The Association Between Genes in Ethnic Cohorts and Acute Lymphoblastic Leukemia

Vivek Nalluri

Abstract

Acute lymphoblastic leukemia (ALL) is a blood cancer which commonly affects children and progresses rapidly. While comparative analyses have been performed between ALL survival rates of children of various races, there remains a lack of knowledge about patterns of commonly expressed genes within different ethnic cohorts who have ALL. Thus, it is critical to identify common genes that are expressed in various ethnic youth cohorts affected by ALL to understand how these genes may be targeted in future ALL therapies.

Introduction

Acute Lymphoblastic Leukemia (ALL), is a lymphoblastic cancer found in the blood and bone marrow. The disease progresses quickly if not treated, producing white blood cells called lymphoblasts, which can crowd out mature white blood cells such as B and T cells (1). Lack of B and T cells in patients with ALL will lead to a weakened immune system and a greater chance of immune attack from foreign antigens from bacteria or viruses. ALL is diagnosed most frequently in children under the age of 18, accounting for more than 30% of all childhood cancers (2). Although there are treatment options for ALL, including chemotherapy and various radiation or targeted therapies, ALL is the most common childhood malignancy with peak diagnosis age being between 2-9 years (3). Comparative studies have been performed on overall ALL survival rate differences between children of various races, with cohorts of Asian descent having the highest survival rate while cohorts of white, hispanic, and black descent have the lowest survival rates (4). Interestingly, childhood ALL was found to be 20% more common in Latino American populations compared to non-Latino white populations (5). What is not known about ALL and cohorts of various ethnicities are the patterns of common genes expressed within those different populations affected by ALL.

Research statement

In this review, common genes that are expressed in various ethnic youth cohorts affected by ALL and mechanism(s) of upregulation of said genes will be investigated.

A comparison of studies reporting ALL outcome in children based on race

Studies conducted by Bhatia et al. and Wiemels et al. have shown a significant fluctuation in the overall survival (OS) and event free survival (EFS) in children of varying race and ethnic groups with ALL. One study conducted by Bhatia et al. surveyed 6703 white, 1071 Hispanic, 506 black, 167 Asian, and 315 children of mixed or other racial origin, over 6.5 years. The study encompassed children placed in clinical trials, while adjusting for known risk factors including age at diagnosis, high initial leukocyte count, treatment era, etc.[4] OS rates at 5 years were about 90% for Asian, 84% for white, 79% for Hispanic, and 75% for black children. EFS rates at 5 years were 75% for Asian, 73% for white, 66% for Hispanic, and 62% for black children. The results showed worse outcomes for Black and Hispanic children compared to White and Asian cohorts. Another genome wide association study performed by Wiemels et al. among 3,623 Californian children with ALL found Latino children to be 20% more susceptible to ALL than non-Latino Whites (5). A potential indicator for Wiemels et al.'s results could be the

majority composition of Latinos within their study at 1,949 of the 3,263 study subjects. In this accord, the study by Bhatia et al. also displays a disproportionate number of study participants that belong to the White community (6,703) in comparison to the study size of the Asian population (167). These improper representations of ethnicities in both studies, can be a potential indicator for the results.

Genes associated with ALL in children

Thr241Met

Looking more specifically at genes reported to be associated with ALL in children, a study performed by Yan et al. published a meta-analysis in 2014 on the association between XRCC3 Thr241Met (rs861539) and leukemia risk with 2920 subjects (1070 cases and 1850 controls) (6). XRCC3, located at chromosome 14q32.3, is involved in DNA repair, and functions in the homologous recombination (HR) of DNA cross-links and double-strand breaks. There were five models assigned in analyses which looked at pooled odds ratios (ORs) to assess the strength of the association between XRCC3 Th r241Met (rs861539) polymorphism and leukemia risk. These models included allele contrast (T vs. C), homozygote (TT vs. CC), heterozygote (TC vs. CC), recessive (TT vs. TC/CC) and dominant (TT/TC vs. CC). A significant association was found within the OR of the Asian population for all 5 models. Table 1, below, highlights comparisons of the five models between populations based on OR, CI, and p-values. The OR for each of the 5 models is significantly higher for the Asian population in comparison to the Caucasian and mixed counterparts, suggesting that within this ethnic cohort, the rs861539 polymorphism may be more prevalent.

Table 1: Comparison of five models between ethnic populations based on OR, CI, and p-value

Models	Population	OR	95% CI	p -value
T vs. C	Asian	5.05	3.74 – 6.81	**0**
	Mixed	0.984	0.88 – 1.10	0.780
	Caucasian	1.17	0.91 – 1.52	0.227
TT vs. CC	Asian	26.29	8.61 – 80.25	**0**
	Mixed	0.99	0.74 – 1.33	0.944
	Caucasian	1.27	0.75 – 2.14	0.373
TC vs. CC	Asian	3.98	2.88 – 5.50	**0**
	Mixed	0.97	0.85 – 1.09	0.573
	Caucasian	1.13	0.88 – 1.45	0.336
TT vs. TC/CC	Asian	18.44	5.98 – 56.84	**0**
	Mixed	1.02	0.75 – 1.41	0.883
	Caucasian	1.24	0.89 – 1.75	0.207
TT/TC vs. CC	Asian	4.35	0.91 – 1.45	**0**
	Mixed	0.98	0.88 – 1.08	0.625
	Caucasian	1.37	0.98 – 1.94	0.258

Table 1: Comparison of five models (T vs. C, TT vs. CC, TC vs. CC, TT vs. TC/CC, and TT/TC vs. CC) between ethnic populations (Asian, mixed, and Caucasian), based on OR, CI, and p-value. Statistically significant p-values are bolded.

Another study published in 2021 by Xie et al. included 16 literature references furthering the study done by Yan et al. The results correlated leukemia risk and XRCC3 Tr241Met within the Caucasian as seen in table 2, below (7).

Table 2: Comparison of five models between ethnic populations based on OR, CI, and p-value

Models	Population	OR	95% CI	p -value
T vs. C	Caucasian	1.20	1.02–1.40	**0.026**
	Asian	1.25	0.62–2.50	0.530
	African	0.91	0.52–1.58	0.727
TT vs. CC	Caucasian	1.35	1.05–1.73	**0.018**
	Asian	2.05	0.74–2.66	0.169
	African	0.51	0.12–2.16	0.361
TC vs. CC	Caucasian	1.07	0.86–1.32	0.123
	Asian	1.01	0.54–1.76	0.643
	African	1.25	0.77–1.98	0.255
TT vs. TC/CC	Caucasian	1.31	1.04–1.64	**0.023**
	Asian	1.95	0.71–5.37	0.194
	African	0.45	0.12–1.69	0.239
TT/TC vs. CC	Caucasian	1.18	0.97–1.44	0.104
	Asian	1.19	0.69–2.05	0.537
	African	1.03	0.57–1.87	0.928

Table 2: Comparison of five models (T vs. C, TT vs. CC, TC vs. CC, TT vs. TC/CC, and TT/TC vs. CC) between ethnic populations (Caucasian, Asian, and African), based on OR, CI, and p-value. Statistically significant p-values are bolded.

The study by Yan et al. only reported a correlation between the Asian population and increased risk of childhood ALL, while the later study done by Xie et al. reported no correlation with the Asian population, but rather with the Caucasian population. Yan et al. used 7 studies while Xie et al reported the use of 16 studies. Furthermore, the study by Yan et al. included the African population and analyzed a larger number of datasets found in ALL literature.

ARID5B

ARID5B is a known transcriptional regulator in the differentiation of B-lymphocyte progenitor cells. Given this role and the significance of B cell dysregulation during leukemic development, a germline variation at the ARID5B locus could alter B-lymphocyte development and play a part in susceptibility to B-ALL. Within ARID5B, certain polymorphisms such as rs7089424 and rs10821936 may have a role in transcriptional regulation. The risk allele for rs10821936, in particular, is thought to eliminate the binding site of transcriptional factor NIT2, which could alter gene expression (8). To understand which cohorts are affected by this germline variation, a study done by Xu et al. tested whether ARID5B polymorphisms were related to treatment outcome in 1,605 children (9). Of the 49 ARID5B SNPs reported in the study, 8 which have already been associated with ALL susceptibility in the white and/or Hispanic populations have been also associated with relapse in ALL. The T allele at one of the identified SNPs (rs6479778) was more frequent in ALL cases than in controls, specifically in white and

Hispanic populations. Patients carrying the T allele were also at higher risk of ALL relapse and lower overall survival. Because the T allele at rs6479778 was more common in Hispanic people (29%) than in white people (14%), this polymorphism was stated to likely contribute to the racial disparity in both the incidence and treatment outcome of ALL. Furthermore, a study performed using 710 individuals from 323 Guatemalan and US Hispanic families by Archer et al. observed two inherited SNPs in ARID5B, rs10821936 and rs7089424, that were associated with childhood ALL. Notably, the magnitude of the effects of rs7089424 and rs10821936 in the study was over 20% higher than estimates calculated in previous genetic association studies among primarily European (non-Hispanic white) populations. This difference suggests that ARID5B variants appear to be more common among Hispanics. The risk allele frequencies (RAFs) observed for both rs10821936 and rs7089424 among both cases and parents in the study were higher than the risk allele frequencies reported among Hispanic and Native American reference populations. The study suggests that the increased frequency of the risk allele for both noted SNPs in Hispanics and Native Americans to be a factor for increased risk of childhood ALL for these populations (8). Finally, a study done by Xu et al. investigated the involvement of ARID5B along with polymorphisms in genes IKZF1, CEBPE, and PIP4K2A in ALL risk. The study concluded that children carrying six to eight copies of risk alleles at a ninefold (95% confidence interval = 6.9 to 11.8) had higher ALL risk relative to those carrying zero to one risk allele at these four single nucleotide polymorphisms (10). The studies show a correlation between ARID5B polymorphisms and increased risk of ALL in Hispanic populations. Hispanics also show an increased frequency of risk alleles. Overall, these studies highlight the increased susceptibility to ALL that Hispanic populations face and suggest the involvement of ARID5B.

FOXO3

A study done by Wang et al. in 2014 investigated the FOXO3 rs4946936 polymorphism for association with increased risk of childhood ALL in the Chinese population. FOXO3 is a member of FOXO3 family which binds to the Forkhead Response Element (FHRE, a core consensus DNA sequence in target genes) to control transcription (11). The study found the T allele to be more common in the group with ALL (30.1%) compared with the control group (25.3%). Furthermore, the combined CT/TT genotype was associated with an increased risk of childhood ALL (OR = 1.35, 95% CI = 1.12-1.63, P = 0.002). The study also found the expression levels of FOXO3a mRNA to be significantly higher in the patients carrying either the CT or TT genotype, as compared to patients carrying the CC genotype, suggesting that FOXO3 is correlated with an increased risk of childhood ALL (12). A study by Yang et al. in 2019 investigated an Asian population and identified 4 FOXO3 SNPs (rs17069665 A>G, rs4945816 T>C, rs4946936 C>T, and rs9400241 A>C) that overlapped with transcription factor binding sites (TFBS) or microRNA (miRNA)-binding sites. After adjusting for age and gender, rs17069665 G allele (GG vs AA/AG: adjusted OR = 1.76; 95% CI = 1.02-3.04, P = 0.043) was significantly related to the increased ALL risk in a recessive model, while rs9400241 C allele (CC/AC vs AA: adjusted OR = 0.80; 95% CI = 0.64-0.99, P = .046) significantly related to decreased ALL risk in a dominant model (11). The study by Yang et al. did not find the FOXO3 rs4946936 polymorphism discussed in the study by Wang et al. to be correlated with increased risk of childhood ALL, but they did identify a correlation with the rs17069665 SNP. Both studies investigated the Chinese population and contained a comparable number of patients in the study; Wang et al. reported 753 ALL patients and 1,088 controls and Yang et al. reported 425 ALL patients and 1339 controls.

Discussion

The following studies were performed to assess how common certain polymorphisms are in ALL cases of specific ethnic populations. The association between XRCC3 Thr241Met (rs861539) and leukemia risk reported by Yan et al. showed that the rs861539 polymorphism may be more prevalent in the Asian cohort compared to the Caucasian and mixed cohort. A later study done by Xie et al. reported no correlation between the rs861539 polymorphism and the Asian population, but rather with the Caucasian population. Other studies reported looking into the ARID5B locus. One study by Xu et al. identified the T-allele of the rs6479778 ARID5B polymorphism to be found more frequently in ALL cases compared to the control group and noted to be more frequent in Hispanics, potentially explaining the highest susceptibility of Latinos to ALL. A further study done by Archer et al. suggests that the increased frequency of the risk allele for both noted SNPs (rs10821936 and rs7089424) in Hispanics and Native Americans to be a factor for increased risk of childhood ALL for these populations. Another locus associated with increased ALL risk, FOXO3, was investigated by two studies (Yang et al. and Wang et al.) targeting its effects within the Asian population. The study by Yang et al. did not find the FOXO3 rs4946936 polymorphism discussed in the study by Wang et al. to be correlated with increased risk of childhood ALL, but they did identify a correlation with the rs17069665 SNP.

It is also notable that XRCC3 Thr241Met, ARID5B, and FOXO3 have also been studied for association with other types of cancers. XRCC3 could potentially be associated with an elevated risk for colorectal cancer, especially with the Asian population (13). The ARID5B locus has also been associated with impacting outcomes for breast cancer (14) and in the development of gastric cancer (15). FOXO3 as a tumor suppressor has also been associated with gastric cancer along with prostate cancer (16).

Apart from ethnicity-related factors reported in these studies, non-ethnicity factors were considered only in the study performed by Wang et al. It is known in literature that other environmental factors such as exposure to radiation, benzene, mercury, and long-term use of antibiotics may correlate with ALL risk. For instance, a hospital-based case-control study by Xie et al. containing 753 children with ALL and 1,088 cancer free controls acknowledged the impact on non-ethnic factors in ALL susceptibility (7). In the children with ALL, 34.5% of their parents identified positive for alcohol consumption, compared to the 21.0% within the control. The study also reported 35.9% of the case patients as having had the house painted during the mother’s pregnancy or after the child's birth, while the control only reported 26.2%. Interestingly, a common substance found in household paints was benzene, which has been reported in a study by Hiraku et al. to cause DNA damage through the formation of hydrogen peroxide (H2O2) before causing internucleosomal DNA fragmentation leading to apoptosis (17). Thus, it would be beneficial in future studies to include environmental factors when reporting ALL risk within ethnic populations based on OR, CI, or p-value to provide a more multifaceted perspective on the various causes of ALL progression.

Acknowledgements

I would like to acknowledge Polygence for providing me with a streamlined platform to perform research on. Along with Polygence, my mentor Christina Mark has been exceptionally supportive throughout the whole process, allowing me the independence to explore the scientific

world, while also guiding me to produce a presentable final product. Her knowledge and passion have been inspirational and crucial to the final outcome.

References

1. Acute Lymphocytic Leukemia. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/acute-lymphocytic-leukemia/symptoms-causes/syc-20369077.
2. Puckett Y, Chan O. Acute Lymphocytic Leukemia. [Updated 2022 Jun 27]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459149/
3. Acute Lymphoblastic Leukemia (ALL). St. Jude Children's Research Hospital. https://www.stjude.org/disease/acute-lymphoblastic-leukemia-all.html
4. Bhatia, et al. Racial and ethnic differences in survival of children with acute lymphoblastic leukemia. The American Society of Hematology. 2002;100:6
5. Wiemels, et al. GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosomes 17q12 and 8q24.21. Nat Commun. 2018 Jan 18;9(1):286.
6. Yan, et al. Association of XRCC3 Thr241Met polymorphism and leukemia risk: evidence from a meta-analysis. 2014; 55(9): 2130–2134
7. Xie, et al. Ethnicity-stratifed analysis of the association between XRCC3 Thr241Met polymorphism and leukemia: an updated meta-analysis. 2021; 14:229
8. Archer NP, Perez-Andreu V, Stoltze U, Scheurer ME, Wilkinson AV, Lin T-N, et al. (2017) Family-based exome-wide association study of childhood acute lymphoblastic leukemia among Hispanics confirms role of ARID5B in susceptibility. PLoS ONE 12(8): e0180488. https:// doi.org/10.1371/journal.pone.0180488
9. Xu, et al. ARID5B Genetic Polymorphisms Contribute to Racial Disparities in the Incidence and Treatment Outcome of Childhood Acute Lymphoblastic Leukemia. J Clin Oncol. 2012;30:751-757.
10. Xu, et al. Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations. J Natl Cancer Inst. 2013;105:733–742.
11. Yang, et al. FOXO3 gene polymorphisms influence the risk of acute lymphoblastic leukemia in Chinese children. J Cell Biochem. 2020;121:2019–2026.
12. Wang, et al. Association of the 3'UTR FOXO3a Polymorphism rs4946936 with an increased risk of childhood acute lymphoblastic leukemia in a Chinese population. Cell Physio Biochem. 2014;34:325-332.
13. Wang Z, Zhang W. Association between XRCC3 Thr241Met polymorphism and colorectal cancer risk. Tumour Biol. 2013 Jun;34(3):1421-9.
14. Zhang J, Hou S, You Z, Li G, Xu S, Li X, Zhang X, Lei B, Pang D. Expression and prognostic values of ARID family members in breast cancer. Aging (Albany NY). 2021 Feb 11;13(4):5621-5637.
15. Lim B, Park JL, Kim HJ, Park YK, Kim JH, Sohn HA, Noh SM, Song KS, Kim WH, Kim YS, Kim SY. Integrative genomics analysis reveals the multilevel dysregulation and oncogenic characteristics of TEAD4 in gastric cancer. Carcinogenesis. 2014 May;35(5):1020-7.
16. Habrowska-Górczyńska DE, Kozieł MJ, Kowalska K, Piastowska-Ciesielska AW. FOXO3a and Its Regulators in Prostate Cancer. Int J Mol Sci. 2021 Nov 20;22(22):12530.

17. Hiraku Y, Kawanishi S. Oxidative DNA damage and apoptosis induced by benzene metabolites. Cancer Res. 1996;56(22):5172-8.