PeterHuang_BINF7001_2024_Assignment2.pdf
ssh s4536711@binf-training.biosci.uq.edu.au

PART A
mkdir assessment2
cd assessment2
ln -s /opt/BINF7001/2024/Assignment2/normal_1.fq.gz .
ln -s /opt/BINF7001/2024/Assignment2/normal_2.fq.gz .
ln -s /opt/BINF7001/2024/Assignment2/tumour_1.fq.gz .
ln -s /opt/BINF7001/2024/Assignment2/tumour_2.fq.gz .
ln -s /opt/BINF7001/2024/Assignment2/a2.segments.bed .
ln -s /opt/BINF7001/2024/Assignment2/a2.segments.bed.gz .

bwa mem -R "@RG\tID:BINF7001\tSM:normal\tPL:ILLUMINA" ref/genome/Homo_sapiens_assembly38.fasta normal_1.fq.gz normal_2.fq.gz > normal.sam
bwa mem -R "@RG\tID:BINF7001\tSM:tumour\tPL:ILLUMINA" ref/genome/Homo_sapiens_assembly38.fasta tumour_1.fq.gz tumour_2.fq.gz > tumour.sam

samtools view -bS normal.sam > normal.bam
samtools view -bS tumour.sam > tumour.bam
samtools sort normal.bam > normal.sorted.bam
samtools index normal.sorted.bam
samtools sort tumour.bam > tumour.sorted.bam
samtools index tumour.sorted.bam

java -jar /opt/picard/picard.jar MarkDuplicates I=normal.sorted.bam O=normal.markdups.bam M=normal.markdups.metrics
samtools index normal.markdups.bam
java -jar /opt/picard/picard.jar MarkDuplicates I=tumour.sorted.bam O=tumour.markdups.bam M=tumour.markdups.metrics
samtools index tumour.markdups.bam

/opt/java7/bin/java -jar /opt/GATK/GenomeAnalysisTK.jar -T RealignerTargetCreator -L a2.segments.bed -R ref/genome/Homo_sapiens_assembly38.fasta -I normal.markdups.bam -o normal.intervals

/opt/java7/bin/java -jar /opt/GATK/GenomeAnalysisTK.jar -T RealignerTargetCreator -L a2.segments.bed -R ref/genome/Homo_sapiens_assembly38.fasta -I tumour.markdups.bam -o tumour.intervals

/opt/java7/bin/java -jar /opt/GATK/GenomeAnalysisTK.jar -T IndelRealigner -L a2.segments.bed -R ref/genome/Homo_sapiens_assembly38.fasta -I normal.markdups.bam -o normal.realigned.bam -targetIntervals normal.intervals --disable_bam_indexing

/opt/java7/bin/java -jar /opt/GATK/GenomeAnalysisTK.jar -T IndelRealigner -L a2.segments.bed -R ref/genome/Homo_sapiens_assembly38.fasta -I tumour.markdups.bam -o tumour.realigned.bam -targetIntervals tumour.intervals --disable_bam_indexing

samtools index normal.realigned.bam
samtools index tumour.realigned.bam

samtools flagstat normal.realigned.bam
592951 total
2514 duplicates
591538 aligned (99.76%)
1413 unaligned (0.24%)

samtools flagstat tumour.realigned.bam
582343 total
2472 duplicates
580928 mapped (99.76%)
1415 + 0 singletons (0.24%)

B.1

Haplotype SNV caller
/opt/gatk-4.4.0.0/gatk HaplotypeCaller \ -R ref/genome/Homo_sapiens_assembly38.fasta \ -I normal.realigned.bam \ -O normal_germline.vcf.gz \ -L a2.segments.bed

Annotate VCF with gnomAD
/opt/jdk-20.0.2/bin/java -jar /opt/snpEff/SnpSift.jar annotate ref/gnomad/af-only-gnomad.hg38.vcf.gz normal_germline.vcf.gz > normal_germline_annotated.vcf.gz

Run SnpEff for functional impact annotation
/opt/jdk-20.0.2/bin/java -jar /opt/snpEff/snpEff.jar -v GRCh38.105 normal_germline_annotated.vcf.gz > normal_germline_annotated.vcf.gz.snpeff.vcf

Count variants with and without an ID
zcat normal_germline_annotated.vcf.gz | grep -v "^#" | awk '{ if($3 != ".") count++ } END { print count " variants with ID" }'
1325 variants with ID
zcat normal_germline_annotated.vcf.gz | grep -v "^#" | awk '{ if($3 == ".") count++ } END { print count " variants without ID" }'
273 variants without ID

B.2

Filter for variants with MODERATE or HIGH impact
cat normal_germline_annotated.vcf.gz.snpeff.vcf | /opt/jdk-20.0.2/bin/java -jar /opt/snpEff/SnpSift.jar filter "((ANN[*].IMPACT has 'MODERATE') | (ANN[*].IMPACT has 'HIGH')) & (na ID) & (QUAL > 30)" > filtered_moderate_high_impact.vcf

Extract INDELs from the VCF
bcftools --gzvcf normal_germline_annotated.vcf.gz --keep-only-indels --recode --stdout > indels_filtered.vcf

Filter INDELs for MODERATE or HIGH impact
/opt/jdk-20.0.2/bin/java -jar /opt/snpEff/SnpSift.jar filter "((ANN[*].IMPACT has 'MODERATE') | (ANN[*].IMPACT has 'HIGH')) & (na ID) & (QUAL > 30)" indels_filtered.vcf > indels_moderate_high_impact.vcf

Generate combined mutation report
cat filtered_moderate_high_impact.vcf indels_moderate_high_impact.vcf > combined_moderate_high_impact.vcf

SNV Table (cat combined_moderate_high_impact.vcf)

Chromozone	Location	Impact	Gene	Amino Acid Change	Allele Frequency
1	207022488	MODERATE	C1orf116	p.Pro426Ser	0.5
2	167251333	MODERATE	XIRP2	p.Ala3092Glu	0.5
6	152249164	MODERATE	SYNE1	p.Ile6523Met	0.5
7	134127432	MODERATE	LRGUK	p.Arg22Ile	0.5
10	95067642	MODERATE	CYP2C8	p.Pro73Leu	0.5
12	12187027	MODERATE	LRP6	p.Ile247Asn	0.5
17	16165085	MODERATE	NCOR1	p.Ala171Val	4.923e-05
20	9566203	MODERATE	PAK5	p.Ser391Ile	0.5
X	41346375	MODERATE	DDX3X	p.Arg487Cys	0.5

B.3

SV calling with Manta
/opt/manta-1.6.0.centos6_x86_64/bin/configManta.py --bam=normal.realigned.bam --referenceFasta=ref/genome/Homo_sapiens_assembly38.fasta --runDir=normal_manta3 --callRegions=a2.segments.bed.gz

normal_manta3/runWorkflow.py -j 1

bcftools view normal_manta3/results/variants/diploidSV.vcf.gz

/opt/jdk-20.0.2/bin/java -jar /opt/snpEff/snpEff.jar -v GRCh38.105 normal_manta3/results/variants/diploidSV.vcf.gz > normal_manta3/results/variants/diploidSV_annotated.vcf.gz

bcftools view normal_manta3/results/variants/diploidSV_annotated.vcf.gz

SV Table

Chr	Location	Type	Reference Sequence	Alternate	Gene
17	16145408	Deletion	Large sequence starting with "AGTGAG..."	Single base "A"	NCOR1 ENSG00000141027
17	39481631	Deletion	Long repeated sequence "GCGCTC..."	Single base "G"	CDK12 ENSG00000167258
17	43043808	Deletion	Single base "T"	<DEL> Placeholder for a large deletion	BRCA1 ENSG00000012048

B.4
bcftools query -f '%CHROM\t%POS\t%REF\t%ALT\t%AF\n' combined_moderate_high_impact.vcf

For most of the entries, the allele frequency around 0.5 suggests that the variant is heterozygous. In the context of normal tissue, this means the individual carries one copy of the reference allele and one copy of the mutant allele for these genes. This is common in germline variants. The low allele frequency for the NCOR1 variant (4.923e-05) suggests it is either a rare germline variant or a sequencing artifact, as it is far below the expected AF for a heterozygous variant.

B.5

Gene	Function	Relevence to Cancer	Reference
C1orf116	Promotes tight junctions and inhibits cell movement	Less expressed in metastatic prostate cancer tumors	(Amend & Pienta, 2018)
XIRP2	Inhibit the proliferation of colon cancer cells in vitro and in vivo	Less expressed in colon cancer tissue	(Zhou et al., 2019)
SYNE1	Encodes a nuclear envelope protein, nesprin-1, critical for connecting the nucleus to the cytoskeleton	Less expressed in ovarian cancer	(Harbin et al., 2023)
LRGUK	LRGUK binds to HOOK2 and functions in sperm tail axoneme extension. LRGUK is vital for haploid germ cell development and male fertility.	Decreased expression leads to abnormal sperm head shaping	(Liu et al., 2015)
CYP2C8	Metabolism of tamoxifen, paclitaxel, and other chemotherapy	Variants influence breast tumour characteristics and disease-free survival in tamoxifen-treated patients.	(Jernström et al., 2009)

LRP6	Regulate cell differentiation, proliferation, and migration and the development of many cancer types.	Overexpressed in colorectal, liver, breast and pancreatic adenocarcinomas. Reducing LRP6 expression inhibits cancer cell proliferation and delays tumour growth.	(Raisch et al., 2019)
NCOR1	Encode a protein that mediates ligand-independent transcription repression of thyroid-hormone by promoting chromatin condensation and preventing access of the transcription machinery.	High expression of *NCOR1* mRNA have a better breast cancer prognosis. NCOR1 mRNA was expressed at higher levels in patients over 50 years of age.	(Zhang et al., 2006)
PAK5	Regulation of cytoskeletal dynamics, proliferation, and cell survival signaling	High expression in colorectal cancer tissues progression. Inhibition of PAK5 led to restrained tumor cell growth.	(Huang et al., 2020)
DDX3X	Involved in translation, cellular signaling, and viral replication.	Overexpressed in lung cancer, breast cancer and other cancers.	(Mo et al., 2021)
CDK12	Regulate transcriptional and post-transcriptional processes, modulating multiple cellular functions.	Genomic alterations in CDK12 have been detected in breast, ovarian, colorectal and pancreatic cancers.	CDK12: an emerging therapeutic target for cancer
BRCA1	Maintaining genomic stability, and it also acts as a tumor suppressor	Mutations in this gene are responsible for more than 80% of inherited breast and ovarian cancers.	NCBI

C.1

Generating raw calls

```
/opt/gatk-4.4.0.0/gatk Mutect2 -R ref/genome/Homo_sapiens_assembly38.fasta -I tumour.realigned.bam -I normal.realigned.bam -normal normal -tumor tumour --germline-resource ref/gnomad/af-only-gnomad.hg38.vcf.gz -O A2.mutect.vcf.gz -L a2.segments.bed

zcat A2.mutect.vcf.gz | grep -v ^# | wc -l
15
```

Generate tumour and normal pileups which will be used for tumour purity estimation:

```
/opt/gatk-4.4.0.0/gatk GetPileupSummaries -I tumour.realigned.bam -V /opt/BINF7001/2024/Prac5_2024/ref/hapmap_3.3.hg38.vcf.gz -L /opt/BINF7001/2024/Prac5_2024/ref/hapmap_3.3.hg38.vcf.gz -O tumour.pileups.table

/opt/gatk-4.4.0.0/gatk GetPileupSummaries -I normal.realigned.bam -V /opt/BINF7001/2024/Prac5_2024/ref/hapmap_3.3.hg38.vcf.gz -L /opt/BINF7001/2024/Prac5_2024/ref/hapmap_3.3.hg38.vcf.gz -O normal.pileups.table
```

Use the pileup summaries to estimate tumour purity (a.k.a. normal contamination):

```
/opt/gatk-4.4.0.0/gatk CalculateContamination -I tumour.pileups.table -matched normal.pileups.table -O paired.contamination.table

cat paired.contamination.table
```

Apply FilterMutectCalls to generate a list of filtered calls:

```
/opt/gatk-4.4.0.0/gatk FilterMutectCalls -V A2.mutect.vcf.gz --contamination-table paired.contamination.table -O A2.mutect.filtered.vcf.gz -R ref/genome/Homo_sapiens_assembly38.fasta

bcftools query -f '%CHROM\t%POS\t%REF\t%ALT\t%FILTER\n' A2.mutect.filtered.vcf.gz | grep -v 'PASS'
bcftools query -f '%FILTER\n' A2.mutect.filtered.vcf.gz | grep -v 'PASS' | wc -l
```

bcftools query -f '%CHROM\t%POS\t%REF\t%ALT\t%FILTER\n' A2.mutect.filtered.vcf.gz

Annotate VCF
/opt/jdk-20.0.2/bin/java -jar /opt/snpEff/SnpSift.jar annotate ref/gnomad/af-only-gnomad.hg38.vcf.gz A2.mutect.filtered.vcf.gz > A2.mutect.filtered.annotated.vcf.gz

Run SnpEff for functional impact annotation
/opt/jdk-20.0.2/bin/java -jar /opt/snpEff/snpEff.jar -v GRCh38.105 A2.mutect.filtered.annotated.vcf.gz > A2.mutect.filtered.annotated.snpeff.vcf.gz

View VCF
zcat A2.mutect.filtered.annotated.vcf.gz
bcftools view A2.mutect.filtered.annotated.vcf.gz | grep -v "^#" | wc -l
bcftools query -f '%CHROM\t%POS\t%REF\t%ALT\t%FILTER\n' A2.mutect.filtered.vcf.gz
bgzip A2.mutect.filtered.annotated.vcf.gz
tabix -p vcf A2.mutect.filtered.annotated.vcf.gz

cat A2.mutect.filtered.annotated.snpeff.vcf.gz | grep -v ^# | grep PASS

C.1 SNV table focuses only the significant variants after filtering out weak evidence.

Chr	Position	REF	ALT	Impact	Gene	Amino Acid Change	VAF
12	25245350	C	T	missense_variant | MODERATE	KRAS	p.Gly12Asp	0.456
13	48307387	C	A	stop_gained | HIGH	RB1	p.Ser82 * (premature stop codon)	0.414
17	7669677	T	A	stop_gained | HIGH	TP53	p.Lys372*	0.440
17	7676214	TG	T	frameshift_variant | HIGH	TP53	p.Gln52fs	0.466
17	39462169	G	A	missense_variant | MODERATE	CDK12	p.Arg33Lys	0.288
17	43045759	C	T	stop_gained | HIGH	BRCA1	p.Trp1858*	0.395

C.1 SV calling with Manta
/opt/manta-1.6.0.centos6_x86_64/bin/configManta.py --bam=tumour.realigned.bam --referenceFasta=ref/genome/Homo_sapiens_assembly38.fasta --runDir=tumor_manta3 --callRegions=a2.segments.bed.gz

tumor_manta3/runWorkflow.py -j 1

bcftools view tumor_manta3/results/variants/diploidSV.vcf.gz
/opt/jdk-20.0.2/bin/java -jar /opt/snpEff/snpEff.jar -v GRCh38.105 tumor_manta3/results/variants/diploidSV.vcf.gz > tumor_manta3/results/variants/diploidSV_annotated.vcf.gz

bcftools view tumor_manta3/results/variants/diploidSV_annotated.vcf.gz

C.1 SV (Structural Variation) table

Chr	Position	REF	ALT	Impact	Gene
13	48299128	T	<DEL>	Deletion | feature_truncation | HIGH	RB1
17	43043808	T	<DEL>	Deletion | feature_truncation | HIGH	BRCA1

C.2

```
bcftools isec -p outdir -n=2 A2.mutect.filtered.annotated.vcf.gz ref/gnomad/af-only-gnomad.hg38.vcf.gz
bcftools view 0000.vcf | grep -v "^#" | wc -l # Variants unique to A2.vcf.gz
bcftools view 0001.vcf | grep -v "^#" | wc -l # Variants unique to gnomAD
cat 0000.vcf
cat 0001.vcf
```

There should be no overlap between somatic variants detected in the tumor sample and germline variants listed in gnomAD. This is because somatic mutations are specific to tumor cells and should not be present in the individual's germline DNA, which is inherited and found in all cells. The purpose of using a germline database like gnomAD is to exclude known germline mutations, so somatic variant detection will identify only mutations unique to the tumor. Based on the two VCF files: 0000.vcf corresponds to a somatic variant detected in the tumor sample (A2.mutect.filtered.annotated.vcf.gz), specifically a G>A mutation at position 39,462,169 on chromosome 17. 0001.vcf represents the same variant in the gnomAD database (af-only-gnomad.hg38.vcf.gz), where it has a very low allele frequency (AF = 2.462e-05). Since this variant appears in both the tumor sample and gnomAD, it is considered a known germline variant that could be found in healthy individual at a low frequency. Occasionally, normal variants may be incorrectly classified as somatic due to sequencing noise.

C.3

In **Sample 2 (tumour tissue)**, the high VAF from the table in C.1 suggests the mutation is somatic and possibly contributes to the cancer, as tumour cells carry both the reference and mutated alleles.

C.4:

KRAS is a gene that encodes a protein belonging to the small GTPase superfamily, which plays a role in regulating cell division. A single amino acid substitution in the KRAS protein can lead to an activating mutation, which causes the protein to be constantly "switched on." This mutation drives uncontrolled cell growth and proliferation, contributing to the development and progression of various cancers. The mutated KRAS protein is commonly associated with pancreatic, colorectal, lung, and thyroid cancers (COSMIC). In pancreatic adenocarcinoma specifically, KRAS mutations are prevalent, with alterations observed in 86.35% of patients (MyCancerGenome). This high frequency emphasizes the significant effect of KRAS mutations in the development of pancreatic cancer and its potential as a therapeutic target.

RB1 is a tumor suppressor gene that serves as a model for understanding how the inactivation of such genes can drive tumor development. Mutations in RB1 are linked to cancers such as breast cancer, and small cell lung carcinoma (NCBI). In this analysis, the SNV table identifies a stop-gained mutation at chromosome 13, position 48,307,387, which leads to the loss of gene function. Additionally, the SV table highlights a deletion in the RB1 gene at chromosome 13, position 48,299,128. Specifically, the p.Ser82* mutation means that at position 82 of the RB1 protein, the serine (Ser) is replaced by a premature stop codon (*), causing the protein to be truncated. This truncation often renders the protein non-functional or only partially functional, contributing to disease progression.

TP53 encodes a tumor suppressor protein with key functions in transcriptional activation, DNA binding, and oligomerization. The protein respond to cellular stresses by regulating the expression of target genes, leading to processes such as cell cycle arrest, apoptosis, senescence, DNA repair, and metabolic changes. Mutations in this gene are linked to breast, colorectal, lung, and sarcoma tumors (COSMIC). A mutation at the 52nd position of the protein causes a frameshift, disrupting the normal amino acid sequence and introducing a premature stop codon. This results in a truncated, likely non-functional protein, which has major biological impacts, especially in tumor suppressor genes like TP53, and contributes to the development of cancers such as breast cancer.

CDK12 facilitates cyclin binding and RNA polymerase II CTD heptapeptide repeat kinase activity. It plays a role in phosphorylating the C-terminal domain of RNA polymerase II, protein autophosphorylation, and the regulation of MAP kinase activity (NCBI). Inhibition of CDK12/CDK13 activates intronic polyadenylation site cleavage, which reduces the expression of essential DNA damage response proteins. This leads to a

"BRCAness" phenotype, characterized by impaired DNA repair mechanisms, thereby enhancing the effects of DNA-damaging chemotherapy and contributing to the progression of triple-negative breast cancer (Quereda et al., 2019).

BRCA1 encodes a 190 kD nuclear phosphoprotein that is important for maintaining genomic stability and functions as a tumor suppressor. Mutations in this gene are responsible for about 40% of inherited breast cancers and over 80% of inherited breast and ovarian cancers (NCBI, COSMIC). The structural variant (SV) table also identifies a deletion at chr17:43043808, indicating that both point mutations and structural changes contribute to the loss of function in key genes like this one. The p.Trp1858* mutation signifies that at position 1858 of the protein, tryptophan is replaced by a stop codon, leading to protein truncation. This results in a non-functional protein, impairing the gene's tumor-suppressing role, which is closely linked to cancer development. Loss of the tumor suppressor RB1 is frequently observed in Triple-Negative Basal-like breast cancer (Jiang et al., 2011).

C.5:

MIRA-1 and PRIMA-1MET are two drugs that specifically target the TP53 gene, which encodes the p53 protein, a well-known tumor suppressor. This protein plays a significant role in controlling the cell cycle and preventing tumor development. Both MIRA-1 and PRIMA-1MET are involved in the p53 pathway, suggesting their mechanism of action likely involves reactivating or modulating the function of p53. This reactivation can restore p53's ability to induce cell cycle arrest or apoptosis in cancer cells that have lost proper p53 function, making these drugs promising therapeutic options in cancers with TP53 mutations (CancerRxGene).

C.6:

As seen from the SNV, SV result tables and the research articles, mutations in BRCA1 and BRCA2 are associated with an increased risk of breast and ovarian cancer. The BRCA1 stop-gained and deletion variants found in this case would significantly impair the gene's tumor suppressor function, leading to loss of DNA repair capability and a heightened risk of cancer development. TP53, another tumor suppressor gene frequently mutated in breast cancer, is involved in preventing uncontrolled cell division. The observed TP53 mutations, especially the stop-gained and frameshift variants, would likely result in a loss of p53 protein function, facilitating unchecked cellular proliferation.

Furthermore, the RB1 gene regulates the cell cycle, also has truncation or deletion mutations in this context. These alterations are commonly associated with breast cancer, where RB1 loss contributes to deregulated cell division. KRAS mutations, though more common in colorectal, pancreatic, and lung cancers, can also be found in breast cancers, further supporting a potential breast cancer diagnosis in this genetic profile.

The combination of BRCA1 and TP53 mutations, along with RB1 deletion, points toward a high likelihood of breast cancer. The presence of structural variants in RB1 and BRCA1, coupled with damaging TP53 missense and frameshift mutations, reinforces the suspicion of a hereditary cancer syndrome, making breast cancer the most probable diagnosis.

Reference List

Amend, S.R. and Pienta, K.J. (2018) 'Abstract 3156: C1orf116, an uncharacterized gene with unknown function, restricts movement capacity of would-be metastatic cells through modulation of tight junctions', *Cancer Research*, 78(13_Supplement), pp. 3156–3156. doi:10.1158/1538-7445.am2018-3156.

Cancerrxgene - Genomics of Drug Sensitivity in Cancer. (n.d.). https://www.cancerrxgene.org/

COSMIC - Catalogue of Somatic Mutations in cancer. (n.d.). https://cancer.sanger.ac.uk/cosmic

Harbin, L. *et al.* (2023) 'Syne1 mutation is associated with increased tumor mutation burden and immune cell infiltration in ovarian cancer (1263)', *Gynecologic Oncology*, 176. doi:10.1016/j.ygyno.2023.06.172.

Huang, S. *et al.* (2020) 'Pak5 facilitates the proliferation, invasion and migration in colorectal cancer cells', *Cancer Medicine*, 9(13), pp. 4777–4790. doi:10.1002/cam4.3084.

Jernström, H. *et al.* (2009) 'CYP2C8 and CYP2C9 polymorphisms in relation to tumour characteristics and early breast cancer related events among 652 breast cancer patients', *British Journal of Cancer*, 101(11), pp. 1817–1823. doi:10.1038/sj.bjc.6605428.

Jiang, Z. *et al.* (2011) 'RB1 and p53 at the crossroad of EMT and triple-negative breast cancer', *Cell Cycle*, 10(10), pp. 1563–1570. doi:10.4161/cc.10.10.15703.

Liu, Y. *et al.* (2015) 'LRGUK-1 is required for basal body and Manchette function during spermatogenesis and male fertility', *PLOS Genetics*, 11(3). doi:10.1371/journal.pgen.1005090.

Lui, G.Y., Grandori, C. and Kemp, C.J. (2018) 'CDK12: An emerging therapeutic target for cancer', *Journal of Clinical Pathology*, 71(11), pp. 957–962. doi:10.1136/jclinpath-2018-205356.

Mo, J. *et al.* (2021) 'DDX3X: Structure, physiologic functions and cancer', *Molecular Cancer*, 20(1). doi:10.1186/s12943-021-01325-7.

National Center for Biotechnology Information. Available at: https://www.ncbi.nlm.nih.gov/

Quereda, V. *et al.* (2019) 'Therapeutic targeting of Cdk12/CDK13 in triple-negative breast cancer', *Cancer Cell*, 36(5). doi:10.1016/j.ccell.2019.09.004.

Raisch, J., Côté-Biron, A. and Rivard, N. (2019) 'A role for the WNT co-receptor LRP6 in pathogenesis and therapy of epithelial cancers', *Cancers*, 11(8), p. 1162. doi:10.3390/cancers11081162.

Zhang, Z. *et al.* (2006) 'NCOR1 mrna is an independent prognostic factor for breast cancer', *Cancer Letters*, 237(1), pp. 123–129. doi:10.1016/j.canlet.2005.05.046.

Zhou, F. *et al.* (2019) 'lncrna XIRP2-AS1 predicts favorable prognosis in colon cancer</p>', *OncoTargets and Therapy*, Volume 12, pp. 5767–5778. doi:10.2147/ott.s215419.