Section 1. Life Science

https://doi.org/10.29013/ELBLS-23-1-3-9

Mingxiao Ma,
Dr. Roberto Aguilar,
Gaston Day School

USING HIGH-THROUGHPUT SNP TECHNOLOGIES TO IDENTIFY VARIANCES ASSOCIATED WITH BLADDER CANCER

Abstract. According to the CDC, about 57,000 men and 18,000 women have bladder cancer, and about 12,000 men and 4,700 women die from it each year in the US. These high penetrance traits and the identification of the bladder cancer-predisposing gene *CCL4,* warrant further investigation. In this study, we looked at single nucleotide polymorphisms (SNPs) and their association with bladder cancer. Whole genome sequencing data was obtained from the Sequence Read Archive and then analysis pipelines were constructed to map the sequences against chromosome 17 followed by indexing and variant calling. Through our analytical tools, we were able to detect variants in bladder cancer tissues that were not present in normal sequence samples. These findings help support previous studies that link CCL4 with genetic susceptibility to bladder cancer.

Keywords: bladder, cancer, CCL4, SNPs.

Introduction

Bladder Cancer

Bladder cancer is a common cancer that occurs when the cells within the bladder hyper proliferate and form tumors which can then metastasize to other parts of the body. This type of cancer commonly begins in the innermost lining of the bladder (urothelium or transitional epithelium). There are different types of bladder cancer which include: 1) urothelial carcinoma, considered a transitional cell carcinoma and is the most common one, 2) squamous cell carcinoma, the subtype encountered only 1–2% of the time, 3) adenocarcinoma, less common at 1%, and 4) small cell carcinoma (less than 1%), and 5) sarcoma (less than 1%) (American Cancer Society, 2019).

According to the CDC, about 57,000 men and 18,000 women have bladder cancer, and about 12,000 men and 4,700 women die from the disease each year in the US. According to American Cancer Society data in 2022, there will be about 81,180 new cases of bladder cancer and 17,100 deaths from bladder cancer. Figure 1 shows the number of bladder cancer cases from 1999 to 2019. As you can see, the number increases in a relative linear relation up to the 60,000 which was reached in 2000. These data seem to support the evidence that bladder cancer is one of the most common cancers in the United States.

Stages

Bladder cancer progression is separated into stages. The staging system used most often is called the Tumor Nodes Metastases System which is based on 3 keys: "T"

used to describe the distance from the main (primary) tumor has grown through the bladder wall to nearby tissues, "N" used to describe any cancer cells spread to lymph nodes around the bladder, "M" usually indicates if the cancer cells have reached other sites such as distant organs (American Cancer Society, 2019).

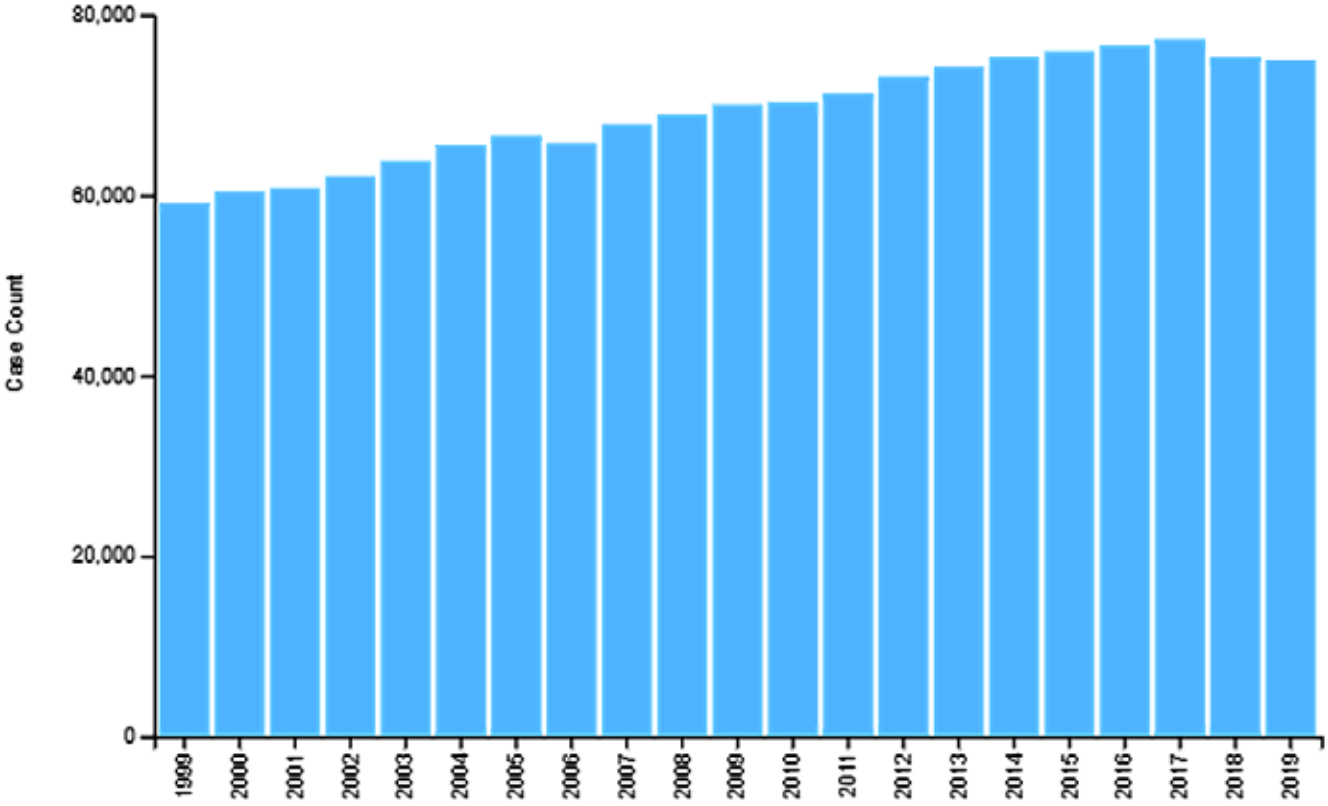

Figure 1: Annual number of new bladder cancer cases from 1999 to 2019 in the United States. This graph shows overall data of new bladder cancer cases from 1999 to 2019. By the year 1999, the case number had already reached 60,000 cases and it kept increasing every year. Obtained from the U. S. Cancer Statistics Working Group. U. S. Cancer Statistics Data Visualizations Tool (2022).

However, there is another numbering system which is not often used that ranges from stage 0 to stage 4. Stage 0 is the earliest stage in which cancer cells are found within the inner layer of the bladder. In stage I, cancer cells can be found growing within the connective tissue beneath the bladder lining. If the cancer cells continue to grow within the connective tissue layer and reach the muscle of the bladder wall, stage II has been reached. Stage III and IV are the two most advanced conditions of bladder cancer, with cancer cells growing through the muscle and into the fat layer, then invading other organs such as the womb, vagina, or the prostate (stage III). In stage IV, the cancer cells can metastasize to nearly all parts of the bladder and then invade other organs (Cancer Research UK, 2018) [5].

Treatment

Treatments of bladder cancer are administered based on the stage it is in. Stage I is most often treated with transurethral resection (TURBT), which is a surgery to remove the tumor from the bladder through the urethra, and intravesical therapy treatment with fulguration 24 hours after surgery (Staff [13]). The side effects of the TURBT include bloody urine, uncomfortable urination, and bladder infection. Intravesical therapy is a major treatment in which the liquid drug is administered into the bladder through a soft catheter and left for up to 2 hours (American Cancer Society [11]). Some common side effects of intravesical therapy include irritation and a burning sensation in the bladder, and bloody urine. The type

of radiation that is often used to combat bladder cancer is named External Beam Radiation Therapy and is used during the early stages of bladder cancer in patients who cannot undergo surgery or chemotherapy due to other complications. This type of radiation therapy can effectively prevent symptoms that are caused by advanced bladder cancer. The possible side effects include: nausea, vomiting, diarrhea, low blood counts, skin changes in the radiation area and bloody urine. Chemotherapy is used to treat bladder cancer where cancer cells are confined to the lining of the bladder and will have a high risk of progression or recurrence to a higher metastasis state. Side effects of chemotherapy include blood in the urine and frequent urination (American Cancer Society [13]).

CCL4

CCL4 (C–C motif chemokine ligand 4) is a gene which is protein-coded located on 17q12, Gene ID6351 (NCBI, 2022). The *CCL4* gene has been found in chimpanzees, rhesus monkeys, dogs, cows, mice, rats, and chickens. The protein encoded by *CCL4* is a mitogen-inducible monokine that is secreted and has chemokinetic and inflammatory functions. Also, It has been known to be one of the major HIV-suppressive factors produced by CD8+ T cells (NCBI [14]). Particularly when it comes to urothelial bladder cancer, one of the most common cancers with a high mortality and recurrence rate, CCL4 plays a role in bladder cancer or its therapy. CCL4 has only been identified as being partially regulated by CC chemokines and being implicated in angiogenesis, tumor development, and metastasis of many malignancies. However, little is known about the function of distinct CC chemokines in bladder cancer.

SNPs

Single nucleotide polymorphisms (SNPs) are the most common type of genetic variation among people (National Library of Medicine, [10]). Each SNP denotes a variation in a single nucleotide, or DNA building block. An example of an SNP is the substitution of the nucleotide thymine (T) with the nucleotide cytosine (C) in a specific DNA stretch. SNPs can be used as genetic markers in analysis, and some SNPs located within a gene may directly affect the protein structure or gene expression levels. SNPs are common throughout the genome and can therefore be used in association studies, which postulate that two closely spaced alleles (a gene and a marker) are inherited together, as indications to find disease-causing genes (Shastry [12]). However, SNPs within genes that control telomere maintenance, mitosis, inflammation, and apoptosis, however, have not been thoroughly studied in bladder cancer.In this investigation, we used bioinformatics tools to identify these gene variances within the *CCL4* gene and other regions of chromosome 17. We believe that any variances found in a bladder cancer clinical cohort that are not found in a normal cohort may have an association with the observed phenotypic tendencies of bladder cancer.

Methods

To identify a good candidate gene, we accessed NCBI's Gene Expression Omnibus (GEO) and sifted through microarray data. Once we settled on a published bladder cancer investigation, we took the accession number and entered it into the GEO2R database to identify genes with high expression. Through the careful analysis of each generated graph, the gene *CCL4* was identified as a candidate with high expression in bladder cancer samples but low expression in normal samples. We then accessed the Sequence Read Archive (SRA) database and identified a bladder cancer clinical cohort composed of a total of 105 individuals. Ensembl's FTP database was used to obtain chromosome 17's sequence to be used as a reference. The fastq-dump tool was used to access the SRA database and download single end reads as fastq files which were then quality checked using the Babraham Institute's FastQC program. The resultant HTML file was then evaluated to ascertain whether the bar graphs principally clustered within the green field. Trimmomatic was used to trim any fastq file reads that fell below a Phred score of 40 and then reanalyzed with FastQC. Bowtie 2 was used to map the reads to the reference chromosome 17. The

sequences were then indexed and locally aligned with the Burrows-Wheeler Aligner software package (bwa index and bwa mem). Next, The Sequence Alignment Map (SAM) tools were used to convert SAM output files into compressed Binary Alignment Map (BAM) files and then sort the files according to specific coordinates. BCF tools was used to call and view the detected SNPs, filter, and then convert the files into variant calling format (VCF). Python's Pandas library was used to eliminate any variances found in both the bladder cancer clinical cohort and the normal cohort. The filtered variances were then compared by chromosomal position to NCBI's Entrez Database to identify the variances' SNP accession number. Ensembl's Variant Effect Predictor was then used to identify the genotypic effect of the identified SNPs.

Results

The analysis of chromosome 17 resulted in the identification of hundreds of variants which were common in at least 70% of the individuals in the bladder cancer cohort. After removing those that were also found in the normal population and were not currently characterized as SNPs, only 14 remained. These were then further evaluated using the Variant Effect Predictor which detected the following effects: 1) Intron variant- impacts alternative splicing by interfering with the splice site recognition (Lin [7]). 2) non coding transcript variant- potentially results in the loss of the start or stop codons (Dhamija [5]). 3) nonsense-Mediated mRNA Decay (NMD) transcript variant- targets on the nonsense-mediated mRNA decay function, which under typical circumstances decreases gene expression mistakes by removing mRNA transcripts with early stop codons. 4) regulatory region variant- variant within the non-coding genomic region that has been demonstrated to be associated with different diseases. 5) upstream and downstream gene variants – variants that can affect a gene's regulatory region and may cause damage to the gene. 6) transcription factor (TF) binding site variant- a sequence variant located within a transcription factor binding site which is involved in the regulation of transcription.

As stated earlier, our analysis detected variances within chromosome 17 totaling 13526. After filtering for characterized SNPs that were not found in the normal cohort, we were left with a total of 14 which are listed in Table 1 below. According to VEP results, ten of the 14 SNPs have at least one effect which may be associated with bladder cancer.

Table 1.– Analysis of the bladder cancer cohort identified 14 SNPs with various genetic effects. Table of the gene's location and the SNPs that correspond with and have effects on bladder cancer: Positions on chromosome 17 were entered in the Entrez database to obtain the SNP accession number. Those variants that were not found to be characterized are marked as "None." Variant effects were listed as determined by the Entrez database

Location	SNPs	Effect
1	***2***	***3***
29959924	rs1243604199	Intron variant/Non coding transcript variant/NMD transcript variant/Regulatory region variant
63057821	rs2035858461	**None**
49418233	rs2070798385	**None**
63048721	rs1305782789	Intron variant
63048685	rs1598312390	Intron variant
22521491	rs80284949	Upstream gene variant/Downstream gene variant/Regulatory region variant/TF binding site variant
63034059	rs2034878591	**None**
49421388	rs778329138	**None**

1	*2*	*3*
49421406	rs1438950654	Intron variant/Non coding transcript variant/Upstream gene variant/Downstream gene variant
62998692	rs2033231845	None
22521506	rs80137394	Upstream gene variant/Downstream gene variant/Regulatory region variant/TF binding site variant
22521503	rs77770974	Upstream gene variant/Downstream gene variant/Regulatory region variant/TF binding site variant
12119297	rs1391768914	Intron variant/Non coding transcript variant/NMD transcript variant/Regulatory region variant/Downstream gene variant
22522609	rs201160689	Upstream gene variant/Downstream gene variant/Regulatory region variant/TF binding site variant

Discussion

Our analysis shows that there is a strong association between the identified 14 SNPs and bladder cancer. Although many variants were identified, not all met the requirements for a common SNP. Still further analysis is needed in order to have a complete picture. Future experiments can involve the analysis of all 23 chromosomes instead of just limiting it to chromosome 17. Further research must also be conducted to correlate these findings with *in vivo* data. Initial DNA can be collected and then purified for sequencing from a clinical cohort. This can be repeated every 10 years as they are simultaneously observed for phenotypic signs of bladder cancer. In order to analyze all 23 chromosomes, at least 30 TB of storage space is required. This coupled with the demand for random access memory places limits on computing power.

Conclusion

Our analyses were able to detect several variances within chromosome 17. After filtering our results for those that were characterized as SNPs on NCBI's Entrez database and those that were not found in normal sequences, we were left with 14 SNPs. Although half were predicted to not have detectable genetic effects, the other half shared genetic consequences ranging from a regulatory region variant to NMD. These SNPs existed in all 105 patients' chromosome 17 and did not exist in normal human mutations. This association presents itself as highly plausible biomarkers that can be used in bladder cancer diagnosis and target treatments.

References:

1. Andrew A. S., Gui J., Sanderson A. C., Mason R. A., Morlock E. V., Schned A. R., Kelsey K. T., Marsit C. J., Moore J. H., & Karagas M. R. (Bladder cancer SNP panel predicts susceptibility and survival. Human genetics. 2009, June). Retrieved August, 16, 2022. From: URL: https: //www.ncbi.nlm.nih.gov/pmc/articles/PMC2763504/#:~: text=Bladder%20cancer%20is%20the%20fourth, assessed%20extensively%20for%20this%20disease
2. Centers for Disease Control and Prevention. Bladder cancer. Centers for Disease Control and Prevention. (2022, July 6). Retrieved August 16, 2022. From: URL: https: //www.cdc.gov/cancer/bladder/index.htm#:~: text=Statistics, women%20die%20from%20the%20disease
3. Database G. C. H.G. (n.d.). CCL4. GeneCards is a searchable, integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. Retrieved August 16, 2022. From: URL: https: //www.genecards.org/cgi-bin/carddisp.pl?gene=CCL4#:~: text=CCL4%20

Gene%20%2D%20C%2DC%20Motif%20Chemokine%20Ligand%204&text=The%20protein%20encoded%20by%20this, has%20chemokinetic%20and%20inflammatory%20functions
4. Deng N., Zhou H., Fan H., & Yuan Y. Single nucleotide polymorphisms and cancer susceptibility. (2017, November 7). Oncotarget. Retrieved August 16, 2022. From: URL: https: //www.ncbi.nlm.nih.gov/pmc/articles/PMC5746410/
5. Dhamija S., Menon M. B. Non-coding transcript variants of protein-coding genes – what are they good for? RNA Biol.– 15(8). 2018.– P. 1025–1031. Doi: 10.1080/15476286.2018.1511675. Epub 2018. Sep 10. PMID: 30146915; PMCID: PMC6161691.
6. Li Y., Chen X., Li D. et al. Identification of prognostic and therapeutic value of CC chemokines in Urothelial bladder cancer: evidence from comprehensive bioinformatic analysis. BMC Urol – 21, 2021.– 173 p. URL: https: //doi.org/10.1186/s12894-021-00938-w
7. Lin H., Hargreaves K. A., Li R. et al. RegSNPs-intron: a computational framework for predicting pathogenic impact of intronic single nucleotide variants. Genome Biol – 20, 2019.– 254 p. URL: https: //doi.org/10.1186/s13059-019-1847-4
8. Li H., Guo J., Cheng G., Wei Y., Liu S., Qi Y., Wang G., Xiao R., Qi W. and Qiu W. Identification and Validation of SNP-Containing Genes with Prognostic Value in Gastric Cancer via Integrated Bioinformatics Analysis. Front. Oncol.– 11. 2021.– 564296 p. Doi: 10.3389/fonc.2021.564296
9. Mayo Clinic. Bladder cancer. Mayo Clinic. (2022, April 19). Retrieved August 16, 2022. From: URL: https: //www.mayoclinic.org/diseases-conditions/bladder-cancer/diagnosis-treatment/drc-20356109#:~: text=Bladder%20cancer%20treatment%20may%20include, progression%20to%20a%20higher%20stage
10. NHS. (n.d.). Side effects of Chemotherapy. NHS choices. Retrieved August 16, 2022. From: URL: https: //www.nhs.uk/conditions/chemotherapy/side-effects
11. Rojano E., Seoane P., Ranea J. A.G., Perkins J. R. Regulatory variants: from detection to predicting impact. Brief Bioinform. Sep 27,– 20(5). 2019.– P. 1639–1654. Doi: 10.1093/bib/bby039. PMID: 29893792; PMCID: PMC6917219.
12. Shastry B. S. SNPs in disease gene mapping, medicinal drug development and evolution. J Hum Genet – 52, 2007.– P. 871–880. URL: https: //doi.org/10.1007/s10038–007–0200-z
13. Siteman Cancer Center. Effects of treatment for bladder cancer. (2021, July 23). Retrieved August 16, 2022. From: URL: https: //siteman.wustl.edu/treatment/cancer-types/bladder/effects-of-treatment
14. The American Cancer Society medical and editorial content team. (n.d.). What is bladder cancer? American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/about/what-is-bladder-cancer.html
15. The American Cancer Society medical and editorial content team. (2022, January 12). Key statistics for Bladder Cancer. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/about/key-statistics.html
16. The American Cancer Society medical and editorial content team. (n.d.). Bladder cancer staging: Bladder Cancer stages. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/detection-diagnosis-staging/staging.html
17. The American Cancer Society medical and editorial content team. (n.d.). Treatment of bladder cancer. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/treating/by-stage.html#:~: text=Chemotherapy%20followed%20by%20radical%20cystectomy, which%20may%20make%20surgery%20easier.

18. The American Cancer Society medical and editorial content team. (n.d.). Bladder cancer surgery. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/treating/surgery.html
19. The American Cancer Society medical and editorial content team. (n.d.). Intravesical therapy for Bladder Cancer. American Cancer Society. Retrieved August 16, 2022. From: URL: https: //www.cancer.org/cancer/bladder-cancer/treating/intravesical-therapy.html
20. Transurethral resection of the bladder: Before your surgery. MyHealth.Alberta.ca Government of Alberta Personal Health Portal. (n.d.). Retrieved August 16, 2022. From: URL: https: //myhealth.alberta.ca/Health/aftercareinformation/pages/conditions.aspx?hwid=ug6073#:~: text=Before%20Your%20Surgery-What%20is%20transurethral%20resection%20of%20the%20bladder%3F,)%20or%20malignant%20(cancer).
21. U. S. National Library of Medicine. (n.d.). What are single nucleotide polymorphisms (snps)?: Medlineplus Genetics. MedlinePlus. Retrieved August 16, 2022. From: URL: https: //medlineplus.gov/genetics/understanding/genomicresearch/snp
22. U. S. National Library of Medicine. (n.d.). CCL4 C–C motif chemokine ligand 4 [homo sapiens (human)]. National Center for Biotechnology Information. Retrieved August 16, 2022. From: URL: https: //www.ncbi.nlm.nih.gov/gene/6351