPs (rs2853677), UCSC Genome Browser (n.d.), retrieved from: http://genome.ucsc.edu/cgi-bin/hgc?hgsid=659866703_tcxVSHSA8MDeWjJqEliY1QAcvD4Q&c=chr5&l=1272078&r=1302078&o=1287078&t=1287079&g=snpediaAll&i=rs2853677, accessed on March 14, 2018.
12. Reference SNP (refSNP) Cluster Report: rs2853677, dbSNP, NCBI (n.d.), retrieved from: https://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=2853677, accessed on March 14, 2018.
13. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000 Jan 7;100(1):57-70. Review. PubMed PMID: 10647931.
14. # 211980 LUNG CANCER, OMIM (n.d.), retrieved from: https://www.omim.org/entry/211980?search=211980%20&highlight=211980, accessed on March 16, 2018.
15. Gridelli C, Rossi A, Carbone DP, Guarize J, Karachaliou N, Mok T, Petrella F, Spaggiari L, Rosell R. Non-small-cell lung cancer. Nat Rev Dis Primers. 2015 May 21; 1:15009. doi: 10.1038/nrdp.2015.9. Review. PubMed PMID: 27188576.
16. Kovalchuk & Kovalchuck (page 130), Epigenetics in Health and Disease, FT Press, 2012.
17. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010 Dec 14;107(50):21931-6. doi: 10.1073/pnas.1016071107. Epub 2010 Nov 24. PubMed PMID: 21106759; PubMed Central PMCID: PMC3003124.
18. Elango N, Yi SV. Functional relevance of CpG island length for regulation of gene expression. Genetics. 2011 Apr;187(4):1077-83. doi:10.1534/genetics.110.126094. Epub 2011 Feb 1. PubMed PMID: 21288871; PubMed Central PMCID: PMC3070517.
19. Mazzolini R, Gonzàlez N, Garcia-Garijo A, Millanes-Romero A, Peiró S, Smith S, García de Herreros A, Canudas S. Snail1 transcription factor controls telomere transcription and integrity. Nucleic Acids Res. 2018 Jan 9;46(1):146-158. doi:10.1093/nar/gkx958. PubMed PMID: 29059385; PubMed Central PMCID: PMC5758914.
20. Luke B, Lingner J. TERRA: telomeric repeat-containing RNA. EMBO J. 2009 Sep 2;28(17):2503-10. doi: 10.1038/emboj.2009.166. Epub 2009 Jul 23. Review. PubMed PMID: 19629047; PubMed Central PMCID: PMC2722245.
21. UniProtKB - O95863 (SNAI1_HUMAN), UniProt (n.d.), retrieved from: http://www.uniprot.org/uniprot/O95863, accessed on March 18, 2018.
22. Bell RJ, Rube HT, Kreig A, Mancini A, Fouse SD, Nagarajan RP, Choi S, Hong C, He D, Pekmezci M, Wiencke JK, Wrensch MR, Chang SM, Walsh KM, Myong S, Song JS, Costello JF. Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. Science. 2015 May 29;348(6238):1036-9. doi:10.1126/science.aab0015. Epub 2015 May 14. PubMed PMID:25977370; PubMed Central PMCID: PMC4456397.
23. Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. Highly recurrent TERT promoter mutations in human melanoma. Science. 2013 Feb 22;339(6122):957-9. doi: 10.1126/science.1229259. Epub 2013 Jan 24. PubMed PMID: 23348506; PubMed Central PMCID: PMC4423787.
24. Sharrocks, A. D. (2013). Faculty of 1000 evaluation for TERT promoter mutations in familial and sporadic melanoma. F1000 - Post-publication peer review of the biomedical literature. doi:10.3410/f.717978928.793473666.
25. Galaxy, The Galaxy Project (n.d.), retrieved from: https://usegalaxy.org, accessed on April 2, 2018.
26. UCSC Genome Browser (n.d.), retrieved from: http://genome.ucsc.edu/, accessed on April 7, 2018.
27. SNPnexus, Barts Cancer Institute (n.d.), retrieved from: http://snp-nexus.org, accessed on April 6, 2018.
28. Variation Mappings, Ensembl (n.d.), retrieved from: https://uswest.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000164363;r=5:1225355-1246189;t=ENST00000324642, accessed on April 6, 2018.