Over 60 Single Nucleotide Polymorphisms Across 8 Chromosomes are Shared by Seven Distinct Cancer Types

Bowen Jiang[1] and Roberto Aguilar[1]
[1]Department of Biology, Western Reserve Academy, Hudson, OH

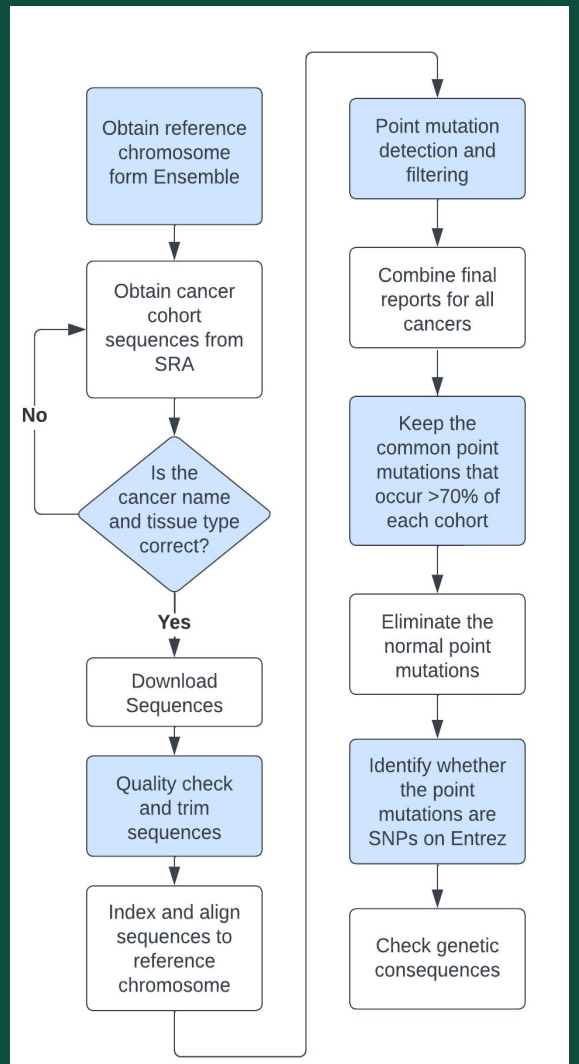
Abstract

Successful cancer treatment relies on understanding its etiology and the early detection of symptomatic patients. The identification of single nucleotide polymorphisms (SNPs) shared between distinct cancers can assist in furthering our understanding and may also serve as possible biomarkers. To investigate this objective, whole genome sequencing data were obtained from the Sequence Read Archive, and then analysis pipelines were constructed to map the sequences against different chromosomes followed by indexing and variant calling. To date, 8 chromosomes have been analyzed and over 60 SNPs found to be shared by 7 distinct cancers that range from breast to prostate cancer. Our analyses have identified cross-cancer SNPs that lay the foundation for future etiological studies and suggest that different cancers share common SNPs that may serve as suitable biomarkers.

Introduction

- About 2 million new cancer cases and 600 thousand cancer deaths are projected to occur in the United States for the year 2023[42]
- The possibility of successfully treating cancer is heavily dependent on its early diagnosis[3, 7, 12, 21, 30, 33, 38, 46]
- Identifying and analyzing the Single Nucleic Polymorphisms (SNPs) from the patient's DNA sequence can improve early diagnosis of cancers[26]
- Over 600 million SNPs have been identified in the human population around the world and they are commonly found in intergenic regions[43]
- Intergenic SNPs may be disease-associated and can serve as cancer biomarkers[10, 17, 25,]
- We formulated a computational workflow that employed the use of bioinformatics programs to analyze 9 different clinical cohorts that represent 7 distinct cancer types

Experimental Design



Cancer Cohort	Chromosome							
	1	2	3	4	5	6	10	17
Breast (S)	141	465	251*	239	375	133	93	271
Gastric (S)	1032	1645	1087	915	978	753	1485	1805
Lung (S)	1137	2404	1608	1538	1689	1332	2136	4457
Liver (S)	1993	3124	1652	2259	2242	1966	2175	2778
Ovarian (SC1)	506	854	616	507	559	439	668	567
Ovarian (SC2)	470	700	553	475	481	366	847	700*
Pancreatic (S)	281	431	305	295	327	233	416	337
Postate (S)	4519	7291	6005	4019	4921	2990	8979	15719
Postate (L)	1145	3097	1316	1483	1205	964	2099	3067
Shared PM	8	37	61	48	5	4	6	19
Shared SNPs	3	24	1*	22	3	2	1	7*

Resultant shared point mutations (PM) and SNPs within nine cancer cohorts across eight chromosomes. Bioinformatics tools were used to conduct a sequence reads analysis of 9 clinical cohorts representing 7 distinct cancers. Each row shows the number of PMs that occur in more than 70% of the analyzed individuals within each clinical cohort. Shared PM indicate the number of PMs that are shared between all cancer clinical cohorts. Shared SNPs were identified based on the shared PM after their elimination if found in the normal group (tissue sample origin: solid = **S**, single cell = **SC**, and liquid = **L**). * The corresponding group was left out of the data merging process.

Conclusion

We identified 63 mutual SNPs on eight distinct chromosomes that are present across nine clinical cohorts. Among the identified SNPs, there are 60 intergenic variants, 2 CTCF transcription factor binding site variants (functional epigenomic signature of cancer), and 1 variant on *IGK* (a cancer-related prognostic immunological biomarker). The total number of shared SNPs ranges from one on chromosome 3 and 10 to twenty-four on chromosome 2. These data support our hypothesis that there are SNPs that possess predictive values for different cancer types and may serve as dependable biological markers.

More Information