European Journal of Biomedical and Life Sciences

2024, No 1–2

PREMIER Publishing

Vienna
2024

European Journal of Biomedical and Life Sciences

Scientific journal

№ No 1–2

ISSN 2310-5674

Proofreading Kristin Theissen

Cover design Andreas Vogel

Additional design Stephan Friedman

Editorial office Premier Publishing s.r.o.
Praha 8 – Karlín, Lyčkovo nám. 508/7, PSČ 18600

E-mail: pub@ppublishing.org

Homepage: ppublishing.org

European Journal of Biomedical and Life Sciences is an international, English language, peer-reviewed journal. The journal is published in electronic form.

The decisive criterion for accepting a manuscript for publication is scientific quality. All research articles published in this journal have undergone a rigorous peer review. Based on initial screening by the editors, each paper is anonymized and reviewed by at least two anonymous referees. Recommending the articles for publishing, the reviewers confirm that in their opinion the submitted article contains important or new scientific results.

Premier Publishing s.r.o. is not responsible for the stylistic content of the article. The responsibility for the stylistic content lies on an author of an article.

Instructions for authors

Full instructions for manuscript preparation and submission can be found through the Premier Publishing s.r.o. home page at: http://ppublishing.org.

Material disclaimer

The opinions expressed in the conference proceedings do not necessarily reflect those of the Premier Publishing s.r.o., the editor, the editorial board, or the organization to which the authors are affiliated.

Premier Publishing s.r.o. is not responsible for the stylistic content of the article. The responsibility for the stylistic content lies on an author of an article.

Included to the open access repositories:

НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА eLIBRARY.RU TOGETHER WE REACH THE GOAL

Scientific Journal Impact Factor

Scientific Journal Impact Factor Value for 2024 – 6.671

Typeset in Berling by Ziegler Buchdruckerei, Linz, Austria.
Printed by Premier Publishing s.r.o., Vienna, Austria on acid-free paper.

The European Journal of Biomedical and Life Sciences 2024, No 1–2
ISSN 2310-5674

PREMIER Publishing
ppublishing.org

Section 1. Clinical medicine

DOI: 10.29013/EJBLS-24-1.2-3-12

ASSOCIATION OF PHYSICAL ACTIVITY WITH DEPRESSION ACROSS DIFFERENT RACES IN ADULTS: INSIGHTS FROM NHANES2015–2018 DATA

Yixuan Liu [1]

[1] The Masters School

Cite: *Yixuan Liu. (2024). Association of Physical Activity with Depression Across Different Races in Adults: Insights from NHANES 2015-2018 Data. The European Journal of Biomedical and Life Sciences 2024, No 1–2* https://doi.org/10.29013/EJBLS-24-1.2-3-12

Abstract

Depression is an affective disorder that brings an economic and emotional burden to individuals and society. This study explored the association between the types of physical activity and the risk of depression, evaluating racial disparities in physical activity levels among adults. Data from the 2015–2018 National Health and Nutrition Examination Survey (NHANES) were utilized, encompassing 2,111 adults. Key variables considered were demographics, physical activity, and depression, as measured by the Physical Activity Questionnaire and the PHQ-9 depression scoring system. Out of the five physical activity categories, vigorous work, moderate work, and walking and cycling activities showed no significant correlation with depression risk. In contrast, moderate recreational and vigorous recreational activities appeared to offer protective effects against depression. Non-Hispanic black individuals notably benefited from moderate recreational activities ($p = 0.021$) in lowering their depression risk. For Non-Hispanic whites, engaging in vigorous recreational ($p = 0.004$) and moderate recreational ($p = 0.002$) activities were advantageous in depression prevention. Among Mexican Americans and "other racial categories", the five types of physical activities did not significantly reduce depression risk. The results suggest that engaging in physical activity can reduce the risk of depression. However, the benefits vary among racial groups based on the intensity and type of physical activity. These results can help inform mental health professionals when making recommendations to patients from different racial groups. Future research should look at the reasons behind the differences among the groups.

Keywords: *physical activity, depression, race, NHANES, PHQ-9*

Introduction

Depression, a widespread mental disorder, is characterized by persistent sadness and a loss of interest in activities previously found enjoyable. This condition not only affects individuals emotionally but also places significant

strain on families and society. The economic implications of depression are profound, including direct medical costs, losses from workplace disruptions like absenteeism, and indirect costs from suicides. In 2018, the economic burden of major depressive disorder in the U.S. reached an estimated $236 billion, a rise of over 35% since 2010 (American Psychiatric Association, 2021). While people suffer from depression, many are not given proper treatment. Data from the National Health and Nutrition Examination Survey, a survey research program that aims to assess the health and nutritional status of adults and children in the United States, provides the opportunity to delve deeper into the topic of the relationship between physical activity and depression, and influence that different racial groups bring to the study result.

Depression coexists with various health issues, including tuberculosis, cardiovascular diseases, and sleep disturbances. Moreover, those with depression often contend with chronic conditions such as arthritis, heart disease, diabetes, stroke, and cancer. While depression can cause physical symptoms and increase the risk for physical illnesses or conditions, illnesses can in turn trigger depression and make the situation worse (Stuart, 2022). Data from the National Library of Medicine indicate that the prevalence of depression increased from 7.3% in 2015 to 8.6% in 2019, suggesting that almost one in 12 U.S. residents experience this disorder (Goodwin et al., 2022). Research has consistently shown a strong association between depression, increased suicide risk, and reduced quality of life (Sakashita & Oyama 2019; Dhar & Barton 2016).

When people exercise, hormones called endorphins are released and reduce the perception of pain by interacting with neurons that respond to pain. Moreover, endorphins act as sedatives, which help reduce stress and anxiety and improve sleep quality (Bruce, 2022). Doing physical activities has been shown to be helpful in reducing depression and making people feel better. The biological benefits of physical activity make it a good option for people to select when coping with their mental issues.

Hallgren et al. carried out a study involving 24,060 participants that demonstrated that replacing sedentary behavior with light to moderate-to-vigorous work activity significantly reduced the risk of depression (Hallgren *et al.*, 2020). An examination of the bidirectional relationship between physical activity and depression in 611,583 adults endorsed the potential of physical activity as a preventive measure against depression (Choi et al., 2019). A meta-analysis emphasized the mental health benefits of physical activity for adults, even when undertaken below the recommended levels for public health (Pearce et al., 2022). Analysis of NHANES data from 2011–2014 indicated that leisure-time physical activity was particularly effective in alleviating depression among US adults (Rutherford et al., 2022). Furthermore, researcher Dina found that the benefits of exercise and physical activity on depressive symptoms were on par with antidepressant treatments (Dinas et al., 2011).

Investigations into the effects of physical activity on depression in both adults and adolescents concluded that regular activity considerably reduces depression risk (Bailey et al., 2018). Dr. Aan Het Rot confirmed the positive correlation between physical activity and enhanced mental well-being in adults (Aan Het Rot et al., 2009). Emerging evidence suggests that even minimal amounts of physical activity can act as a safeguard against depression (Teychenne et al., 2008). Research focusing on older adults highlighted that various forms of exercise, including aerobic, resistance, and mind-body exercises, were linked to fewer depressive symptoms (Zhang et al., 2021). Lastly, Mumba et al. found that both vigorous and moderate recreational physical activities had notable effects on depression scores among older adults (aged 50 and above) (Mumba et al., 2021). It further mentioned the importance of considering the potential differences between various racial groups, which aligns with one of the focuses of our study.

While numerous studies have explored the impact of physical activity on depression, many focus on the effects of physical activity on depression, together with other mental disorders and diseases. A few researchers take the five physical activity levels as a standard for research and directly study the relationships between the time of doing physical activities and the chance of suffering from depressive symptoms. Moreover, there remains a limited body

of research examining the differential effects across different racial groups in the US. Given the substantial socio-economic and health disparities among these groups, the influence of physical activity on depression in the general population may not directly apply to specific racial demographics. Hence, it becomes imperative to evaluate how different types of physical activity affect depression within each racial group.

Methods

Data Collection & Analysis

This research utilized data sourced from the National Health and Nutrition Examination Survey (NHANES), a nationally representative cross-sectional survey orchestrated by the Centers for Disease Control and Prevention (CDC). NHANES employs a stratified multistage design rooted in random sampling techniques. Trained interviewers gathered questionnaire data in participants' homes. Meanwhile, the mobile examination center (MEC) phlebotomists collected blood samples. These samples were subsequently refrigerated or frozen and dispatched to laboratories for analysis. Additionally, the MEC managed physical examinations. The current analysis uses data from the NHANES cycles of 2015–2016 and 2017–2018.

Out of the 11,268 participants in the NHANES2015–2018 cohort, exclusions were made for those under 20 years old (529 participants) and those with indeterminable depression statuses (1,066 participants). The NHANES review process further omitted entries with missing, rejected, or ambiguous data on physical activity and sedentary time, as well as those reporting over 24 hours, encompassing total physical activity time, sedentary periods, and sleep duration. Participants with absent or rejected covariates were also excluded (7,562 in total). Ultimately, a refined cohort of 2,111 participants, with comprehensive data on primary outcomes, exposure, and relevant variables, was retained for analysis.

Definition of Major Depression and Physical Activities

For the purposes of this research, participants were classified as having depression based on their responses to the PHQ-9 questionnaire (Lagerros et al., 2006); specifically, if they completed all items and achieved a threshold score.

Physical activity patterns were ascertained using the Global Physical Activity Questionnaire (Lagerros et al., 2009). Physical activities were delineated into five categories, determined by participants' responses:

Vigorous Work Activity: Defined as tasks that cause a pronounced increase in breathing or heart rate for an extended period. Vigorous Recreational Activities: including leisure, fitness, and high-intensity sports that result in a significant escalation in respiration or heart rate. Moderate Work Activity: Activities characterized by a modest elevation in breathing or heart rate. Moderate Recreational Activities: Activities inducing a light rise in respiratory or heart rate for a set duration. Walking or Bicycling: Engaging in either of these for transportation purposes, such as commuting to work, school, or shopping. The reported activity durations were aggregated to calculate the cumulative physical activity duration. This was derived from the multiplication of the frequency and duration of each physical activity type reported by participants.

Covariates

Data was collected using a structured household questionnaire, capturing gender (male, female), age, education level (Less than high school, HS grad/GED, Some college/AA, College/above), marital status (categorized as married or unmarried, with the latter encompassing widowed, divorced, separated, never married, and cohabiting), race (categories included Mexican American, Non-Hispanic Black, Non-Hispanic White, and other races), and employment status from the previous week. Smoking history was ascertained with the question, "Have you smoked at least 100 cigarettes in your entire life?" (yes/no). Body Mass Index (BMI) was calculated by dividing weight (kg) by the square of height (m^2). Hypertension status was determined based on self-reported diagnosis, antihypertensive medication use, or if measured blood pressure exceeded specified thresholds (systolic ≥ 140 mm Hg and/or diastolic ≥ 90 mm Hg).

Statistical Analysis

The baseline characteristics of participants were categorized by depression status.

Descriptive statistical analyses were applied to demographic data (such as gender, age, education level, race, marital status, and employment), behavioral metrics (like total physical activity duration), body measurements (specifically, BMI), and medical conditions (such as hypertension and obesity). Categorical variables were expressed as frequency (%), and chi-square tests were used for non-ordinal categorical variables. Continuous variables were presented as mean ± standard error, with the Student's t-test determining statistical relevance. Considering the multistage sampling design and oversampling, we employed MEC weights in our models, supplemented with variables for primary sampling units (PSUs) and the Masked Variance Unit Pseudo-stratum. These weight variables enable generalization to the entire civilian noninstitutionalized US population. The association between depression risk and physical activity was examined through weighted logistic regression models. Odds ratios (ORs) with 95% confidence intervals (CIs) were provided, along with p-values to assess the association's magnitude and significance. Multivariable logistic regression models, inclusive of demographics, behaviors, and medical conditions as covariates, were devised to adjust for potential confounders. A p-value below .05 was deemed statistically significant. All computations were executed in the R software.

Results

Table 1 shows the sample characteristics of a total of 2111 respondents between the years 2015–2018. Overall, depression is more common in females (12.9%) than males (7.4%, p-value = 0.0097). Participants who answered "Yes" to smoking were more likely to have major depression than those who did not smoke (15.6% vs 4.7%, p-value, 0.0001). Major depression was more likely to occur among participants with normal weight (12.7%), without a job or business (13.3%), and lowest poverty income ratio (16.3% for PIR < 1.3). There were no significant differences between participants with or without depression for hypertension and total physical activity time.

Table 1. *Sample characteristics of demographics, medical conditions, and physical activities by depression status*

		Depression status				
		No		Yes		p value
		n(%)	SE	n(%)	SE	
Hypertension	No	1223(91.2)	1.1	124(8.8)	1.1	0.1727
	Yes	682(88.4)	1.9	82(11.6)	1.9	
Gender	male	1268(92.6)	1.8	104(7.4)	1.8	0.0097
	Female	637(87.1)	0.9	102(12.9)	0.9	
Smoking status	No	1044(95.3)	0.8	65(4.7)	0.8	<0.0001
	Yes	861(84.3)	1.7	141(15.6)	1.7	
Obesity status	Normal/underweight	818(87.3)	1.6	108(12.7)	1.6	0.0206
	Obesity	22(97.1)	3.2	1(2.9)	3.2	
	Overweight	1065(92.8)	1.1	97(7.2)	1.1	
employment	Looking for work/not working	847(86.7)	1.5	131(13.3)	1.5	<0.001
	With a job or business	1058(92.7)	1.1	75(7.3)	1.1	
Poverty level	<1.3	623(83.7)		106(16.3)		<0.001
	1.3~3.5	799(89.7)		78(10.3)		
	>3.5	483(95.4)		22(4.6)		
		mean	SE	mean	SE	
Physical activity time (minutes)		1007.2	39.8	890.9	116	0.358

Note: SE = standard error, poverty level <1.3 = poverty, 13.~3.5 = normal, >3.5 = rich

Table 2 shows the crude (univariable logistic regression) and adjusted (multiple logistic regression) models of various physical activity types as predictors and depression as the outcome. In the crude univariable logistic regression model, physical activities such as vigorous and moderate work activities were not significantly associated with depression (OR and 95% CI, 0.95 [0.55,1.58], p-value = 0.8104 for vigorous work activities; OR and 95% CI, 1.09[0.76,1.55], p-value = 0.6561 for vigorous work activities). Similarly, in multiple logistic regression models adjusted for other covariates, the odds ratios (OR) for vigorous (OR = 0.74, 95% CI = [0.42, 1.32], p-value = 0.3276) and moderate work activities (OR = 0.98, 95% CI = [0.66, 1.45], p-value = 0.9189) were not significantly associated with depression. Walk or bicycle activities were significantly associated with depression (p-value = 0.0066), with an OR of 1.64 (95% CI = [1.18, 2.30]), indicating having walk or bicycle activities was associated with depression. However, after adjusting for other covariates, the association between walking or bicycle activities and depression was attenuated (OR = 1.37, 95% CI = [0.95, 1.96], p =0.1072). Vigorous recreational activities were significantly associated with depression (OR = 0.46 95% CI = [0.28, 0.75], p-value = 0.0041) in the crude model. As in the multiple logistic regression models, the relationship between depression and vigorous recreational activities is still significant, with odds ratios of 0.48 (95% CI [0.31, 0.73], and p-value = 0.0037).

Table 2. *Associations Between Individual Types of Physical Activities and Depression: Crude and Adjusted Logistic Regression Models*

	Crude model		Adjusted model	
Type of PA	**OR[LCI, UCI]**	**p value**	**OR[LCI, UCI]**	**p value**
Vigorous work	0.95 [0.55,1.58]	0.8104	0.74 [0.42, 1.32]	0.3276
Moderate work	1.09[0.76,1.55]	0.6561	0.98[0.66, 1.45]	0.9189
Walk or Bicycle	1.64[1.18, 2.30]	0.0066	1.37[0.95,1.96]	0.1072
Vigorous Recreational	0.46[0.28,0.75]	0.0041	0.48[0.31,0.73]	0.0037
Moderate Recreational	0.39[0.27, 0.58]	<.001	0.47[0.32,0.71]	0.0024

Note: OR = odds ratio, LCI = lower confidence intervals, UCI = upper confidence intervals

The subgroup analysis of the association between various physical activity types and depression by race groups in multiple logistic regression models is shown in Table 3. The vigorous work activities do not have a significant protective effect on depression among Mexican-American groups (OR=0.44, 95% = = [0.44,1.25], p-value = 0.122). Similarly, vigorous work activities influence Non-Hispanic black (OR = 0.64, 95% = [0.64, 1.85], p-value = 0.409), white (OR = 0.75, 95% = = [0.75, 1.48], p-value = 0.413) and other races (OR = 0.48, 95% = [0.84, 1.79], p-value = = 0.643) group respectively by decreasing the possibilities of getting depression. However, none of the associations between vigorous work activities and depression were statistically significant.

Table 3. *Associations Between Individual Types of Physical Activities and Depression: subgroup analysis by racial groups*

	Mexican American		Black	
Type of PA	**OR[LCI, UCI]**	**p value**	**OR[LCI, UCI]**	**p value**
Vigorous Work	0.44[0.44, 1.25]	0.122	0.64[0.64, 1.85]	0.409
Moderate Work	1.46[1.46, 5.24]	0.561	1.21[1.21, 2.47]	0.592
Walk or Bicycle	0.27[0.27, 2.26]	0.225	1.32[1.32, 2.46]	0.384
Vigorous Recreational	1.43[1.43, 5.30]	0.594	0.45[0.45, 1.14]	0.093
Moderate Recreational	1.36[1.36, 4.74]	0.631	0.42[0.42, 0.88]	0.021

Type of PA	White OR[LCI, UCI]	p value	Other OR[LCI, UCI]	p value
Vigorous Work	0.75[0.75, 1.48]	0.413	0.84[0.84,1.79]	0.643
Moderate Work	0.88[0.88, 1.45]	0.629	1.24[1.24, 2.87]	0.607
Walk or Bicycle	1.65[1.65, 2.77]	0.58	0.76[0.76, 1.97]	0.572
Vigorous Recreational	0.48[0.48, 0.79]	0.004	0.35[0.35, 1.04]	0.059
Moderate Recreational	0.37[0.37, 0.68]	0.002	0.99[0.99, 2.10]	0.975

Note: PA = physical activity, OR = odds ratio, LCI = lower confidence intervals, UCI = upper confidence intervals

For moderate work activities, Non-hispanic white is the only group that shows protective effect of moderate work activities(OR = 0.88, 95% = [0.88, 1.45], p value = 0.629). Whereas among Mexican Americans (OR = 1.46, 95% = [1.46, 5.24], p-value = 0.561), Non-Hispanic blacks (OR = 1.21, 95% = [1.21, 2.47], p-value = 0.592) and other races(OR = 1.24, 95% = [1.24, 2.87], p-value = 0.607), moderate work activities increased the likelihood of getting depression. However, moderate work activities were not a significant factor for depression.

Walking and cycling activities are not significantly associated with depression in Mexican American(OR = 0.27, 95% = [0.27, 2.26], p-value = 0.225) and Non-hispanic white(OR = 1.65, 95% = [1.65, 2.77], p-value = 0.58). Meanwhile, non-hispanic black(OR = 1.32, 95% = [1.32, 2.46], p-value = 0.384) and other races(OR = 0.76, 95% = [0.76, 1.97], p-value = 0.607) are relatively less associated by walking and cycling activities in depression.

The vigorous recreational activities are not significantly associated with depression in Mexican Americans (OR = 1.43, 95% = [1.43, 5.30], p-value = 0.594), Non-Hipanic Black (OR = 0.45, 95% = [0.45, 1.14], p-value = 0.093), and Other racial groups(OR = 0.35, 95% = [0.35, 1.04], p-value = 0.059). However, in the Non-Hispanic White group, vigorous recreational activities are significantly associated with a lower likelihood of having depression (OR = 0.48, 95% = [0.48, 0.79], p-value = 0.004).

Moderate recreational activities are significantly reducing the likelihood of having depression among Non-Hispanic White (OR = 0.37, 95% = [0.37, 0.68], p-value = 0.002) and Non-Hipanic Black (OR = 0.42, 95% = [0.42, 0.88], p-value = 0.021). While in Mexican American and other racial groups, the associations between moderate recreational activities and depression were not statistically significant (OR = 1.36, 95% = [1.36, 4.74], p-value = 0.631 and OR = 0.99, 95% = [0.99, 2.10], p-value = 0.975 for Mexican American and Other races respectively).

Discussion

The primary objective of this research was to examine the association between various forms of physical activity (including vigorous work-related activity, vigorous recreational activities, moderate work activity, moderate recreational activities, and walking/bicycling) and the prevalence of depression. In addition, considering potential racial disparities in the susceptibility to depression, this study delved deeper into the moderating effect of race on the relationship between the mentioned physical activities and depression.

To the best of my knowledge, this is the first study that studies the variable effects of physical activities on different racial groups. The findings indicate that specific physical activities can either reduce or amplify the risk of depression among the four racial groups analyzed. For instance, in Mexican Americans, even though the results weren't statistically significant, engaging in vigorous work activities, as well as walking and bicycling, appears to lessen the risk of depression. Conversely, moderate, vigorous recreational, and moderate recreational activities may increase their predisposition to depression. In contrast, the non-Hispanic Black population benefits from moderate recreational activities but seems to have an increased risk of depression when involved in moderate work activities and walking or cycling. Non-Hispanic Whites exhibit a markedly reduced risk of depression with moderate recreational activities, which

appear to act as protective factors. Other racial groups did not demonstrate a significant correlation between any form of physical activity and depression. Nonetheless, vigorous recreational activities could potentially offer protective benefits, although the statistical significance is marginal.

My study determined that, among all participants, those without depression engaged in physical activity for an average of 116.3 minutes more per week compared to their counterparts with depression. This observation aligns with findings by Pearce et al., suggesting that adults who meet the recommended physical activity duration are less prone to depression (Pearce et al., 2022). Choi et al. also explored the bidirectional relationships between physical activity and depression in adults, reinforcing our conclusion about the protective nature of physical activity against depression (Choi et al., 2019). Similarly, research by Emily R. Rutherford et al. established that individuals engaging in over 150 minutes of moderate to vigorous work activity per week are less likely to experience depression (Rutherford et al., 2022). Our findings also receive backing from a study by Mats Hallgren et al. on the relationships between various sedentary behaviors, physical activity, and depression. This research recommended substituting passive sedentary actions with mentally active sedentary behaviors or moderate-to-vigorous work activity to diminish the risk of depression in adults (Hallgren *et al.*, 2020). Concurrently, Marije Aan Het Rot et al. emphasized that adhering to the CDC's recommendation of at least 30 minutes of moderate-intensity physical activity on most days significantly alleviates depressive symptoms (Aan Het Rot et al., 2009).

Additionally, a study conducted by Su Zhang et al. delved into the connection between physical activity and depression in older adults. They highlighted an inverse relationship between engagement in physical activities and the onset of depressive symptoms, underscoring the benefits of exercise in reducing depression risks for this demographic (Zhang et al., 2021). Mercy Ngosa Mumba et al. undertook research on the same demographic and not only corroborated our findings but also emphasized that, for many older adults, physical activity is a more cost-effective approach to mitigating depression risks than medical treatments. This study also suggested potential variations in depression rates across different racial groups, a key focus of our research (Mumba et al., 2021). The therapeutic potential of physical activity in addressing depression is also grounded in science. As Aaron Kandola et al. determined, physical activity exerts its antidepressant effects through various biological and psychosocial mechanisms within the brain, forging a protective shield against depression (Kandola et al., 2019). A similar conclusion was drawn in research by Megan Teychenne et al., which underscored an inverse correlation between physical activity and depression (Teychenne et al., 2008). Andreas Ströhle's study resonated with our findings, suggesting that physical activities bolster resilience against depression in adults (Ströhle, 2009). Intriguingly, research by A. P. Bailey et al. indicated that young adults and adolescents with depression could benefit significantly from physical activity as a primary intervention, suggesting its therapeutic potential in alleviating their symptoms (Bailey et al., 2018).

While the aforementioned studies corroborate our findings, emphasizing the benefits of physical activity in reducing depression risks for adults, there remains a noticeable gap in research addressing the potential disparities across different racial groups. This avenue presents an opportunity for researchers to delve deeper, seeking to understand the nuances of how and why individuals from various racial backgrounds might exhibit differing susceptibilities to depression.

This study boasts several significant strengths. Primarily, the analysis is underpinned by a robust participant sample. The model adjustments ensure the comprehensive inclusion of pertinent covariates, aided by an exhaustive survey. The physical activity questionnaire, validated in prior studies(Lagerros et al., 2006), astutely differentiates between various types of physical activities (Lagerros et al., 2006). Furthermore, our subgroup analysis segmented by race/ethnicity offers granular insights into how diverse physical activity types might uniquely affect different racial groups.

However, the research paper has limitations. Firstly, participants' self-reporting

of their physical activities could introduce recall errors during questionnaire completion. Secondly, our depression outcomes, defined by these self-reported questionnaires, might be prone to underestimation. This is due to potential disparities in participants' interpretation of the questions, which could result in imprecise responses and, consequently, skewed final scores. Lastly, inherent limitations are tied to the NHANES database, which is based on a cross-sectional population survey. By design, NHANES can only determine associations between factors, not the causative mechanisms underlying them. As a result, this study cannot conclusively establish causality.

Conclusion

Our study investigates the protective role of physical activities in mitigating the risk of depression among adults. Distinctively, the differential impacts of various physical activity types on depression risk were evident across the four racial groups studied. While these insights are promising, the results should be approached with prudence due to the inherent limitations highlighted. Future research, ideally with a more diverse racial categorization and a more expansive sample, is imperative to comprehensively understand the nuanced interplay between race, physical activity, and depression risk. Such endeavors could inform targeted interventions for specific racial groups, optimizing both treatment and prevention strategies.

Acknowledgments

I am deeply grateful to Dr. Xin Yang for his invaluable guidance, expertise, and unwavering support throughout this research. I also want to express my appreciation for the effective and helpful support of Mington workers. Their mentorship significantly influenced the direction of this study.

References

Marije aan het Rot, M., Collins, K. A., & Fitterling, H. L. (2009). Physical exercise and depression. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine: A Journal of Translational and Personalized Medicine*,– 76(2).–P. 204–214.

Bailey, A. P., Hetrick, S. E., Rosenbaum, S., Purcell, R., & Parker, A. G. (2018). Treating depression with physical activity in adolescents and young adults: a systematic review and meta-analysis of randomised controlled trials. *Psychological medicine*,– 48(7).– P. 1068–1083.

Bruce, D. F. (2022, April 1). *Exercise and depression: Endorphins, reducing stress, and more*. WebMD. URL: https://www.webmd.com/depression/exercise-depression#1–1

Centers for Disease Control and Prevention. (2017, December). *Nhanes 2015–2016: Mental health – depression screener data documentation, codebook, and frequencies*. Centers for Disease Control and Prevention. URL: https://wwwn.cdc.gov/Nchs/Nhanes/2015–2016/DPQ_I.htm

Centers for Disease Control and Prevention. (2020, February). *Nhanes 2017–2018: Mental health – depression screener data documentation, codebook, and frequencies*. Centers for Disease Control and Prevention. URL: https://wwwn.cdc.gov/Nchs/Nhanes/2017–2018/DPQ_J.htm

Chen, T. C., Parker, J. D., Clark, J., Shin, H. C., Rammon, J. R., & Burt, V. L. (2018). National health and nutrition examination survey: estimation procedures, 2011–2014.

Chen, Te-Ching et al. (2020). National Health and Nutrition Examination Survey, 2015–2018: sample design and estimation procedures.– 184 p.

Choi, K. W., Chen, C. Y., Stein, M. B., Klimentidis, Y. C., Wang, M. J., Koenen, K. C., & Smoller, J. W. (2019). Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample mendelian randomization study. *JAMA psychiatry*,– 76(4).– P. 399–408.

Dhar, A. K., & Barton, D. A. (2016). Depression and the link with cardiovascular disease. *Frontiers in psychiatry*,– 7.– 33 p.

Dinas, P. C., Koutedakis, Y., & Flouris, A. D. (2011). Effects of exercise and physical activity on depression. *Irish journal of medical science*,– 180.– P. 319–325.

Goodwin, R. D., Dierker, L. C., Wu, M., Galea, S., Hoven, C. W., & Weinberger, A. H. (2022). Trends in U.S. Depression Prevalence From 2015 to 2020: The Widening Treatment Gap. *American journal of preventive medicine*,– 63(5).– P. 726–733. URL: https://doi.org/10.1016/j.amepre.2022.05.014

Hallgren, M., Nguyen, T. T., Owen, N., Stubbs, B., Vancampfort, D., Lundin, A., Dunstan, D., Bellocco, R., & Lagerros, Y. T. (2020). Cross-sectional and prospective relationships of passive and mentally active sedentary behaviours and physical activity with depression. *The British journal of psychiatry: the journal of mental science*,– 217(2).– P. 413–419. URL: https://doi.org/10.1192/bjp.2019.60

Kandola, A., Ashdown-Franks, G., Hendrikse, J., Sabiston, C. M., & Stubbs, B. (2019). Physical activity and depression: Towards understanding the antidepressant mechanisms of physical activity. *Neuroscience & Biobehavioral Reviews*,– 107.– P. 525–539.

Keating, X. D., Zhou, K., Liu, X., Hodges, M., Liu, J., Guan, J., ... & Castro-Piñero, J. (2019). Reliability and concurrent validity of global physical activity questionnaire (GPAQ): a systematic review. *International journal of environmental research and public health*,– 16(21).– 4128 p.

Lagerros, Y. T., Mucci, L. A., Bellocco, R., Nyrén, O., Bälter, O., & Bälter, K. A. (2006). Validity and reliability of self-reported total energy expenditure using a novel instrument. *European journal of epidemiology*,– 21.– P. 227–236.

Lagerros, Y. T., Bellocco, R., Adami, H. O., & Nyrén, O. (2009). Measures of physical activity and their correlates: the Swedish National March Cohort. *European journal of epidemiology*,– 24.– P. 161–169.

Mirel, L. B., Mohadjer, L. K., Dohrmann, S. M., Clark, J., Burt, V. L., Johnson, C. L., & Curtin, L. R. (2013). National Health and Nutrition Examination Survey: estimation procedures, 2007–2010. *Vital and health statistics. Series 2, Data evaluation and methods research*,– (159).– P. 1–17.

Mumba, M. N., Nacarrow, A. F., Cody, S., Key, B. A., Wang, H., Robb, M., Jurczyk, A., Ford, C., Kelley, M. A., & Allen, R. S. (2021). Intensity and type of physical activity predicts depression in older adults. *Aging & mental health*,– 25(4).– P. 664–671. URL: https://doi.org/10.1080/13607863.2020.1711861

Pearce, M., Garcia, L., Abbas, A., Strain, T., Schuch, F. B., Golubic, R., ... & Woodcock, J. (2022). Association between physical activity and risk of depression: a systematic review and meta-analysis. *JAMA psychiatry*.

Rutherford, E. R., Vandelanotte, C., Chapman, J., & To, Q. G. (2022). Associations between depression, domain-specific physical activity, and BMI among US adults: NHANES2011–2014 cross-sectional data. *BMC Public Health*,– 22(1).– P. 1–9.

Sakashita, T., & Oyama, H. (2019). Developing a hypothetical model for suicide progression in older adults with universal, selective, and indicated prevention strategies. *Frontiers in psychiatry*,– 10.– 161 p.

Ströhle, A. (2009). Physical activity, exercise, depression and anxiety disorders. *Journal of neural transmission*,– 116.– P. 777–784.

Stuart, A. (2022, September 27). *Physical effects & symptoms of depression*. WebMD. URL: https://www.webmd.com/depression/how-depression-affects-your-body

Team, T. H. (2017, August 3). *Depression risks: Medical, social, and substance factors*. Healthline. URL: https://www.healthline.com/health/depression/risk-factors

Teychenne, M., Ball, K., & Salmon, J. (2008). Physical activity and likelihood of depression in adults: a review. *Preventive medicine*,– 46(5).– P. 397–411.

The economic cost of depression is increasing; direct costs are only a small part. (2021, May 27). URL: https://www.psychiatry.org:443/news-room/apa-blogs/the-economic-cost-of-depression-is-increasing

Wei, J., Lu, Y., Li, K., Goodman, M., & Xu, H. (2022). The associations of late-life depression with all-cause and cardiovascular mortality: The NHANES2005–2014. *Journal of affective disorders*,– 300.– P. 189–194.

World Health Organization. (n.d.). *Depression*. World Health Organization. URL: https://www.who.int/health-topics/depression
Zhang, S., Xiang, K., Li, S., & Pan, H. F. (2021). Physical activity and depression in older adults: the knowns and unknowns. *Psychiatry research*,– 297.– 113738 p.
Email: Liuyix07@gmail.com

submitted 12.06.2024;
accepted for publication 26.06.2024;
published 22.07.2024

Contact: Liuyix07@gmail.com

The European Journal of Biomedical and Life Sciences 2024, No 1–2
ISSN 2310-5674

PREMIER Publishing
ppublishing.org

DOI:10.29013/EJBLS-24-1.2-13-23

UNDERSTANDING INFLUENZA SUSCEPTIBILITY: AN ANALYSIS OF DEMOGRAPHIC, SOCIOECONOMIC, AND HEALTH FACTORS USING NHANES DATA

Zitong Zheng [1]

[1] Shenzhen Middle School

Cite: *Zitong Zheng. (2024). Understanding Influenza susceptibility: An analysis of demographic, socioeconomic, and health factors using NHANES data. The European Journal of Biomedical and Life Sciences 2024, No 1–2.* https://doi.org/10.29013/EJBLS-24-1.2-13-23

Abstract

Background: Influenza continued to be a significant global health concern in recent 2 years, leading to significant economic burdens and public health challenges globally. Thus, understanding the complex risk factors associated with influenza susceptibility is crucial for effective prevention and mitigation strategies.

Objects: Our study evaluated the association of several risk factors with infection related to influenza, including socioeconomic, demographic and medical conditions.

Method: We examined data from the National Health and Nutrition Examination Survey (NHANES) spanning 5 survey cycles, from 2005 to 2018, to provide a comprehensive assessment of influenza susceptibility factors. Both univariate and multivariate logistic regression models were utilized to analyze the relationships between influenza outcomes and various factors in the general population. In our model, we also conducted Chi-square test for independence to evaluate the interaction between explanatory variables.

Results: In the multivariate logistic regression analysis, the risk of an influenza infection varies in different races: Compared to Hispanic, influenza susceptibility is significant within Non-Hispanic Black, Non-Hispanic Asian and multiracial groups (OR =1.96, 95% CI [1.5,2.55], *P* value <0.0001 for Non-Hispanic Black, OR = 1.43, 95% CI [1.11,1.84], *P* value = 0.0064 for Non-Hispanic Asian and multiracial). While Non-Hispanic White (OR = 1.34, 95% CI [0.92,1.97], *P* value = 0.134) shows no statistical significance associated with influenza. For medical conditions, individuals with Chronic Bronchitis or Asthma displayed higher probability of flu infection (OR = 1.34, 95% CI [1.01, 1.79], *P* value = 0.0483 for asthma, and OR = 1.77, 95% CI [1.14, 2.75], *P* value = .0137 for chronic bronchitis).

Conclusion: Ethnicity, respiratory diseases (Chronic Bronchitis and Asthma) were associated with higher odds of flu infection. We found no interaction between race and respiratory illness in their association with influenza. Targeted interventions addressing socioeconomic disparities and respiratory health are crucial for reducing the burden of influenza morbidity and mortality.

Keywords: *influenza, risk factors, racial disparity, respiratory diseases, logistic regression*

Introduction

Influenza, commonly known as the flu, is a contagious respiratory illness transmitted through airborne droplets during coughing or sneezing, or by contact with contaminated surfaces. It is caused by four main types of viruses: influenza A, influenza B, influenza C, and influenza D. Influenza A and B are responsible for seasonal flu outbreaks globally (CDC, Burden of Influenza, Centers for Disease Control and Prevention (2024). For instance, the H1N1 influenza pandemic, caused by an influenza A virus, now becomes a regular human flu virus and continues to circulate seasonally worldwide. Treatments of influenza include Oseltamivir, also known as Tamiflu, which is an antiviral medication against influenza A and B (Flu and Older Adults — NFID, Https://Www.Nfid.Org/ (n.d.). 2024), as well as humidifiers (Klein S.L., Hodgson A., Robinson D.P., 2012). Additionally, vaccination remains the most effective preventive measure against influenza. Given the capacity for flu viruses to mutate, new vaccines are developed annually to target prevalent strains (Fiscella K., Dressler R., Meldrum S., Holt K., 2007).

Despite these treatments, influenza imposes a significant worldwide economic burden each year. According to the Centers for Disease Control and Prevention (CDC), the annual economic cost of seasonal influenza in the United States ranges from $11 to $18 billion, encompassing direct costs such as medical treatment, hospitalizations, and medications, as well as indirect costs like lost productivity due to missed workdays and school absences (CDC, Burden of Influenza, Centers for Disease Control and Prevention (2024). Furthermore, influenza stands as a challenge to public health, particularly impacting vulnerable populations such as young children, elderly individuals, and those with underlying medical conditions. Severe complications including pneumonia, bronchitis, and exacerbation of chronic conditions like asthma, diabetes, or chronic obstructive pulmonary disease (COPD) can arise from influenza infections (Aligne C.A., 2016).These complications and comorbidity can result in hospitalizations and even fatalities. The World Health Organization (WHO) estimates that seasonal influenza epidemics result in approximately 3–5 million cases of severe illness globally each year, with respiratory deaths attributed to flu-related causes ranging from 290,000 to 650,000 annually (Vaughan E., Tinker T., 2009).

Therefore, a comprehensive understanding of the complex risk factors associated with influenza is imperative. Given its rapid transmission and potential for severe illness and mortality, identifying and studying these risk factors is crucial for developing effective strategies to mitigate the impact of influenza outbreaks and pandemics. This research aims to explore a variety of influenza risk factors, spanning individual, and societal dimensions. Through synthesizing empirical evidence on associated risk factors, our goal is to provide insights that can inform targeted interventions and policy initiatives, thereby reducing the burden of influenza morbidity and mortality.

Several factors collectively contribute significantly to the susceptibility to influenza, encompassing demographic variables, socioeconomic status, and underlying health conditions. Previous research has elucidated the associations between these factors and flu infection. Age and gender are pivotal determinants, with adults aged 65 years and older being identified as a high-risk group for developing severe complications from influenza (Flu and Older Adults — NFID, Https://Www.Nfid.Org/ (n.d.). 2024), while females typically experience greater morbidity and mortality during influenza outbreaks and pandemics (Klein S.L., Hodgson A., Robinson D.P., 2012). In addition, Fiscella's study highlighted racial and ethnic disparities that contribute to poorer health outcomes and higher mortality rates among minority groups (Fiscella K., Dressler R., Meldrum S., Holt K., 2007). From a socioeconomic perspective, overcrowded living conditions serve as a vehicle for respiratory infection transmission, including the flu, and can exacerbate disease severity (Aligne C.A., 2016). Low-income communities often experience worse health conditions and have a higher likelihood of premature mortality during epidemics (Adler N.E., Newman K., 2002). Moreover, individuals with lower educational levels tend to exhibit inferior receptiveness to precautionary health information (Vaughan E., Tinker T., 2009), thereby increasing their susceptibility to infection during pandemics.

Furthermore, medical conditions also account for vulnerability in flu pandemic. On one hand, people with regular health care had a higher probability of receiving influenza vaccination than those without health care coverage, which reduces their vulnerability in flu infection (Gurel-Headley M., Mamisashvili M., CarlLee S., Reece S., Chapman C., Kraleti S., Andersen J.A., Selig J.P., Willis D.E., Li J., McElfish P.A. 2023). On the other hand, respiratory conditions such as chronic bronchitis are associated with flu infections (Stark J.E., Heath R.B., Curwen M.P., 1965), while individuals with asthma are at a higher risk of having complications from the flu, such as pneumonia (Flu (Influenza), Asthma & Allergy Foundation of America (n.d.)). Besides, individuals with heart disease are more likely to develop complications from the flu, including pneumonia, bronchitis and lung failure (National Health and Nutrition Examination Survey (NHANES) — Healthy People 2030).

Building upon the existing research investigating various risk factors, this study addresses a significant gap regarding the influence of asthma and chronic bronchitis on flu risk across different racial groups. Our approach involves building logistic regression models to analyze the relationships between flu outcomes and the predictors of asthma, chronic bronchitis, and race. Additionally, we tested potential interactions between race and these respiratory conditions. By synthesizing evidence from population-representative data, we aspire to provide insights that can guide targeted interventions and policy measures to alleviate the impact of influenza morbidity and mortality. However, there remains a crucial need for further study on the intersection of social determinants of health, particularly in the context of influenza preparedness, response strategies, and post-flu recovery efforts. Future research should delve deeper into these determinants to develop effective solutions, and assist decision-makers and service organizations in addressing the needs of populations adversely affected by pandemic influenza.

Methods

2.1. Study population

The National Health and Nutrition Examination Survey (NHANES) is a cross-sectional, population-based health survey of the non-institutionalized U.S. civilian population conducted by the U.S. National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (National Health and Nutrition Examination Survey (NHANES) — Healthy People 2030). The purpose of this survey is to evaluate the nutritional status and health of adults and children every two years. The NHANES program enrolled a nationally representative sample of the U.S. population every survey cycle and included interviews, examinations, and laboratory tests of serum and urine samples. We obtained the data from NHANES for the years 2005–2008 and 2011–2018 which span 5 cycles of surveys.

In each NHANES data cycle, each subsample was representative of the complete population, and the corresponding weight for the subsample was calculated after considering the additional step of sampling, the unequal probability of selection, and the non-response rate (Centers for Disease Control and Prevention, 2017). A total of 71,058 people participated in the study throughout seven cycles, with 39,221 of them aged ≥ 20 years. Individuals with missing HTN and urine OH-PAHs data were not included in the final design. Finally, 8951 participants of the NHANES2003–2016 were included in the final examinations (Supplementary Fig. S1). The NHANES representatives gathered information about prospective confounders using a self-reported questionnaire. Body parameters were measured in the Mobile Examination Center (MEC) by NHANES staff who had been trained. The NHANES procedure was approved by the National Center for Health Statistics (NCHS) Institutional Review Committee, and signed informed consent forms were acquired.

2.2 Dependent variable

The dependent variable in this study is whether an individual has influenza or pneumonia during the past 30 days. The question is administered to participants as, “Did you have flu, pneumonia, or ear infections that started during those 30 days?” Participants who answered “1” were recorded as «Yes,» and “2” as «No.»

2.3 Independent variables

The independent variables in this study were selected based on risk factors reported

in the literature, and classified into five broad categories: 1) demographic *i.e.* age, gender, race (Hispanic, Non-Hispanic Black, Non-Hispanic White, and other races); 2) socioeconomic *i.e.* education (less than high school, High school graduate, College, College graduate); 3) Disease: Diabetes, Chronic bronchitis, Coronary heart disease; and 4) other characteristics *i.e.* Family size, Marital Status (Married, Never Married), Routine Healthcare.

2.4 Statistical analysis

The distributions of demographic variables in total sample and population with and without flu were described using the numbers of sample (n) and proportions of each category (%). The association between the individual variables and outcome were determined by weighted Chi Squared tests. In order to determine the odds ratio and 95% confidence intervals between independent variables and outcome, we conducted univariate and multiple logistic regression models using gender, education, age, marital status, family size, diabetes, coronary heart disease, asthma, chronic bronchitis. For univariate models, only one independent variable was used to fit the model, whereas the multivariate logistic regression model also included all other covariates to adjust for the potential confounding effects. The interaction effects between the statistically significant main effects were also tested but no significance was found. The Chi-squared tests and logistic regression models were adjusted to incorporate weighted measures, ensuring that the estimates reflected more precise standard errors.

Multivariate logistic regression analysis was used to adjust for the various independent variables. Also, we calculated the *P* value, and odds ratios (OR) with their corresponding 95% confidence intervals (CI). A *P* value less than 0.05 was considered as statistical significance. All analyses were conducted using the R version (Using the R Statistical Computing Environment to Teach Social Statistics Courses).

Results

Table 1 shows the sample characteristics of study in total and by whether participants have had the flu during the past 30 days. Factors such as gender, education, age, marital status, family number, diabetes and coronary heart disease were not significantly associated with flu. For the race of all participants, 5.8% were Non-Hispanic White, 10.3% were Non-Hispanic Black, 73.7% were Hispanic, and 10.2% were other races (Non-Hispanic Asian and multiracial). However, the proportions for each race were significantly different between participants with and without flu ($P < .001$). For instance, the proportion of Non-Hispanic Black in the group with flu was about 16.9%, which was significantly higher than that in the non-flu group (10%). When considering diseases, we found that for all the participants, 14.5% people have asthma, and 85.5% people don't have asthma. However, the proportion of people having asthma among people having flu (20.1%) was significantly higher ($P = .005$) than people having asthma among people having no flu (14.3%). For 13546 participants, 7.5% people reported to have chronic bronchitis. But people having chronic bronchitis had a significantly larger proportion among people having flu (12.8%) than people having no flu (7.3) with a *P* value of 0.016.

Table 1. *Sample characteristics in all participants and groups by flu status.*

Variables	All n(%)	Flu Yes	No	P value
Cycle				0.814
2005–2006	2422(18.3)	137(21.6)	2285(18.2)	
2007–2008	2798(17.3)	123(15.1)	2675(17.4)	
2011–2012	1964(15.4)	84(14.7)	1880(15.5)	
2013–2014	2267(17.1)	107(18.6)	2160(17.1)	
2015–2016	2108(15.9)	101(15.3)	2007(16)	
2017–2018	1987(15.9)	76(14.8)	1911(16)	

Variables	All n(%)	Flu Yes	No	P value
Gender				0.313
Male	6476(46.4)	256(43.5)	6220(46.5)	
Female	7070(53.6)	372(56.5)	6698(53.5)	
Education				0.156
Less than high school	1885(7.1)	114(8.8)	1771(7.1)	
College	4108(33.8)	157(31.3)	3951(33.9)	
College graduate	5132(45.2)	220(43.2)	4912(45.3)	
High school graduate	2421(13.9)	137(16.8)	2284(13.8)	
Age				0.086
20–34	1829(14.8)	85(14.7)	1744(14.8)	
35–49	3193(28.6)	157(31.2)	3036(28.4)	
50–65	4328(33.3)	224(35.7)	4104(33.2)	
above 65	4196(23.3)	162(18.3)	4034(23.5)	
Poverty income ratio				0.057
<1.3	4635(22.9)	271(28.9)	4364(22.7)	
1.3–3.5	3078(34.6)	116(30.9)	2962(34.8)	
>3.5	5833(42.4)	241(40.3)	5592(42.5)	
Race				<0.001
Hispanic	6521(73.7)	223(64)	6298(74.1)	
Non-Hispanic Black	2651(10.3)	182(16.9)	2469(10)	
Non-Hispanic White	1431(5.8)	77(6.7)	1354(5.7)	
Other races	2943(10.2)	146(12.5)	2797(10.1)	
Marital Status				0.376
Divorced	2174(15.5)	110(16.8)	2064(15.5)	
Married	8916(71.2)	378(68.3)	8538(71.3)	
Never Married	1700(9.2)	92(9.6)	1608(9.2)	
Other	756(4.1)	48(5.3)	708(4.1)	
Family size				0.006
1	2549(16.6)	122(17.4)	2427(16.5)	
2	4246(34.5)	148(26.7)	4098(34.8)	
3–6	6170(46.1)	325(51.6)	5845(45.8)	
>7	581(2.9)	33(4.4)	548(2.9)	
Diabetes				0.365
No	11196(86.8)	496(85)	10700(86.9)	
Yes	2350(13.2)	132(15)	2218(13.1)	
Routine healthcare				0.815
No	1759(12.6)	83(13)	1676(12.6)	
Yes	11787(87.4)	545(87)	11242(87.5)	
Asthma				0.005
No	11633(85.5)	498(79.9)	11135(85.7)	
Yes	1913(14.5)	130(20.1)	1783(14.3)	
Chronic Bronchitis				0.016
No	12568(92.5)	560(87.2)	12008(92.7)	
Yes	978(7.5)	68(12.8)	910(7.3)	

Variables	All n(%)	Flu Yes	No	P value
Coronary heart disease				0.874
No	12779(95.3)	593(95.4)	12186(95.3)	
Yes	767(4.7)	35(4.6)	732(4.7)	

Table 2 presents the results of univariate logistic regression models examining the relationship between each independent variable and flu occurrence. Among the variables analyzed, several were found to be statistically significant in their association with flu, including race, poverty income ratio (PIR), and medical conditions related to the respiratory system (asthma and chronic bronchitis). Specifically, a higher poverty income ratio (> 3.5) was associated with a lower likelihood of flu occurrence (OR=0.74, 95% CI [0.57, 0.97], P = .031). For race, Non-Hispanic Black and other races (Non-Hispanic Asian and multiracial) had a higher prevalence of flu compared to Hispanic (OR =1.96, 95% CI [1.5, 2.55], P < .001 for Non-Hispanic Black, OR = 1.43, 95% CI [1.11, 1.84], P = .006 for other races, and OR = 1.34, 95% CI [0.92, 1.97], P = .134 for Hispanic). Furthermore, asthma showed a statistically significant association with flu, with a 51% higher probability of flu occurrence among individuals with asthma compared to those without asthma (95% CI [1.19, 1.91], P = .001), and chronic bronchitis was also significantly associated with flu, with an 85% higher odds among individuals with chronic bronchitis compared to those without (OR = 1.85, 95% CI [1.24, 2.75], P = .003), indicating a heightened risk of flu infection among individuals with this condition.

Table 2. *Univariate analyses of flu status and risk factors*

Variables	OR*	LCL**	UCL**	P value
Cycle				
2005–2006		Ref.		
2007–2008	0.73	0.44	1.19	0.203
2011–2012	0.8	0.44	1.44	0.453
2013–2014	0.92	0.58	1.45	0.715
2015–2016	0.81	0.49	1.32	0.393
2017–2018	0.78	0.47	1.3	0.335
Gender				
Female		Ref.		
Male	0.89	0.7	1.12	0.314
Education				
Less than high school		Ref.		
College	0.75	0.53	1.06	0.104
College graduate	0.77	0.57	1.04	0.091
High school graduate	0.98	0.74	1.3	0.889
Age				
20–34		Ref.		
35–49	1.1	0.79	1.53	0.558
50–65	1.08	0.75	1.55	0.677
above 65	0.78	0.56	1.09	0.153

Variables	OR*	LCL**	UCL**	P value
poverty income ratio				
<1.3		Ref.		
1.3–3.5	0.7	0.49	1	0.051
>3.5	0.74	0.57	0.97	0.031
Race				
Hispanic		Ref.		
Non-Hispanic Black	1.96	1.5	2.55	<0.001
Non-Hispanic White	1.34	0.92	1.97	0.134
Other races	1.43	1.11	1.84	0.006
Marital Status				
Divorced		Ref.		
Married	0.88	0.64	1.22	0.448
Never Married	0.96	0.65	1.41	0.841
Other	1.21	0.81	1.79	0.357
Family size				
1		Ref.		
2	0.73	0.53	1.01	0.06
3–6	1.07	0.79	1.46	0.663
>7	1.47	0.96	2.26	0.081
Diabetes				
No		Ref.		
Yes	1.16	0.85	1.6	0.35
Routine health-care				
No		Ref.		
Yes	0.96	0.67	1.37	0.813
Asthma				
No		Ref.		
Yes	1.51	1.19	1.91	0.001
Chronic bron-chitis				
No		Ref.		
Yes	1.85	1.24	2.75	0.003
Coronary heart disease				
No		Ref.		
Yes	0.97	0.62	1.5	0.876

** Odds ratio (OR)*
** Limits of the 95% confidence interval. Lower confidence limit (LCL) and upper confidence limit (UCL)*

Table 3 displays the results of multiple logistic regression models including all variables to adjust for potential confounding effects. Variables such as gender, education, age, marital status, family number, diabetes, routine healthcare, and coronary heart disease were not significantly associated with flu after adjusting for other covariates. However, the effects of PIR on flu were attenuated after adjusting for all the covariates, indicating the effects of this variable might be confounded. Among these variables, three variables remained significant risk factors for flu in the multivariate logistic model: Race, Asthma, Chronic bronchitis. Non-Hispanic Black and other race (Non-Hispanic Asian, multiracial) were significantly associated with higher susceptibility to flu compared to Hispanic (OR = 1.94, 95% CI [1.44, 2.62], $P < .001$ for Non-Hispanic Black, and OR =1.34, 95% CI [1.05, 1.71], $P = .02$ for Non-Hispanic Asian and multiracial, respectively). Additionally, asthma and chronic bronchitis were also independent risk factors of flu (OR = 1.34, 95%, CI [1.01, 1.79], $P = .048$ for asthma, and OR = 1.77, 95% CI [1.14, 2.75], $P = .014$ for chronic bronchitis, respectively).

Table 3. *Multivariate logistic regression analysis of risk factors associated with influenza*

Variables	OR	LCL	UCL	P value
Cycle				
2005–2006		Ref.		
2007–2008	0.71	0.45	1.14	0.16
2011–2012	0.78	0.43	1.4	0.411
2013–2014	0.88	0.56	1.37	0.569
2015–2016	0.77	0.48	1.22	0.268
2017–2018	0.74	0.44	1.24	0.262
Gender				
Female		Ref.		
Male	0.92	0.72	1.19	0.53
Education				
Less than high school		Ref.		
College	0.99	0.69	1.42	0.964
High school graduate	1.18	0.87	1.61	0.342
College graduate	1.02	0.73	1.42	0.912
Age				
20–34		Ref.		
35–49	1.12	0.81	1.57	0.496
50–65	1.22	0.86	1.75	0.268
above 65	0.91	0.61	1.35	0.628
poverty income ratio				
<1.3		Ref.		
1.3–3.5	0.91	0.64	1.3	0.603
>3.5	0.89	0.66	1.19	0.437
Race				
Hispanic		Ref.		
Non-Hispanic Black	1.94	1.44	2.62	<0.0001
Non-Hispanic White	1.32	0.9	1.91	0.156
Other races	1.34	1.05	1.71	0.02
Marital Status				
Divorced		Ref.		
Married	0.93	0.68	1.29	0.677
Never Married	1.07	0.7	1.63	0.759
Other	1.05	0.71	1.57	0.801

Variables	OR	LCL	UCL	P value
Family size				
1		Ref.		
2	0.82	0.6	1.11	0.204
3–6	1.08	0.76	1.53	0.659
>7	1.31	0.83	2.07	0.25
Diabetes				
No		Ref.		
Yes	1.11	0.79	1.55	0.542
Routine healthcare				
No		Ref.		
Yes	1.06	0.73	1.54	0.744
Asthma				
No		Ref.		
Yes	1.34	1.01	1.79	0.048
Chronic bronchitis				
No		Ref.		
Yes	1.77	1.14	2.75	0.014
Coronary heart disease				
No		Ref.		
Yes	1.03	0.63	1.68	0.907

In conclusion, our study identified race, asthma, and chronic bronchitis as major independent predictors of flu. While we also tested the interaction among asthma, race, and chronic bronchitis, no significant interaction effects were found between any of these variables (data not shown).

Discussion

This study investigated the association between influenza and various risk factors encompassing demographic variables (gender, age, race, family size), socioeconomic status indicators (education, marital status, access to routine healthcare) and medication conditions (diabetes, chronic bronchitis, coronary heart disease, and asthma). We found that risk factors including race, chronic bronchitis and asthma are significantly associated with influenza susceptibility. Specifically, individuals of Non-Hispanic Black and other races (Non-Hispanic Asian and multiracial) exhibited a notably higher likelihood of flu compared to Hispanics. Other two respiratory diseases like asthma and chronic bronchitis also showed statistical significance in association with flu, displaying an increased susceptibility to flu infection.

Whereas some studies exploring the risk factors of influenza focused on socioeconomic risk factors (Mamelund S.-E., Shelley-Egan C., Rogeberg O., 2021; Pandemic influenza and socioeconomic disparities: Lessons from 1918 Chicago. 2024) or emphasized the relationship between respiratory disease and flu (Peteranderl C., Herold S., Schmoldt C., 2016), our research examined multiple risk factors simultaneously using logistic regression models and a large dataset representing the non-institutionalized US population. Notably, contrary to some prior studies that linked lower literacy to higher flu prevalence (3–1–4–2_dm_high_risk_populations.pdf, (n.d.)) our study did not find significant differences in flu prevalence based on education levels. This discrepancy may stem from variations in survey design and demographic characteristics of the study samples. Regarding race, our study's findings regarding racial disparities in flu infection differed from some previous findings. One study of the 1918 pandemic indicated smaller racial disparities in flu-related mortality (Eiermann M., Wrigley-Field E., Feigenbaum J. J., Helgertz J., Hernandez E., Boen C. E., 2022) (and another study highlighted higher hospitalization rates among Blacks and Native Americans during the H1N1 pandemic compared to Whites (Quinn S. C., Kumar S., Freimuth V. S., Musa D., Casteneda-Angarita N.,

Kidwell K., 2011), whereas our study suggested that Non-Hispanic Black and other races (Non-Hispanic Asian and multiracial) shows significantly high susceptibility to influenza. For socioeconomic conditions, although one study found a strong association between income level and probability of having flu (Gaskin C. M., Woods D. R., Ghosh S., Watson S., Huber L. R., 2023), our regression model did not find a direct association between income and flu, suggesting a potential confounding variable related to race influencing this relationship. Additionally, overcrowding has been previously linked to flu transmission (Aligne C. A., 2016), however, our model showed no association between numbers of family members and flu prevalence.

Nevertheless, our research findings aligned with prior studies regarding the relationship between medical conditions and influenza. One study revealed that asthmatic children suffer from more frequent and severe virus-induced illnesses (Olenec J. P., Kim W. K., Lee W.-M., Vang F., Pappas T. E., Salazar L. E. P., Evans M. D., Bork J., Roberg K., Lemanske R. F., Gern J. E., 2010), while another study identified asthma as a risk factor for hospitalization during the influenza pandemic (Frontiers | Influenza in Asthmatics: For Better or for Worse? (n.d.). 2024). These findings are consistent with the significant impact of asthma observed in our study. Moreover, research by Stark, Heath, and Curwen suggested that chronic bronchitis increases susceptibility during flu pandemics (Stark J. E., Heath R. B., Curwen M. P., 1965), which is in accordance with our model.

Furthermore, our study boasts several strengths. Firstly, we conducted a cross-sectional exploration on the multiple risk factors of influenza, encompassing socioeconomic, demographic, and medical factors, contributing to a more comprehensive understanding of influenza risk factors. Moreover, our data were derived from a large national representative database, enhancing the generalizability of our findings to diverse populations, including minorities. Furthermore, we explored potential interaction effects among various factors, although no significant interactions were detected among the primary risk factors.

However, our study has limitations. On one hand, certain factors such as working conditions, pregnancy, language, culture (linked to ethnicity), disabilities, and other pre-existing medical conditions were not accounted for in our analysis. On the other hand, as our study is cross-sectional, the temporal causality cannot be inferred from the analysis in this study.

In brief, our findings revealed varying levels of susceptibility to influenza associated with racial disparities and respiratory diseases. We anticipate that policymakers and healthcare professionals will enact precautionary measures to mitigate ethnic disparities and respiratory-related illness during pandemics. Further investigation is warranted to elucidate the underlying mechanisms, including molecular and genetic pathways, through which respiratory conditions like asthma and COPD influence influenza outcomes.

References

CDC, Burden of Influenza, Centers for Disease Control and Prevention (2024). URL: https://www.cdc.gov/flu/about/burden/index.html (accessed March 18, 2024).

Flu and Older Adults – NFID, URL: https://Www.Nfid.Org/ (n.d.). URL: https://www.nfid.org/infectious-diseases/flu-and-older-adults/ (accessed April 2, 2024).

Klein S. L., Hodgson A., Robinson D. P. Mechanisms of sex disparities in influenza pathogenesis, J Leukoc Biol 92 (2012) 67–73. URL: https://doi.org/10.1189/jlb.0811427.

Fiscella K., Dressler R., Meldrum S., Holt K. Impact of influenza vaccination disparities on elderly mortality in the United States, Prev Med 45 (2007) 83–87. URL: https://doi.org/10.1016/j.ypmed.2007.03.007.

Aligne C. A. Overcrowding and Mortality During the Influenza Pandemic of 1918: Evidence From US Army Camp A. A. Humphreys, Virginia, American Journal of Public Health 106 (2016) 642. URL: https://doi.org/10.2105/AJPH.2015.303018.

Adler N. E., Newman K. Socioeconomic Disparities In Health: Pathways And Policies, Health Affairs 21 (2002) 60–76. URL: https://doi.org/10.1377/hlthaff.21.2.60.
Vaughan E., Tinker T. Effective Health Risk Communication About Pandemic Influenza for Vulnerable Populations, Am J Public Health 99 (2009) S324–S332. URL: https://doi.org/10.2105/AJPH.2009.162537.
Gurel-Headley M., Mamisashvili M., CarlLee S., Reece S., Chapman C., Kraleti S., Andersen J. A., Selig J. P., Willis D. E., Li J., McElfish P. A. Associations between Influenza Vaccination and Health Care Access among Adults in the United States, Vaccines (Basel) 11 (2023) 416. URL: https://doi.org/10.3390/vaccines11020416.
Stark J. E., Heath R. B., Curwen M. P. Infection with Influenza and Parainfluenza Viruses in Chronic Bronchitis, Thorax – 20. 1965. – P. 124–127.
Flu (Influenza), Asthma & Allergy Foundation of America (n.d.). URL: https://aafa.org/asthma/asthma-triggers-causes/respiratory-infections-flu-cold-asthma/flu-influenza/ (accessed April 2, 2024).
National Health and Nutrition Examination Survey (NHANES) – Healthy People 2030 | health.gov, (n.d.). URL: https://health.gov/healthypeople/objectives-and-data/data-sources-and-methods/data-sources/national-health-and-nutrition-examination-survey-nhanes (accessed March 31, 2024).
Mamelund S.-E., Shelley-Egan C., Rogeberg O. The association between socioeconomic status and pandemic influenza: Systematic review and meta-analysis, PLOS ONE – 16. (2021). e0244346. URL: URL: https://doi.org/10.1371/journal.pone.0244346.
Pandemic influenza and socioeconomic disparities: Lessons from 1918 Chicago | PNAS, (n.d.). URL: https://www.pnas.org/doi/full/10.1073/pnas.1616537113 (accessed April 18, 2024).
Peteranderl C., Herold S., Schmoldt C. Human Influenza Virus Infections, Semin Respir Crit Care Med – 37. (2016).– P. 487–500. URL: https://doi.org/10.1055/s-0036–1584801.
3-1-4-2_dm_high_risk_populations.pdf, (n.d.). URL: https://www.redcross.ca/crc/documents/3-1-4-2_dm_high_risk_populations.pdf (accessed April 21, 2024).
Eiermann M., Wrigley-Field E., Feigenbaum J. J., Helgertz J., Hernandez E., Boen C. E. Racial Disparities in Mortality During the 1918 Influenza Pandemic in United States Cities, Demography 59 (2022) 1953–1979. URL: https://doi.org/10.1215/00703370–10235825.
Quinn S. C., Kumar S., Freimuth V. S., Musa D., Casteneda-Angarita N., Kidwell K. Racial Disparities in Exposure, Susceptibility, and Access to Health Care in the US H1N1 Influenza Pandemic, Am J Public Health 101 (2011) 285–293. URL: https://doi.org/10.2105/AJPH.2009.188029.
Gaskin C. M., Woods D. R., Ghosh S., Watson S., Huber L. R. The Effect of Income Disparities on Influenza Vaccination Coverage in the United States, Public Health Rep 138 (2023) 85–90. URL: https://doi.org/10.1177/00333549211069190.
Olenec J. P., Kim W. K., Lee W.-M., Vang F., Pappas T. E., Salazar L. E.P., Evans M. D., Bork J., Roberg K., Lemanske R. F., Gern J. E. Weekly monitoring of children with asthma for infections and illness during common cold seasons, Journal of Allergy and Clinical Immunology 125 (2010) 1001–1006.e1. URL: https://doi.org/10.1016/j.jaci.2010.01.059.
Frontiers | Influenza in Asthmatics: For Better or for Worse?, (n.d.). URL: URL: https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2018.01843/full (accessed April 19, 2024).

submitted 27.04.2024;
accepted for publication 11.05.2024;
published 22.07.2024

Contact: Theresazhengq@gmail.com

The European Journal of Biomedical and Life Sciences 2024, No 1–2
ISSN 2310-5674

PREMIER Publishing
ppublishing.org

Section 2. Life sciences

DOI:10.29013/EJBLS-24-1.2-24-35

UNVEILING THE COMPLEX SNP LANDSCAPE IN LYMPHOMA THROUGH COMPREHENSIVE ANALYSIS FOR FUTURE FUNCTIONAL INTERPRETATION AND THERAPEUTIC APPLICATIONS

Ethan Zhang [1]

1

Cite: *Ethan Zhang. (2024). Unveiling the Complex SNP Landscape in Lymphoma Through Comprehensive Analysis for Future Functional Interpretation and Therapeutic Applications. The European Journal of Biomedical and Life Sciences 2024, No 1–2* https://doi.org/10.29013/EJBLS-24-1.2-24-35

Abstract:

The successful management of lymphoma increasingly hinges on a deep understanding of its genomic underpinnings, especially for early detection and targeted treatment strategies. The current study focused on the identification and functional categorization of single nucleotide polymorphisms (SNPs) specifically in lymphoma patients. Utilizing whole-genome sequencing data sourced from the Sequence Read Archive, we employed a rigorous analytical pipeline involving sequence alignment against the Homo sapiens chromosome 4 reference genome, followed by indexing, and variant calling. Our investigation led to the discovery of 1072 SNPs, a significant portion of which remain functionally uncharacterized. Among the categorized SNPs, variants were identified that have potential implications for gene regulation, splicing, and protein function. Statistical analyses revealed a significant association between these SNPs and the lymphoma cohort. Our findings offer nuanced insights into the complex genetic landscape of lymphoma and serve as a foundational platform for future research aimed at functional characterization and potential clinical applications. These identified SNPs could serve as potential biomarkers and contribute to the development of more effective diagnostic and therapeutic strategies for lymphoma.

Keywords: *Single Nucleotide Polymorphisms (SNPs), Lymphoma, TET2*

Introduction

Lymphoma

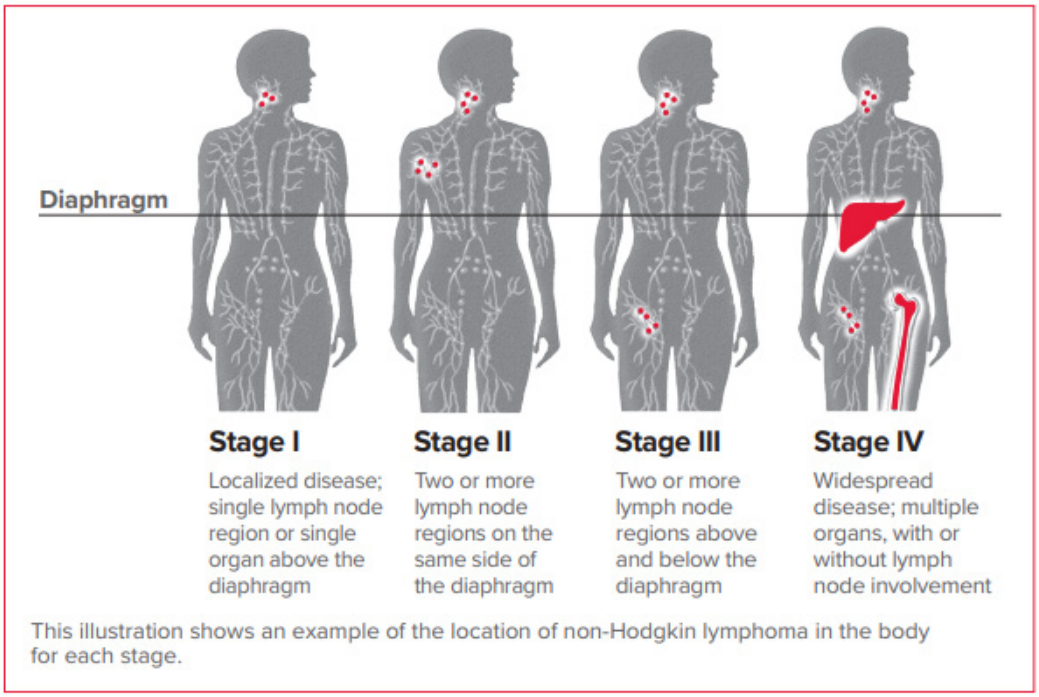
Lymphoma is a form of cancer that originates in the cells of the lymphatic system. The disease is categorized into two main types: Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL). Of the two, NHL is one of the most prevalent cancers in the United States, accounting for approximately 4% of all new cancer cases (“Key Statistics for Non-Hodgkin Lymphoma”, 2023). It is characterized by its

heterogeneous nature, comprising various malignancies that arise from the clonal proliferation of B-cell, T-cell, and natural killer cell subsets of lymphocytes at different stages of maturation. NHL represents around 4% of all malignancies and exhibits an overall survival rate of approximately 72% once diagnosed (Jamil, 2023). Likewise, the projected incidence of HL in the year 2023 is expected to encompass approximately 8,830 novel cases. Although constituting a relatively modest proportion of all newly diagnosed cancer cases, accounting for a mere 0.5%, HL bears a distinctive imprint on mortality rates. Projections indicate an estimated 900 deaths attributed to this condition every year, with 540 mortalities among males and 360 among females ("Cancer Stat Facts", 2023).

Stages of Lymphoma

Lymphoma is systematically categorized into discrete stages according to the extent of its dissemination within the body. These progressive stages serve as pivotal navigational tools for the processes of diagnosis, therapeutic strategizing, and prognostic assessment. Stage I lymphoma, delineates a condition in which cancer is localized within a confined region, *in situ* (see Figure 1). This can manifest as either a solitary cancerous lymph node or as an occurrence within a specific lymphoid organ, such as the thymus, or even within a discrete segment of an extra lymphatic organ. The malignancy's reach is strictly constrained at this juncture. Transitioning to Stage II lymphoma, the scope of disease involvement slightly broadens. At this stage, the two lymph nodes which are located on the same side of the diaphragm may be affected ("What Are the Different Stages of Lymphoma", 2023).

Figure 1. *Stages of Lymphoma.* *This diagram illustrates the four stages of Lymphoma, detailing the progression and spread of the disease. Stage I shows localized involvement of a single lymph node region, while Stage II involves two or more lymph node regions on the same side of the diaphragm. In Stage III, cancer spreads to lymph node regions on both sides of the diaphragm, and Stage IV indicates the disseminated involvement of one or more extra lymphatic organs or tissues ("NHL Staging", n.d.)*



Alternatively, the disease may traverse from one lymph node to an adjoining organ. Though the propagation surpasses that of Stage I, it continues to be largely circumscribed within a single side of the diaphragm. Progressing to Stage III, an amplified dissemination pattern emerges. This disease often infiltrates multiple lymph nodes distrib-

uted both above and below the diaphragm, with potential engagement of the spleen. The spanning of both sides of the diaphragm reflects an escalated systemic presence of the disease. Ultimately, Stage IV denotes the utmost progression of lymphoma, signaling extensive metastasis. The malignancy breaches the confines of the lymphatic system, infiltrating distant organs scattered throughout the body. This may encompass two or more remote organs, such as the liver or lungs. The pervasive and systemic nature of the disease's dispersion in Stage IV underscores its substantial impact, frequently necessitating comprehensive and multifaceted therapeutic interventions ("What Are the Different Stages of Lymphoma", 2023).

Treatments

Treatment for lymphoma may include radiation, chemotherapy, or a combination of both. It may also include immunotherapy or other new treatments. The treatment that is best for patients will depend on many factors, such as the type of lymphoma the patients have. Hodgkin lymphoma is also called Hodgkin's disease. It may be treated with chemotherapy, a combination of chemotherapy and radiation, immunotherapy, stem cell transplantation, or other treatments. Many treatment options are now available for non-Hodgkin lymphoma. These include traditional chemotherapy, targeted drugs, and novel therapies that are available only through clinical trials. Patients with aggressive B cell non-Hodgkin lymphoma is often treated with R-CHOP chemotherapy, consisting of cyclophosphamide, doxorubicin, vincristine, prednisone, and rituximab, with alternatives considered for potential heart damage. T-Cell NHL, such as T-lymphoblastic lymphoma/leukemia, involves intensive chemotherapy and may include intrathecal chemotherapy, maintenance therapy, or stem cell transplant, with close monitoring for complications such as tumor lysis syndrome ("R-CHOP", 2022; "Treating Non-Hodgkin Lymphoma", 2023). For individuals with HIV infections, who face an elevated risk of aggressive NHL, the prognosis has improved significantly, particularly with the use of highly active antiretroviral therapy (HAART). HAART enables better tolerance to chemotherapy and immunotherapy and has somewhat alleviated challenges related to low blood cell counts, allowing for the administration of full doses of chemotherapy. Despite this progress, caution is exercised in chemotherapy for HIV-infected patients with lymphoma, and blood counts are closely monitored. The interaction between HIV infection and lymphoma prognosis remains complex, with modern anti-HIV therapy contributing to improved outcomes ("Treating Non-Hodgkin Lymphoma", 2023).

Immunotherapy has emerged as a vital treatment modality for both B-cell and T-cell NHL. In B-cell lymphomas, monoclonal antibodies targeting the CD20 antigen, such as rituximab (Rituxan), obinutuzumab (Gazyva), ofatumumab (Arzerra), and ibritumomab tiuxetan (Zevalin), have been proven effective. Their administration, typically via intravenous infusion, requires careful monitoring for side effects ranging from mild infusion-related reactions to more severe manifestations like skin conditions and blood cell count anomalies (Halim & Maher, 2020; "Treating Non-Hodgkin Lymphoma", 2023). Assessment for prior hepatitis B infection and vigilance for post-treatment infections are also imperative. For T-cell lymphomas, innovative immunotherapeutic agents such as T-cell engaging bispecific antibodies (e.g., mosunetuzumab (Lunsumio), epcoritamab (Epkinly), glofitamab (Columvi)) have been developed. These antibodies simultaneously target the CD3 protein on T cells and specific antigens on lymphoma cells, enhancing the immune response against the cancer. Monitoring for potential complications, including cytokine release syndrome and neurological issues, is essential, and the treatment approach must be tailored to the specific lymphoma type and the patient's overall health ("Treating Non-Hodgkin Lymphoma", 2023).

Chemotherapy is a cornerstone in lymphoma treatment, employing various drug categories such as alkylating agents to damage DNA, corticosteroids to reduce inflammation, platinum drugs to inhibit DNA replication, purine analogs to disrupt DNA and RNA, anti-metabolites to impede cell growth, anthracyclines to intercalate DNA, and other drugs to interfere with cell division. Standard regimens may be supplemented with immu-

notherapy drugs to target specific cancer cells, enhancing efficacy ("Treating Non-Hodgkin Lymphoma", 2023). Targeted drug therapies have also been developed, aiming to inhibit specific cellular processes. Proteasome inhibitors block enzyme complexes within cells to control cell division, while histone deacetylase (HDAC) inhibitors modulate gene activity. BTK inhibitors can target specific proteins involved in B-cell lymphoma cell growth and survival. These targeted therapies, administered through various routes such as intravenous infusion or oral capsules, present a sophisticated approach for treating specific NHL types, especially when standard treatments are insufficient ("Treating Non-Hodgkin Lymphoma", 2023).

Single Nucleotide Polymorphisms

Single Nucleotide Polymorphisms (SNPs) constitute a fundamental form of genetic variation among individuals, characterized by alterations at a singular nucleotide position within the genomic DNA sequence. With a prevalence that exceeds 1% across the human genome, SNPs occur approximately once in every 1000 base pairs and significantly contribute to the genetic diversity observed among distinct human populations ("Single Nucleotide Polymorphisms", n.d.; "What are single nucleotide polymorphisms", 2022). SNPs may manifest as either transitions (purine to purine or pyrimidine to pyrimidine) or transversions (purine to pyrimidine or vice versa), and their effects can range from being silent (synonymous), leading to no change in amino acid sequence, to causing changes in amino acid composition. The biological ramifications of SNPs extend to various key genetic processes, such as modulation of promoter activity, alteration of messenger RNA (mRNA) stability, and influence on subcellular localization. As a result, SNPs have the potential to precipitate various pathological conditions by affecting gene expression or function, thereby underscoring their significance in both normal physiological variation and disease susceptibility (Degtyareva *et al.*, 2021; "What are single nucleotide polymorphisms", 2022).

Investigating the intricate link between SNPs and cancer, particularly in the context of lymphoma, reveals that SNPs within promoter regions can prompt alterations in the number of methylation loci, consequently impacting gene expression and elevating the risk of cancer. Beyond methylation dynamics, promoter region SNPs exert their influence through promoter activity modulation, transcription-factor binding, and histone modifications, inclusive of DNA methylation (Deng *et al.,* 2017; Jiraskova *et al.*, 2019). Notably, SNPs can also exert effects on exonal regions, where they possess the capacity to suppress gene transcription and translation, thereby perturbing cancer susceptibility. SNPs occurring within genes associated with DNA mismatch repair, cell cycle regulation, metabolism, and immunity further contribute to genetic susceptibility to cancer. Understanding these multifaceted relationships between SNPs and cancer is pivotal for unraveling the intricacies of cancer etiology and pathogenesis (Deng *et al.,* 2017; Jiraskova *et al.*, 2019).

The utilization of SNPs as potential markers for early lymphoma diagnosis offers promising avenues in cancer diagnostics. By leveraging SNP markers and arrays, comprehensive genome screening at a global scale may enable the identification of chromosomal anomalies encompassing copy number variants, DNA amplifications, deletions, and loss of heterozygosity. In contrast to earlier technologies such as cytogenetics or gene candidate approaches, which demonstrated limited sensitivity, the utilization of SNP markers can present a robust alternative. Genetic aberrations, including copy number variations (CNVs) and translocations, have been established to exhibit distinct associations with specific human cancers, notably prevalent in hematological malignancies such as lymphomas (Shahrabi *et al.*, 2020; Etebari *et al.*, 2015). The potential to establish specific markers for such cases holds the key to identifying early manifestations of respective cancers. Notably, the expedience and cost-effectiveness of SNP microarrays facilitate rapid and comprehensive investigations, thereby enabling the unveiling of genetic anomalies implicated in oncogenesis. Recent applications of SNP arrays encompass genome-wide screening, fostering enhanced correlations between genetic patterns and phenotypes, encompassing disease etiology. This underscores the pivotal role of SNP arrays in unraveling the intricate landscape

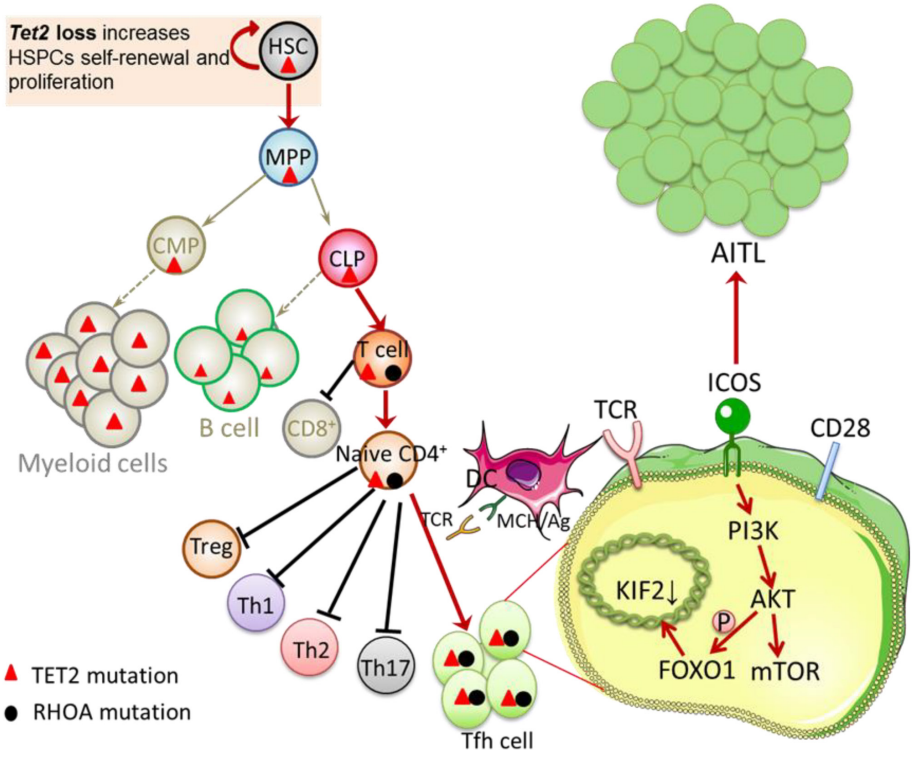
of cancer genetics and their potential contribution to early cancer diagnosis (Shahrabi *et al.*, 2020; Etebari *et al.*, 2015).

Genetics

The TET2 gene, situated at chromosome 4 position 4q24 with accession number NM_001127208.3 (NCBI, 2023), has emerged as a central topic of investigation in contemporary molecular biology. Responsible for encoding a protein with complex and elusive functionality, this gene has piqued the scientific community's interest due to its potential involvement in transcription regulation, the initial phase of protein synthesis. Its role is particularly prominent in hematopoietic stem cell differentiation, a process that encompasses the formation of vital blood cell types, such as erythrocytes (red blood cells), leukocytes (white blood cells), and thrombocytes (platelets) ("Tet2 Gene: Medlineplus Genetics.", 2014; Shastry, 2009). In addition to these functions, the TET2 protein has been identified as possessing tumor-suppressive properties, acting to inhibit abnormal cellular proliferation, thus aligning it with the characteristics of established tumor suppressor proteins.

Within this context, Figure 2 serves to provide a comprehensive visual depiction of the proposed mechanism of action of TET2, offering a tangible and illustrative insight into its functional role within cellular processes. This figure illustrates the impact of TET2 and RHOA mutations on the body, showing a complex interaction that leads to enhanced self-renewal and proliferation of hematopoietic stem/progenitor cells. This cooperative effect suppresses $CD8^+$ T cell differentiation while favoring $CD4^+$ T cells and T follicular helper (Tfh) cells, leading to aberrant activation and transformation of Tfh cells. ICOS plays a critical role through specific pathways, including the ICOS-PI3K-mTOR signaling pathway and AKT activation, contributing to Tfh lineage differentiation and the transformation characteristics of angioimmunoblastic T-cell lymphoma (AITL) (Phillip, 2022; Hu, *et al.*, 2022).

Figure 2. ***TET2 and RHOA Mutation Impact on T Cell Development and AITL.*** *TET2 loss and RHOA G17V mutations synergistically enhance HSPC self-renewal and proliferation. They also suppress CD8+ T cell differentiation, favoring CD4+ T cells and leading to Tfh cell expansion. ICOS plays a co-stimulatory role via the ICOS-PI3K-mTOR and ICOS-PI3K-AKT-FOXO1 pathways, essential for Tfh cell differentiation and AITL progression*



Gene's Implication in Lymphoma

The Tet2 gene is characterized by a prevalent mutational occurrence across a diverse array of hematopoietic malignancies, including both myeloid and lymphoid cancers, as well as various solid tumor types ("Tet2", 2017; Jiang, 2020). Within the realm of cancer cells and specific immune cell subsets such as T cells, B cells, and macrophages, Tet2 assumes an essential role in the epigenetic modification of DNA. This process is a vital aspect governing gene regulatory mechanisms and can influence the activity of genes within malignant cells and certain immune cell populations. Notably, the epigenetic modifications known as 5hmC and 5mC. 5mC is the methylation of the fifth position of the pyrimidine ring of cytosine. 5hmC is a product of 5mC demethylation by the Ten-Eleven Translocation family proteins. 5hmC was found to regulate many cellular and developmental processes, including the pluripotency of embryonic stem cells, neuron development, and tumorigenesis in mammals (Shi *et al*., 2017). 5hmC and 5mC are orchestrated by Tet2, and are consistently present under both physiological and pathological circumstances (Jiang, 2020).

In the context of hematological neoplasms, the impairment or abrogation of Tet2 function primarily leads to the abnormal production of 5mC, thus disrupting the epigenetic landscape. This perturbation is strongly associated with the progression of solid tumors, with frequent Tet2 mutations being a hallmark feature of hematopoietic malignancies (Cong, 2021; Jiang 2020) Experimental evidence from murine models has revealed that disruption of Tet2, or concurrent disruption of both Tet2 and Tet3, precipitates the spread of myeloid or lymphoid cell populations, culminating in the formation of aggressive tumor (Jiang, 2020). These insights emphasize the critical role of Tet2 in shaping the pathogenesis of both hematopoietic and solid cancers, thereby illuminating potential avenues for therapeutic interventions.

Through a synthesis of current research, this paper aims to provide a targeted examination of the TET2 gene, focusing specifically on the association of SNPs that may be found not only on TET2 but the entirety of chromosome 4 and their potential connection to lymphoma. Given the complexity and crucial nature of this gene, the present study underscores the need for an extensive analysis to decipher the underlying relationships between the TET2 gene's variations and the onset or progression of lymphoma. This research may pave the way for novel therapeutic interventions and contribute valuable insights into personalized medicine strategies.

Methods

In the current study, we undertook an analysis of sequence reads obtained from lymphoma patients and control subjects to identify genomic variants and meticulously examine their potential implications. The reference genome for *Homo sapiens* chromosome 4, accessed from Ensembl (Release 104) (Cunningham *et al*., 2021), was utilized as a benchmark for accurate alignment and interpretation of genomic alterations detected in our study. Our analytical pipeline incorporated the selection of Sequence Read Archive (SRA) sequences based on library construction strategy and experimental design. SRA files were obtained using the fastq-dump utility of the SRA Toolkit v2.10.7 (NCBI, 2021), with subsequent analyses performed via the command-line interface. An initial assessment of sequence read quality was conducted using FastQC (v0.11.9) (Andrews, 2010), and sequence trimming was performed using Trimmomatic (v0.39) (Bolger *et al*., 2014), focusing on regions with compromised base quality (Phred score < 33). This process improved the overall sequence data quality by removing low-quality bases.

Sequence alignment to the reference human genome and indexing of the reads were achieved using Bowtie2 (v2.4.2) (Langmead & Salzberg, 2012). The resultant Sequence Alignment/Map (SAM) file was converted to a Binary Alignment/Map (BAM) file using SAMtools (v1.11) (Li *et al*., 2009). Coverage depth at each genomic position was determined through the SAMtools 'depth' command, identifying regions likely to contain genetic variants. SNPs were identified using the binary variant call format (BCF) utilities from the SAMtools suite. The generated BCF file underwent stringent filtering via BCFTools (v1.11), retaining only high-confidence variants. These SNPs were further

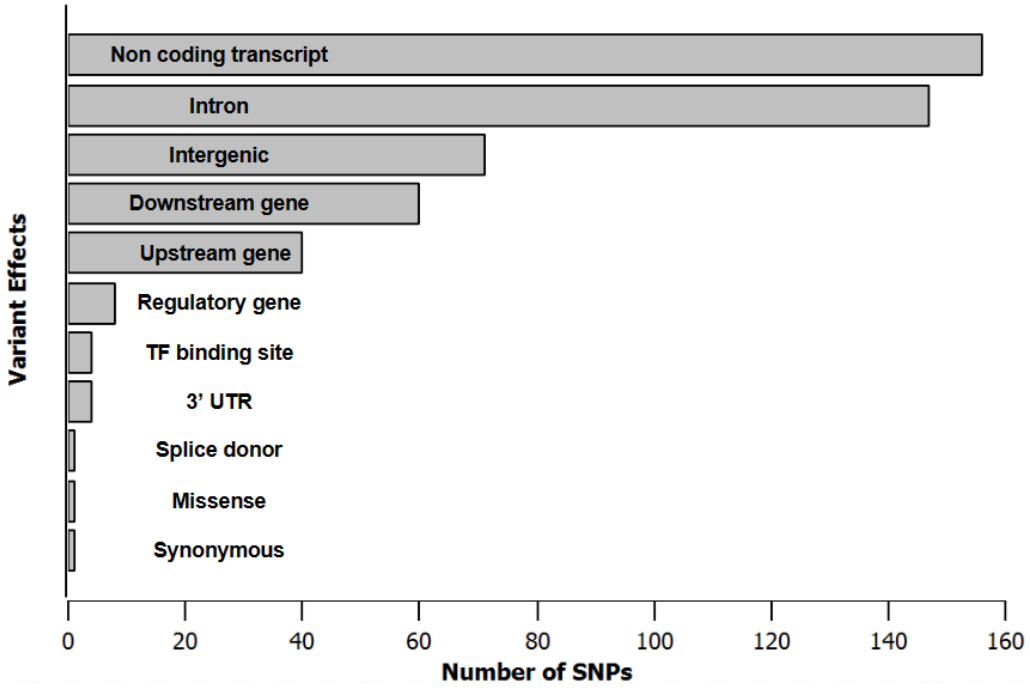
validated by comparing them to a reference dataset of SNPs from a healthy population using the Pandas library in Python. This approach allowed us to focus on lymphoma-specific genomic alterations.

Statistical analyses assessed the association between identified variants in the lymphoma and control cohorts, employing Chi-square tests for categorical variables evaluation. Contingency tables, constructed from the CSV file, exhibited subject counts with or without specific variants across both cohorts. The Python SciPy library was used to calculate the chi-square statistic and corresponding p-value, revealing a significant association ($p < 0.05$) between the identified variants and the lymphoma cohort. We further linked SNP accession numbers to their genomic consequences using a Python script we developed. SNP accession numbers were extracted from the SAMtools-generated BCF file, and the Ensembl REST API provided pertinent information regarding each SNP, including genomic location and allelic configuration. The parsed JSON responses facilitated data extraction, subsequently merged with genomic consequences and stored in a CSV file for further analysis. Our Python script allowed efficient information extraction and reduced manual error risk. All scripts used in this study are publicly available on our GitHub repository (https://github.com/crisprmax/SNP-identifier-Python), including comprehensive documentation and guidelines.

Results

In our endeavor to unravel the genetic heterogeneity and discern the functional roles of SNPs within the genome of a lymphoma clinical cohort, Figure 3 (referenced below) stands as a critical juncture in our investigative journey. This figure offers an overview, capturing both the distribution and the functional categories of the SNPs identified in our study. We analyzed a total of 1,072 SNPs, categorizing them based on their presumed functional roles. Intriguingly, a significant proportion of these variants have yet to be functionally characterized, posing an exciting avenue for future research. To maintain the integrity of the graphical representation, these uncharacterized variants were intentionally omitted, thereby enabling a more discernible visualization of the less abundant, but potentially functionally significant, variant categories.

Figure 3. ***Different Variant Effects on Lymphoma Clinical Cohort.*** *This figure illustrates the distribution and functional categorization of SNPs found within a clinical cohort of lymphoma patients. A total of 1072 SNPs were analyzed, with the majority remaining uncharacterized*



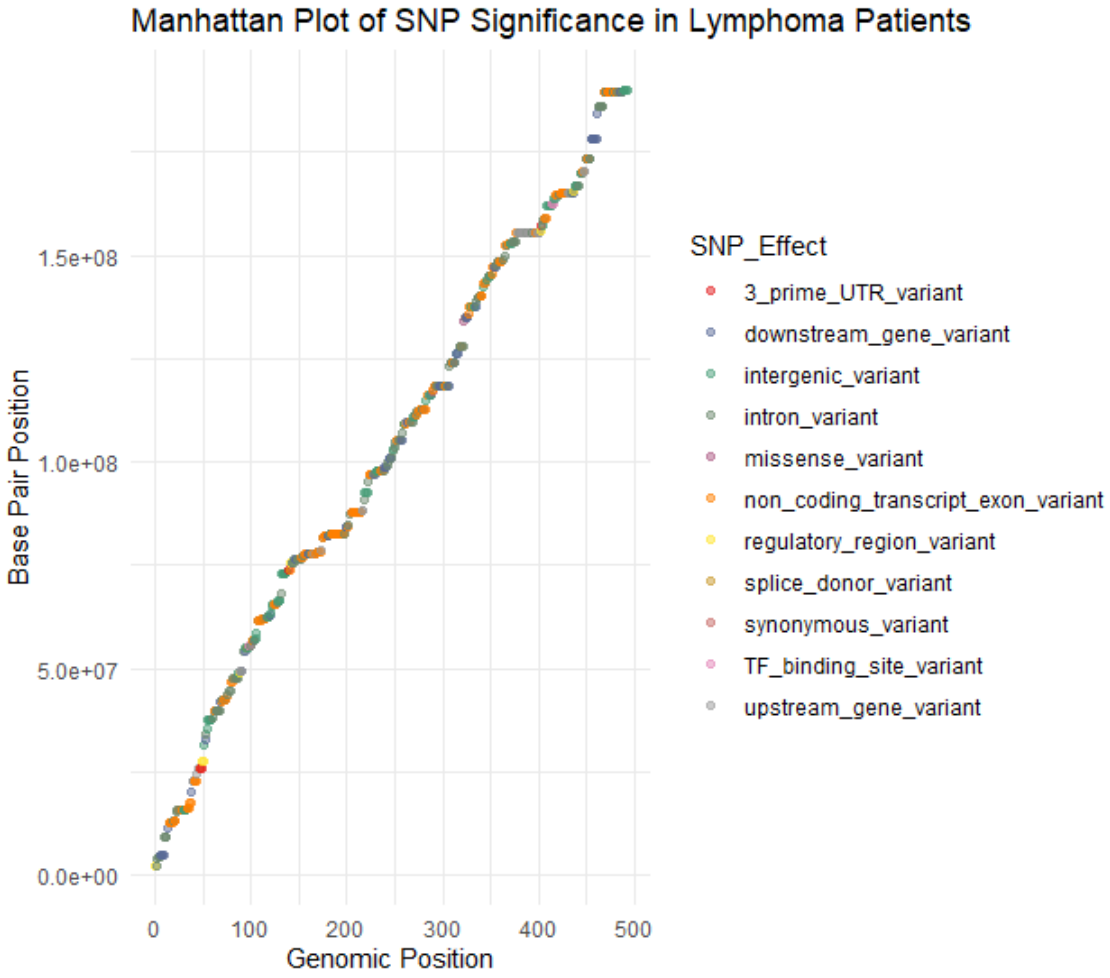
Within the dataset of characterized SNPs, multiple functional categories were represented, highlighting the diverse genetic landscape of the lymphoma clinical cohort (Table 1). Specifically, 156 SNPs were classified as non-coding transcript exon variants, indicating variations in regions that are transcribed but do not encode proteins. A further 147 SNPs were identified as intron variants, which could potentially influence splicing mechanisms. Additionally, 71 SNPs were categorized as intergenic variants, implicating regions between genes. The dataset also included 60 downstream and 40 upstream gene variants, suggesting a potential impact on the regulation of adjacent genes. Notably, 8 SNPs were found in regulatory regions and an additional 4 in transcription factor binding sites, emphasizing their likely role in gene regulation. Four SNPs were located within the 3' untranslated region (UTR), pointing to potential post-transcriptional regulatory effects. Lastly, the analysis unveiled a small number of functionally impactful SNPs: a single splice donor variant with potential implications for splicing processes, one missense variant that could alter protein function, and a lone synonymous variant with minimal impact on protein structure or function. These comprehensive insights underscore the diversity of SNPs present within the genome of the unknown organism, emphasizing potential functional repercussions that necessitate further exploration.

Table 1. ***Categorization and Descriptions of SNP Effects in Lymphoma Clinical Cohort****. This table enumerates the various types of SNPs identified within a clinical cohort of lymphoma patients, alongside brief descriptions outlining their potential functional implications*

SNP Effects	Description
Non-coding Transcript Exon Variants	Variations in regions that are transcribed but do not encode proteins.
Intron Variants	Potential influence on splicing mechanisms.
Intergenic Variants	Implicating regions between genes.
Downstream Gene Variants	Potential impact on the regulation of adjacent genes.
Upstream Gene Variants	Potential impact on the regulation of adjacent genes.
Regulatory Region Variants	Likely role in gene regulation.
Transcription Factor Binding Site Variants	Likely role in gene regulation.
3' Untranslated Region (UTR) Variants	Potential post-transcriptional regulatory effects.
Splice Donor Variant	Potential implications for splicing processes.
Missense Variants	Could alter protein function.
Synonymous Variant	Minimal impact on protein structure or function.

To better elucidate the complexities of these data, we employed a Manhattan Plot (Figure 4). Our aim was to provide an overview of both the spatial arrangement and presumptive functional significance of SNPs in our clinical cohort. The x-axis of the plot arranges the SNPs according to their respective genomic coordinates on chromosome 4, offering a seamless view of their genomic distribution. Conversely, the y-axis demarcates the precise base pair positions for each individual SNP within this chromosome, thus enriching our comprehension of their genomic context. Each data point on the plot represents a unique SNP, with its y-axis placement signifying its specific base pair location on chromosome 4. To further enhance interpretability, we incorporated a color-coding scheme based on a scale that quantifies SNP Effect values. This chromatic stratification serves to spotlight SNPs that warrant particular attention, especially those with potential functional or regulatory roles.

Figure 4. ***Manhattan Plot of SNP Significance in Lymphoma Patients.*** *This plot provides an intricate overview of the spatial distribution and potential functional implications of SNPs specific to lymphoma patients. Each point populating the plot represents an individual SNP, with its vertical placement serving as a proxy for its base pair position on chromosome 4. When applicable, the coloration of each point is governed by a color scale that corresponds to its SNP Effect value*



Our study provides an analytical framework for the investigation of SNPs within a lymphoma clinical cohort. By using advanced visualization techniques, such as the Manhattan Plot, we have been able to delineate the spatial distribution and infer potential functional roles of these genetic variants. The color-coding system employed significantly aids in identifying SNPs with potential functional or regulatory implications, serving as a roadmap for targeted genomic studies in the future. Our findings not only enhance the current understanding of the genetic landscape in lymphoma but also pave the way for more in-depth explorations that could have clinical relevance.

Discussion:

The TET2 gene has been intricately linked to epigenetic modifications, cellular differentiation, and tumor suppression, signifying its critical role in hematopoiesis and cancer development. While we did not discover SNPs directly within TET2, the genetic variants that were identified in this study still offer insights into the broader genetic landscape associated with lymphoma patients. These variants can potentially serve as markers for early diagnosis and contribute to the development of personalized medicine strategies, which could have a significant impact on patient outcomes.

The prevalence of lymphoma, particularly NHL, stands as a significant contributor

to cancer-related morbidity and mortality, underscoring the urgent need for continued research efforts aimed at improving diagnostic tools, treatment modalities, and patient outcomes. The staging process of lymphoma, illustrated in Figure 1, further highlights the progressive nature of the disease and emphasizes the critical importance of early detection and intervention. Complementing this, the role of immunotherapy and targeted therapies in lymphoma treatment is elucidated. The potential of monoclonal antibodies and innovative immunotherapeutic agents is particularly noteworthy. These advancements in treatment strategies not only hold promise for enhancing patient responses but also for reducing adverse effects, thereby reinforcing the necessity of further research and clinical trials.

The systematic investigation of SNPs in our study offers a nuanced understanding of the genetic heterogeneity underlying lymphoma. Figure 3 serves as a cornerstone, providing an exhaustive overview of the potential functional implications of these genetic variants. The notable prevalence of «unknown» variants amplifies the imperative for continued research to elucidate their specific roles and consequences in the pathogenesis of lymphoma. Complementing this, the methodological rigor of our study stands as a testament to the robustness of our SNP analysis. Leveraging advanced bioinformatics tools and publicly available datasets, we have employed meticulous alignment, sequence trimming, and variant calling processes. This ensures not only the accuracy of our findings but also the reliability of the SNPs identified, thereby laying a solid foundation for future research in this critical area of lymphoma genetics.

Conclusion:

Our study aimed to identify and categorize SNPs in lymphoma patients. Utilizing a robust and comprehensive analytical pipeline that integrated multiple bioinformatics tools and statistical methods, we were able to identify several lymphoma-specific SNPs of particular interest. A notable finding is the identification of 1072 SNPs, the majority of which remain uncharacterized. These «unknown» SNPs present an intriguing area for future research, particularly in understanding their functional roles and potential involvement in lymphoma. Our study also revealed a substantial number of categorized SNPs, including 156 non-coding transcript exon variants, 147 intron variants, and 71 intergenic variants. These variants have potential implications for gene splicing, regulation, and even intergenic interactions.

Importantly, we also identified variants that are likely to have more immediate functional implications. These include 60 downstream and 40 upstream gene variants that may influence the regulation of neighboring genes, 8 regulatory region variants, and 4 transcription factor binding site variants that suggest potential regulatory roles. This investigation serves as an intricate tapestry of the SNP landscape in lymphoma, woven together through a rigorous analytical framework. Our findings not only offer a nuanced insight into the intricate genetic landscape associated with lymphoma but also serve as a foundational platform for future research endeavors focused on the functional interpretation and potential therapeutic applications of these genetic variants in the clinical management of the disease.

References:

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Retrieved from URL: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*,– 30(15).– P. 2114–2120. URL: https://doi.org/10.1093/bioinformatics/btu170

Cong, B., Zhang, Q., & Cao, X. (2021). The function and regulation of TET2 in innate immunity and inflammation. *Protein & cell*,– 12(3).– P. 165–173. URL: https://doi.org/10.1007/s13238-020-00796-6

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., & Flicek, P. (2021). Ensembl 2021. *Nucleic Acids Research*,– 49(D1).– D884-D891. URL: https://doi.org/10.1093/nar/gkaa942

Degtyareva, A. O., Antontseva, E. V., & Merkulova, T. I. (2021). Regulatory SNPs: Altered Transcription Factor Binding Sites Implicated in Complex Traits and Diseases. *International journal of molecular sciences*,– 22(12).– 6454 p. URL: https://doi.org/10.3390/ijms22126454

Etebari, M., Navari, M., & Piccaluga, P. P. (2015). SNPs Array Karyotyping in Non-Hodgkin Lymphoma. *Microarrays (Basel, Switzerland)*,– 4(4).– P. 551–569. URL: https://doi.org/10.3390/microarrays4040551

Halim, L., & Maher, J. (2020). CAR T-cell immunotherapy of B-cell malignancy: the story so far. *Therapeutic advances in vaccines and immunotherapy*,– 8, 2515135520927164. URL: https://doi.org/10.1177/2515135520927164

Hu, L., Zhang, X., Li, H., Lin, S., & Zang, S. (2022, November 20). *Targeting TET2 as a therapeutic approach for angioimmunoblastic T cell lymphoma*. MDPI. URL: https://www.mdpi.com/2072-6694/14/22/5699

Jamil A., Mukkamalla S.K.R. Lymphoma. [Updated 2023 July 17]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: URL: https://www.ncbi.nlm.nih.gov/books/NBK560826/

Jiang S. (2020). Tet2 at the interface between cancer and immunity. *Communications biology*,– 3(1).– 667 p. URL: https://doi.org/10.1038/s42003–020–01391–5

Jiraskova K., Hughes D.J., Brezina S., Gumpenberger T., Veskrnova V., Buchler T., Schneiderova M., Levy M., Liska V., Vodenkova S., et al. Functional Polymorphisms in DNA Repair Genes Are Associated with Sporadic Colorectal Cancer Susceptibility and Clinical Outcome. *International Journal of Molecular Sciences*. 2019; 20(1):97. URL: https://doi.org/10.3390/ijms20010097

Key statistics for Non-Hodgkin Lymphoma. American Cancer Society. (2023). URL: http://www.cancer.org/cancer/types/non-hodgkin-lymphoma/about/key-statistics.html#:~:text=The%20American%20Cancer%20Society%27s%20estimates,11%2C780%20males%20and%208%2C400%20females).

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*,– 9(4).– P. 357–359. URL: https://doi.org/10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*,– 25(16). 2078–2079. URL: https://doi.org/10.1093/bioinformatics/btp352

Lymphoma-patient version. National Cancer Institute. (2023). URL: http://www.cancer.gov/types/lymphoma

National Center for Biotechnology Information (NCBI). (2021). SRA Toolkit v2.10.7. Retrieved from URL: https://github.com/ncbi/sra-tools

NHL staging. https://www.lls.org/lymphoma/non-hodgkin-lymphoma/diagnosis/nhl-staging. (2023). URL: https://www.lls.org/lymphoma/non-hodgkin-lymphoma/diagnosis/nhl-staging

Non-hodgkin lymphoma stages: Staging Non-Hodgkin Lymphoma. Staging Non-Hodgkin Lymphoma | American Cancer Society. (n.d.). URL: *https://www.cancer.org/cancer/types/non-hodgkin-lymphoma/detection-diagnosis-staging/staging.html*

Philip, M., Schietinger, A. *CD8+ T cell differentiation and dysfunction in cancer. Nat Rev Immunol* – 22. – P. 209–223 (2022). URL: *https://doi.org/10.1038/s41577-021-00574-3*

R-CHOP: *Combination chemotherapy cancer treatment*. City of Hope. (2022, June 14). URL: https://www.cancercenter.com/treatment-options/r-chop#:~:text=Doctors%20use%20R%2DCHOP%20to, other%20cancer%20drugs%20and%20treatments

Shahrabi, S., Ghanavat, M., Behzad, M. M., Purrahman, D., & Saki, N. (2020). *CD markers polymorphisms as prognostic biomarkers in hematological malignancies. Oncology reviews,*– 14(2).– 466 p. URL: *https://doi.org/10.4081/oncol.2020.466*

Shastry B.S. (2009). *SNPs: impact on gene function and phenotype. Methods in molecular biology (Clifton, N.J.),*– 578.– P. 3–22. URL: *https://doi.org/10.1007/978-1-60327-411-1_1*

Shi, D.-Q., Ali, I., Tang, J., & Yang, W.-C. (2017, July 5). *New insights into 5hmc DNA modification: Generation, distribution and function*. Frontiers. URL: https://www.frontiersin.org/articles/10.3389/fgene.2017.00100/full

Single nucleotide polymorphisms (snps). Genome.gov. (n.d.). URL: *https://www.genome.gov/genetics-glossary/Single-Nucleotide-Polymorphisms*

Stages of lymphoma. moffitt. (2023). URL: http://www.moffitt.org/cancers/lymphomas-hodgkin-and-non-hodgkin/diagnosis/stages

Tet2. Tet2 – My Cancer Genome. (2017). URL: https://www.mycancergenome.org/content/gene/tet2

Treating Non-Hodgkin Lymphoma: Non-Hodgkin Lymphoma Treatments. Non-Hodgkin Lymphoma Treatments | American Cancer Society. (2023). URL: https://www.cancer.org/cancer/types/non-hodgkin-lymphoma/treating.html

U. S. National Library of Medicine. (2023). Tet2 gene: Medlineplus Genetics. MedlinePlus. https://medlineplus.gov/genetics/gene/tet2/

U. S. National Library of Medicine. (2023). *Tet2 Tet methylcytosine dioxygenase 2 [homo sapiens (human)] – gene – NCBI*. National Center for Biotechnology Information. URL: https://www.ncbi.nlm.nih.gov/gene/54790

submitted 15.05.2024;
accepted for publication 29.05.2024;
published 22.07.2024

Contact: liliww2@gmail.com

The European Journal of Biomedical and Life Sciences 2024, No 1–2
ISSN 2310-5674

PREMIER Publishing
ppublishing.org

DOI:10.29013/EJBLS-24-1.2-36-45

FINE-TUNING CHROMOPROTEIN EXPRESSION FOR BIOPRODUCTION IN HALOMONAS SPP

Ruiyan Huang [1]

[1] 300 Garrison Forest Rd, Owings Mills, MD21117

Cite: *Ruiyan Huang. (2024). Fine-Tuning Chromoprotein Expression for Bioproduction in Halomonas spp. The European Journal of Biomedical and Life Sciences 2024, No 1–2.* https://doi.org/10.29013/EJBLS-24-1.2-36-45

Abstract

As bioproduction, especially microbial production, of substances became spotlighted, attention has been raised on developing tools to combat the unpredictable cellular synthetic activities that are often posed by the microbial cell factories. Fine-tuning tools, as one way to ameliorate cell factories for production, have become popular and effective for improving biosynthesis processes by achieving relatively stable regulatory purposes. In this study, a tunable expression system was created and utilized to control the level of expression of five single genes in the non-model bacteria *Halomonas bluephagenesis*, a member of *Halomonas spp*. and a sustainable and low-cost bioengineering and biomanufacturing platform for "next-generation industrial biotechnology" (NGIB). To better visualize results of the inducible flux regulation of this multiple inducible expression system which is designed to control five genes simultaneously, chromoprotein single genes were expressed and regulated through various combinations of induction concentration. The results demonstrated the system's ability to regulate expression of multiple individual genes at the same time. The intensely visible chromoprotein production of varied colors and the colorful bacteria paintings from the visual results also proved the viability of this system in *Halomonas bluephagenesis*, the utilization of which contributes to a more cost-effective and waste-reducing solution to resource waste and cost issues of current industrial biotechnology. Overall, this study reveals the feasibility of the desirable tuning system for efficacious expression regulation of target genes and allows for further exploration and metabolic flux optimization in the bioproduction of other substances to allow more stable bioproduction processes in synthetic biology.

Keywords: *Chromoproteins, Halomonas bluephagenesis, Next generation industrial biotechnology, Metabolic flux engineering, Synthetic Biology*

Introduction

Many elements of our nature present fascinating and multitudinous colors, and marine coral reefs being a significant one of them in adding hues to the oceans. Coral reefs derive their colors from both fluorescent proteins (FPs) and chromoproteins (CPs) which are homologous to the green fluorescent protein (GFP) (1). Thanks to their macroscopic luminescence properties,

chromoproteins absorb visible ambient light and give off strong colors visible to the naked eye (2). Therefore, CPs play crucial parts as biomarkers in finding food contaminants, landmines and biowarfare agents, and biosensors were adapted. They are also popular being used as dye or paint replacements for artistic creations in BioArt, While FPs are broadly utilized in *in vivo* bioimaging of molecules, live cells, organelles, and manufactured compounds, CPs, as a subset of the fluorescent protein family, are also used for bioimaging purposes and have advantages over FPs such as its ability to be detected in an instrument-free way (3–4). Whereas the detection of FPs usually require fluorometers, flow cytometers, or ultraviolet light (UV) lamps which can result in issues such as background fluorescence, photobleaching, and UV damage of the sample (2). The utilization of CPs leaves out these possible concerns that FP imaging can raise and the necessity of expensive lab equipment, making CPs attractive for a wide range of applications such as visual detection of gene expression without equipment in synthetic biology development and industrialized biomanufacturing where costs, production resources and required experimental proficiency of labor are major considerations (2, 5). Therefore, CPs are selected as visual identifiers of bioproduction in this study.

In modern days, most CPs require expression in bacteria such as *Escherichia coli* (2). While such model bacteria have been extensively used for recombinant protein production (6), relatively fewer efforts have been made to explore the tuning of multiple inducible genes in other engineered and improved non-model bacteria like *Halomonas bluephagenesis*. Although various attempts have been made to explore intense production of novel CPs and general bioengineering tools in model bacteria like *E. coli* (Bao et al., 2020; Meyer et al., 2019; Pang et al., 2020), bioprocesses in these bacteria often necessitates the use of stainless steel bioreactors, strict sterilization, and costly separation procedures and practiced engineers who are able to perform the processes under sterile conditions, often diminishing competitiveness of resultant industrial biotechnology products (Chen & Jiang, 2018). On the other hand, *Halomonas bluephagenesis*, an extremophilic non-model halophile possessing tolerations for high salt and pH condition, has been continuously engineered and improved in the last few decades as a chassis (Ye & Chen, 2021). Its strong growth rate under high salt and pH levels enables efficacious biosynthesis and fermentation processes under non-sterile conditions (Tan et al., 2011), particularly characterized in the Next Generation Industrial Biotechnology (NGIB), greatly reducing the difficulties led by microbial contamination, the consumption and possible waste of energy and freshwater resources for sterilization, slow growth of production organisms, and the expensive separation, which are existing issues of current industrial biotechnology (CIB) that does not make it more competitive in contrast to the chemical engineering industry (Chen & Jiang, 2018). Thus the use of *Halomonas* as a biological chassis can raise the cost-effectiveness of production and contribute to the sustainable and low-consumption development of synthetic biology research in the long run (Yu et al., 2019). Therefore, it is crucial to create and establish tools for efficacious control and tuning of biosynthesis in this previously scarcely explored non-model bacteria, *Halomonas bluephagenesis*.

Research objective

This study aimed to explore the feasibility of inducible production regulation systems in this non-model chassis bacteria *H. bluephagenesis* by constructing and testing an induction system containing multiple inducible operons that enables the simultaneous regulations of the expression of five independent chromoprotein genes, and using the direct visualization of the color and intensity resulted from the targeted chromoprotein genes' expressions as indication of its level of throughput controlled by the system. The hypothesis is that by changing the concentration of inducers added to induce each of the five genes encoding for CPs, the color intensity of the produced will vary, suggesting the successful fine-tuning of the gene expressions of multiple genes in *H. bluephagenesis*.

Materials and methods

Plasmid Design

For the purpose of this study, the plasmid 'CP' was constructed by Gibson Assembly. The plasmid holds inducible operons with promoters, terminators, the five corresponding open-source chromoprotein genes, *spisPink*, *fwYellow*, *amajLime*, *amilCP*, and *gfasPurple*, as target sequences expressing distinct colors of pink, yellow, green, blue, and purple under the induction with corresponding inducers IPTG, OHC14, OC6, arabinose, and vanillic acid, respectively.

Figure 1. *Schematic plasmid design illustration of the inducible chromoprotein expression flux regulation pathways. Characterization of constitutive promoters, insulators, ribosomal binding sites (RBS), and genes of interest (GOI), and terminator, presented in order from left to right, for the five independently regulated pathways, each labeled in a different color*

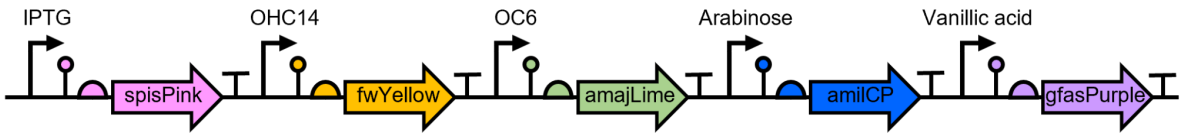
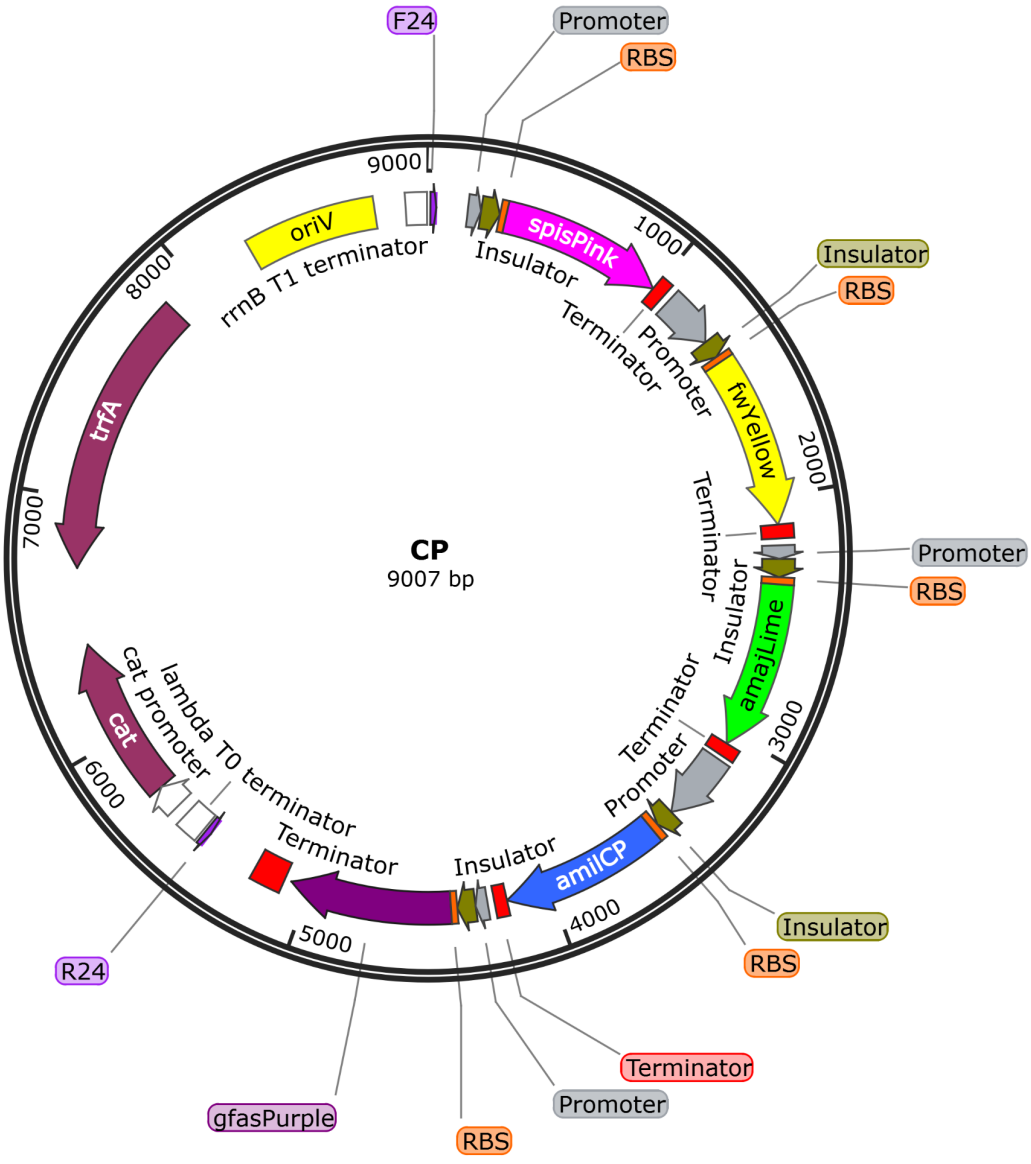


Figure 2. *Vector design illustration of the plasmid 'CP'*



Plasmid Transformation and Cell Proliferation

The plasmid 'CP' was transformed into the chemically competent *E. coli* cells of strain s17–1. The *E. coli* s17–1 cells holding the CP plasmid and the chassis *H. bluephagenesis* strain TDR2 cells were incubated at 37 °C and 200 rpm in a 20mL LB medium of 10g/L NaCl concentration and in a 20mL LB medium of 60g/L NaCl added with 20μL chloramphenicol antibiotic, respectively. The conjugated *E. coli* cells with the designed plasmid were grown in a 20mL LB liquid medium at 37 °C with the addition of 10g/L NaCl and 20μL liquid chloramphenicol antibiotic. The halomonas strain was grown in a 20mL LB medium with 60g/L NaCl at 37 °C for 12 hours with no additional antibiotic.

Conjugation of the CP Plasmid into H. bluephagenesis TD cells

The two bacteria cultures of *H. bluephagenesis* TDR2 and *E. coli* s17–1 harboring plasmid CP, respectively, are then centrifuged at 5000 rpm and the liquid supernatant was discarded. The precipitations from both tubes were mixed and added into 50 μL of LB liquid medium of 10g/L NaCl concentration. The mixture is then dropped on a petri dish and grown at 37 °C for 8–12 hours to allow conjugation of CP plasmid into the TDR2 cells.

Afterwards, recombinant *H. bluephagenesis* TDCP cells harboring the CP plasmid were scrubbed up and evenly spread on three separate agar plates. In order to purify the remaining bacteria mixture and to sift out the *E. coli* cells and halomonas cells that do not contain the designed plasmid, the bacteria culture that we suppose contains the designed plasmid were evenly spread on the three petri dishes with agar-solidified LB medium with 60g/L NaCl and chloramphenicol addition. The bacteria on these agar plates were grown at 37 °C for 48 hours.

Monoclonal purification and Conjugation Verification

The monoclonals grown on agar plate were then picked up using sterile wooden toothpicks and purified on agar plates with chloramphenicol and of 60g/L NaCl concentration using the streaking method and incubated at 37 °C. Each clone was numbered on the petri dish to make sure that we use the clones of correct DNA length for verification and induction later performed in this study.

In order to verify that the monoclonal cells were successfully conjugated, PCR was performed to amplify the recombinant TDCP cells and a gel electrophoresis was done to visualize the size of the target conjugated TD DNA. The clone bands showing lengths of about 6000 bp, the desired DNA length of the conjugated TD cells, indicate the presence of successful conjugation of CP plasmid with TDR2 cells.

Pipette tips were used to transfer such clones from the agar plates into the wells of a 96-wells deep-well plate. In each well, the pipette tip attaching the TDCP cells were put into 1mL LB medium of 60g/L NaCl concentration with chloramphenicol antibiotic added.

Chromoprotein (CP) Induction in TD cells for novel colors and microbial paintings

The recombinant *H. bluephagenesis* TDCP cells were induced to produce chromoproteins in three 96-wells deep-well plates. On each plate, 32 wells were designed to hold different multiple protein combinations (**Table 1**) and the other 15 wells show single protein induction for the five proteins, each at induced three different inducer concentrations (**Table 2**). The rest of the wells were left empty.

Table 1. *32 combinations generated from the following inducer concentrations: 10, 0 mg/L IPTG, 10^{-6}, 0M OHC14, 10^{-6}, 0 M OC6, 10^{-3}, 0M Ara (Arabinose), 10^{-5}, 0 M Van (Vanillic acid). Each cell in the bolded perimeters of this table corresponds to one of the 32 wells where combined expression of colors occur*

	IPTG = 10 mg /L			
	OHC14 10–6 M		**OHC14 0M**	
	1	**2**	**3**	**4**
A	OC6 10–6M Ara 10–3M	OC6 10–6M Ara 0M	OC6 10–6M Ara 10–3M	OC6 10–6M Ara 0M

	IPTG = 10 mg /L			
	OHC14 10–6 M		**OHC14 0M**	
	1	**2**	**3**	**4**
	Van 10–4M	Van 10–4M	Van 10–4M	Van 10–4M
	OC6 10–6M	OC6 10–6M	OC6 10–6M	OC6 10–6M
B	Ara 10–3M	Ara 0M	Ara 10–3M	Ara 0M
	Van 0M	Van 0M	Van 0M	Van 0M
	OC6 0M	OC6 0M	OC6 0M	OC6 0M
C	Ara 10–3M	Ara 0M	Ara 10–3M	Ara 0M
	Van 10–4M	Van 10–4M	Van 10–4M	Van 10–4M
	OC6 0M	OC6 0M	OC6 0M	OC6 0M
D	Ara 10–3M	Ara 0M	Ara 10–3M	Ara 0M
	Van 0M	Van 0M	Van 0M	Van 0M
	IPTG=0mg /L			
	OHC14 10–6M		**OHC14 0M**	
	1	**2**	**3**	**4**
	OC6 10–6M	OC6 10–6M	OC6 10–6M	OC6 10–6M
E	Ara 10–3M	Ara 0M	Ara 10–3M	Ara 0M
	Van 10–4M	Van 10–4M	Van 10–4M	Van 10–4M
	OC6 10–6M	OC6 10–6M	OC6 10–6M	OC6 10–6M
F	Ara 10–3M	Ara 0M	Ara 10–3M	Ara 0M
	Van 0M	Van 0M	Van 0M	Van 0M
	OC6 0M	OC6 0M	OC6 0M	OC6 0M
G	Ara 10–3M	Ara 0M	Ara 10–3M	Ara 0M
	Van 10–4M	Van 10–4M	Van 10–4M	Van 10–4M
	OC6 0M	OC6 0M	OC6 0M	OC6 0M
H	Ara 10–3M	Ara 0M	Ara 10–3M	Ara 0M
	Van 0M	Van 0M	Van 0M	Van 0M

Table 2. *Three different inducer concentrations each added to induce each of the five individual CPs. Each shaded cell in the table corresponds to each of the 15 wells designed for individual chromoprotein expression*

Chromoprotein name	**Inducer**	**Concentration (inducer/ddH_2O)**		
spisPink	IPTG	10 mg/L	5 mg/L	0 mg/L
fwYellow	OHC14	10^{-6}M	10^{-8}M	0M
amajLime	OC6	10^{-6}M	10^{-8}M	0M
amilCP	Arabinose	10^{-3}M	10^{-5}M	0M
gfasPurple	Vanillic acid	10^{-4}M	10^{-5}M	0M

In each well, the induction mix was prepared by adding 1mL LB medium of 60 g/L NaCl concentration, 5µL liquid medium containing recombinant TDCP cells, each inducer solution of different concentrations, and chloramphenicol antibiotic to see the effects of the simultaneous induction of multiple colors that result in 32 mixed colors in total (**Figure 4**).

The deep-well plates were put into a micro-perforated plate thermostated oscillator and incubated at 37 °C and 1000rpm. After 48 hours, the plates were centrifuged for 15 minutes at 4700 rpm. The liquid supernatants were discarded and TDCP cells presenting visible colors as a result of chromoprotein expression were precipitated at the bottom of each well. Pictures were taken from the bottom-view of the deep-well plates to gain a vi-

sual of the level of production of chromoproteins centrifuged at the bottom of wells.

Several colors were chosen for microbial paintings imitating the starry sky, flowers, and Chinese landscape painting on agar plates. Because the area where colors should cover vary in sizes, the appropriate amount of inducer needed to bring out these colors at desired intensities were calculated by approximating the volume of solidified LB medium with 60g/L NaCl carried in sections of the agar plate and the corresponding volume of inducer. The calculated amounts of each inducer needed were spread on plain agar plates first, then TDCP bacteria mediums were evenly spread in designated areas after the inducer solutions were fully absorbed. The cells were grown at 37 °C.

Cell lysis and Protein Concentration Measurement

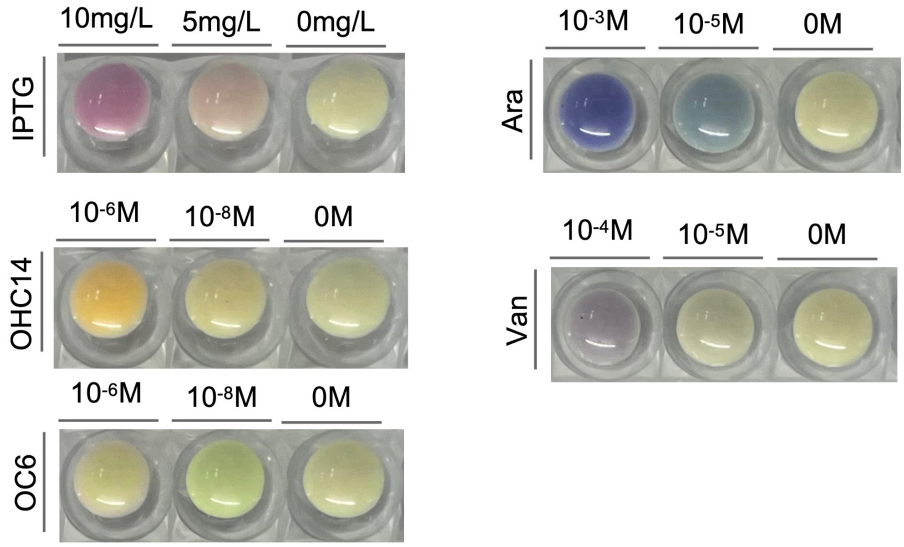
To further investigate the effects of the constructed induction system, the cells in the 15 wells of each deep-well plate that contains single chromoprotein induction were lysed and the proteins were extracted to measure the group protein concentration generated. The lysis solution includes 500μL 0.1M EDTA, 500μL 100X protease inhibitor cocktail for bacterial cell extracts, 1mL 50mg/mL lysozyme, and 50mL Beyo Lytic™ Bacterial Active Protein Extraction Reagent. 1mL of the lysis solution was added to the TDCP cells in wells where only the genes of one chromoprotein were induced in order to see the effect of different inducer concentrations on the color intensity produced. The lysate solution and cell mixture were centrifuged at 25 °C and 4700 rpm for 20 minutes after being incubated at 37 °C and 1000 rpm for 10 minutes. 2μL of the protein precipitate from each well selected was measured using the Scopes method in the Nanodrop A280 program with baseline correction of 340 nm. The protein concentrations were recorded and analyzed by calculating the percentage differences in group protein concentrations between the samples in which different levels of inducer were added.

Results and discussion

Overall, the direct visualization of the visibly and intensely colored recombinant *H. bluephagenesis* TDCP cells suggests that the constructed tunable expression system was successful in regulating the expression of multiple genes in this non-model chassis bacteria.

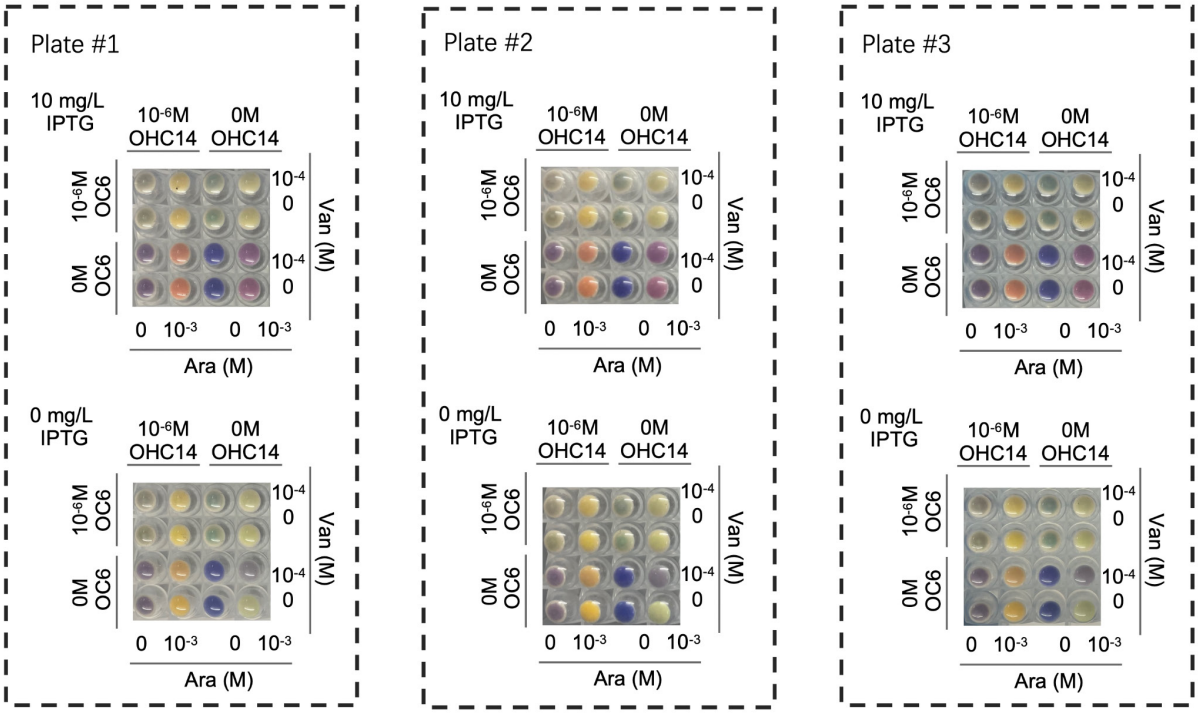
This induction tuning system constructed for multiple inducible expressions in this non-model bacteria produced vibrantly-colored bacteria after induction, resulting in the successful expression regulation of five chromoproteins in combinations and individually, as shown in the figures below.

Figure 3. *Matrix of recombinant H. bluephagenesis TDCP in which each of the five chromoproteins were induced individually and induced by three distinct concentrations of inducer*



While the wells with 0M inducer concentration acted as negative controls of no chromoprotein production, it is observed that the product cells present a general increase in color intensities along with the increase in concentration of the separately added inducers (**Figure 3**). This visualization directly suggests that the production of chromoproteins is regulated by the induction system as hypothesized, demonstrating the effectiveness of the systems in controlling the relative degree to which each of the genes are independently expressed. It is also observed, however, that the results of the OC6-induced expression stands as an exception to this trend. It is worth noticing that the observed color intensity and the measured group protein concentration of the *amajLime* gene induced by OC6 both show a lower chromoprotein production under a higher-concentration (10^{-8} M) inducer induction than that under lower-concentration (10^{-8} M) inducer induction (**Figure 3**). One possible explanation for this low production of *amajLime* chromoprotein that resulted from a relatively high-level induction with OC6 inducer is metabolic burden, a longstanding issue in engineered recombinant cells in biotechnology.

Figure 4. *Matrix of tuned recombinant H. bluephagenesis TDCP with expressed chromoprotein induced by 10, 0 mg/L IPTG, 10^{-6}, 0M OHC14, 10^{-6}, 0 M OC6, 10^{-3}, 0M Ara (Arabinose), 10^{-5}, 0 M Van (Vanillic acid), resulting in 32 colors of combination on each of the three deep-well plates*



In the visual results of inductions in the deep-well plates and agar plates (Figure 4), it is also shown that the mixed induction of the CPs in each well produced a well-blended mix of colors that complies with the common-sense resultant colors of mixing colored pigments. The various mixed colors presented proves the successful activation of multiple CP gene expressions simultaneously, further establishing the induction system's effect in regulating gene expression in *H. bluephagenesis*. Furthermore, the colors produced on the agar plates are the same colors desired by researchers, who selected and used the corresponding inducer concentrations that, in liquid culture in the deep well plates, produced those specific desired colors as reference. The successful production of colored proteins on the agar plates proves the recombinant cells' consistency in responding to the different inducers and their respective varying concentrations in both liquid and solid culture (Figure 4 and Figure 5).

Figure 5. *Cell paintings as results of chromoprotein expression in H. bluephagenesis on agar plates with different inducers concentrations in select parts of the plate that made shapes with various colors constituting the paintings*

In summary, the results described above lead to the conclusions of 1) a trend showing the general increased in color intensity resulted from the augmentation of inducer concentration thus the successful fine-tuning of chromoprotein expression flux, and 2) the feasibility of the induction systems in expression of multiple chromoprotein genes at the same time and in producing various mixed colors. Overall, the fine-tuning of genes for biosynthesis of substances like chromoproteins in the extremophilic but previously seldomly investigated non-model halophile *H. bluephagenesis* were made possible in this study.

The fine and precise simultaneous regulation of multiple gene expressions in the extremophilic halophile *H. bluephagenesis* bacteria consequently widens the door for future biological exploration of the behaviors and capabilities of this non-model bacteria. This study, thus, endows the prospect of using this bacteria for controllable bioproduction of other products other than chromoproteins, shedding light on further research in possible recombinant protein production in this bacteria for medical, industrial, or academic applications. Most importantly, the possibility successful application of this non-model bacteria as chassis for recombinant engineering, which is explored in this study, would have significant contribution to the sustainability of prospective scientific research with this bacteria, especially in terms of saving laboratory consumables and water and energy resources, allowed by its extremophilic and environmental toleration characteristics and its needlessness and unnecessity for sterile experimental conditions (Ye & Chen, 2021).

Besides its applications in biological research, the bacteria has been proven useful in producing producing colorful agar paintings that allow for countless number of colors that can be produced by the inducing multiple chromoprotein genes simultaneously for the artistic community (Liljeruhm et al., 2018), as combinations of inducer concentrations are shown to be able to produce novel mixtures of colors made up of combinations of chromoproteins.

Conclusion

This study successfully constructed and tested an induction system of five independently inducible genes by demonstrating and proving the hypothesis that the induction of this newly designed tunable expression system in the non-model extremophilic bacteria chassis *H. bluephagenesis* is viable and concentration-dependent, as shown in the resulting levels of chromoprotein production. Thus, a new inducible flux control system feasible is provided for this previously scarcely explored but valuable and environ-

mentally beneficial bacteria for synthetic biology research.

Future Direction

This experiment can be improved, on one hand, by inserting epitope tag genes, which were absent in the design of this study, onto the target plasmid so that the resultant chromoproteins can be purified and separated from the bacteria's group protein, for easier and precise quantification of the chromoproteins than the naked eye can distinguish. The quantification data can then be used to more accurately compare protein concentrations caused by each specific concentration of inducers, allowing for clearer determinations of flux yield as a result of various levels of induction. Although group protein measurements were done in this study, they were not accurate indicators of chromoproteins alone; thus, it is important to extract single proteins to determine the exact effects the tuning system has on bioproduction for future efforts.

In addition, future expansions on the success of these induction systems can be explored to regulate and hopefully optimize the flux of the bioproduction of other substances other than chromoproteins. Along with the benefits of reducing sterilization cost and waste, allowed by its extremophilic characteristics, investigating production using chassis could be a significant step in achieving the goals of Next Generation Industrial Biotechnology (NGIB) and a more sustainable and environmentally friendly future of biological research (Chen et al., 2022; Liljeruhm et al., 2018).

References

Alieva, N. O., Konzen, K. A., Field, S. F., Meleshkevitch, E. A., Hunt, M. E., Beltran-Ramirez, V., Miller, D. J., Wiedenmann, J., Salih, A., & Matz, M. V. (2008). Diversity and Evolution of Coral Fluorescent Proteins. *PLoS ONE*,– 3(7).– e2680. URL: https://doi.org/10.1371/journal.pone.0002680

Liljeruhm, J., Funk, S. K., Tietscher, S., Edlund, A. D., Jamal, S., Wistrand-Yuen, P., Dyrhage, K., Gynnå, A., Ivermark, K., Lövgren, J., Törnblom, V., Virtanen, A., Lundin, E. R., Wistrand-Yuen, E., & Forster, A. C. (2018). Engineering a palette of eukaryotic chromoproteins for bacterial synthetic biology. *Journal of Biological Engineering*,– 12(1). 8 p. URL: https://doi.org/10.1186/s13036–018–0100–0

Bao, L., Menon, P. N. K., Liljeruhm, J., & Forster, A. C. (2020). Overcoming chromoprotein limitations by engineering a red fluorescent protein. *Analytical Biochemistry*,– 611.– 113936 p. URL: https://doi.org/10.1016/j.ab.2020.113936

Ahmed, F. H., Caputo, A. T., French, N. G., Peat, T. S., Whitfield, J., Warden, A. C., Newman, J., & Scott, C. (2022). Over the rainbow: Structural characterization of the chromoproteins gfasPurple, amilCP, spisPink and eforRed. *Acta Crystallographica Section D Structural Biology*,– 78(5).– P. 599–612. URL: https://doi.org/10.1107/S2059798322002625

Chen, G.-Q., Zhang, X., Liu, X., Huang, W., Xie, Z., Han, J., Xu, T., Mitra, R., Zhou, C., Zhang, J., & Chen, T. (2022). Halomonas spp., as chassis for low-cost production of chemicals. *Applied Microbiology and Biotechnology*,– 106(21).– P. 6977–6992. URL: https://doi.org/10.1007/s00253–022–12215–3

Lozano Terol, G., Gallego-Jara, J., Sola Martínez, R. A., Martínez Vivancos, A., Cánovas Díaz, M., & De Diego Puente, T. (2021). Impact of the Expression System on Recombinant Protein Production in Escherichia coli BL21. *Frontiers in Microbiology*,– 12.– 682001 p. URL: https://doi.org/10.3389/fmicb.2021.682001

Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J., & Voigt, C. A. (2019). Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nature Chemical Biology*,– 15(2).– P. 196–204. URL: https://doi.org/10.1038/s41589–018–0168–3

Chen, G.-Q., & Jiang, X.-R. (2018). Next generation industrial biotechnology based on extremophilic bacteria. *Current Opinion in Biotechnology*,– 50.– P. 94–100. URL: https://doi.org/10.1016/j.copbio.2017.11.016

Liljeruhm, J., Funk, S. K., Tietscher, S., Edlund, A. D., Jamal, S., Wistrand-Yuen, P., Dyrhage, K., Gynnå, A., Ivermark, K., Lövgren, J., Törnblom, V., Virtanen, A., Lundin, E. R., Wistrand-Yuen, E., & Forster, A. C. (2018). Engineering a palette of eukaryotic chromoproteins for bacterial synthetic biology. *Journal of Biological Engineering*,– 12(1).– 8 p. URL: https://doi.org/10.1186/s13036–018–0100–0

Lozano Terol, G., Gallego-Jara, J., Sola Martínez, R. A., Martínez Vivancos, A., Cánovas Díaz, M., & De Diego Puente, T. (2021). Impact of the Expression System on Recombinant Protein Production in Escherichia coli BL21. *Frontiers in Microbiology*,– 12.– 682001 p. URL: https://doi.org/10.3389/fmicb.2021.682001

Pang, Q., Han, H., Liu, X., Wang, Z., Liang, Q., Hou, J., Qi, Q., & Wang, Q. (2020). In vivo evolutionary engineering of riboswitch with high-threshold for N-acetylneuraminic acid production. *Metabolic Engineering*,– 59.– P. 36–43. URL: https://doi.org/10.1016/j.ymben.2020.01.002

Tan, D., Xue, Y.-S., Aibaidula, G., & Chen, G.-Q. (2011). Unsterile and continuous production of polyhydroxybutyrate by Halomonas TD01. *Bioresource Technology*,– 102(17).– P. 8130–8136. https://doi.org/10.1016/j.biortech.2011.05.068

Ye, J.-W., & Chen, G.-Q. (2021). *Halomonas* as a chassis. *Essays in Biochemistry*,– 65(2).– P. 393–403. URL: https://doi.org/10.1042/EBC20200159

audreyryhuang@outlook.com

submitted 25.05.2024;
accepted for publication 08.06.2024;
published 22.07.2024

Contact: audreyryhuang@outlook.com

The European Journal of Biomedical and Life Sciences 2024, No 1–2
ISSN 2310-5674

PREMIER Publishing
ppublishing.org

Section 3. Preventive Medicine

DOI:10.29013/EJBLS-24-1.2-46-50

TO STUDY THE EFFECTIVENESS OF IMMUNE PROTECTION AFTER VACCINATION AGAINST VIRAL HEPATITIS B

Tadjiyev Botir Mirxashimovich [1], ***Matyakubov Mansurbek Baxtiyar o'g'li*** [1]

[1] Republican Specialized Scientific and Practical Medical Center of Epidemiology, Microbiology, Infectious and Parasitic Diseases, Uzbekistan, Tashkent

Cite: *Tadjiyev B.M., Matyakubov M.B. (2024). To study the effectiveness of immune protection after vaccination against viral hepatitis B. The European Journal of Biomedical and Life Sciences 2024, No 1–2.* https://doi.org/10.29013/EJBLS-24-1.2-46-50

Abstract

The aim of the study was to assess the immunity of vaccinated children against hepatitis B. Vaccinated children have a high level of protection against hepatitis B, which confirms the importance of vaccination in the prevention of infectious diseases among children. Further research will deepen our understanding of the effectiveness of vaccination and its role in ensuring public health.

Keywords: *viral hepatitis B, gender differences, immunity, vaccine*

Introduction

Infectious liver infections are a serious global public health problem, having a significant negative impact on the health of the population and the health system as a whole (Nelson N. P. et al,. 2016).

Hepatitis B is caused by the hepatitis B virus (HBV), which affects the liver, causes inflammation and death of its cells. HBV infection can be acute or chronic, and the severity of the disease varies from asymptomatic (no symptoms) to severe.

According to statistics provided by the World Health Organization (WHO), 820 thousand people die every year from complications caused by viral hepatitis B. Of this number, more than 650 thousand are associated with the development of cirrhosis or hepatocellular carcinoma, and approximately 130 thousand cases are associated with acute viral hepatitis (Nayagam S. et al,. 2016 and Wang H. et al,. 2018). In the world today, 296 million people are infected with chronic viral hepatitis, while 1.5 million new cases are infected annually. Mortality from hepatocellular carcinoma ranks third among oncological diseases in the world (Sung H et al,. 2020).

The vaccine contains a modified hepatitis B virus surface antigen (HBsAg), which is artificially created using a special yeast culture. This antigen cannot cause disease, but the human immune system perceives it as a pathogen and reacts by producing protective antibodies (AHBs).

The protective efficacy of hepatitis B vaccination is associated with the induction of anti-HBs antibodies, but also includes stimulation of T cell memory.3 The level of anti-HBs equal to 10 mMU per ml, determined 1–3 months after the last dose of the initial course of vaccinations, is considered as a reliable marker of protection against infection.

Method

A thorough study of 280 blood samples using enzyme immunoassay was conducted at the Republican Center for Epidemiology, Microbiology, Infectious and Parasitic Diseases. Various methods have been used for this purpose, including clinical, diagnostic and statistical approaches. In Tashkent, the capital of Uzbekistan, 280 children aged 0 to 8 years were tested for hepatitis B antigen and a special enzyme immunoassay. The first screening of the film in Russian was dedicated to 280 films, including «Express Test» and «Concise Translation». Let's do it with a brief analysis. It was very important to draw attention to HBs. The main group consisted of children who had been vaccinated against hepatitis B and who had vaccination data, after which an additional statistical analysis was carried out.

For the rapid distribution of HBsAg, a test system «COMBINED RAPID TEST for HBs» was created by the American manufacturer Aria (CTK Biotech, Inc., publication date 2025–01–29). After the proclamation of the Republic of Belarus, a concise analytical center was created to create an antigen using the v – HBS antigen test system in Vectohep from vector-the best.

To measure the level of protection against hepatitis B, enzyme immunoassay (ELISA) was used using the Vector-Best reagent kit, whose certification complies with ISO 13485 and is valid until May 12, 2024, to detect the concentration of antibodies to HBs. The blood for the study was delivered to the laboratory in a specialized refrigerated container, and then the serum was stored at –20 °C. Before the rapid analysis and enzyme immunoassay for antigens and antibodies, the serum was left to approach room temperature for 60 minutes.

Results

As part of the laboratory study, the blood of 280 children from birth to 8 years old from Tashkent was analyzed. 280 serum tests were performed for the presence of HBs antigen (using the express method and ELISA). Next, we tested these samples for antibodies to HBs through enzyme immunoassay. The main group for analysis included children who had been vaccinated against hepatitis B, and due to the availability of vaccination cards, we conducted an additional statistical analysis of these cards.

According to the results of our study, it was revealed that 21 children (7.5%) did not have vaccination documents, 15 children (5.4%) were vaccinated once, 7 children (2.5%) received two vaccinations, 28 children (10%) went through three vaccinations, and the vast majority, 209 children (74.6%), were vaccinated four times.

Table 1. *Percentage of children surveyed by gender*

	categories	abs.	%	95% DI
gender	boys	145	518	458–578
	girls	135	482	422–542

Note. total N=280

There were more boys in the study group (52%, or 145 young participants) compared to girls (48%, or 135 participants) (Table 1).

During the analysis of children with the absence of HBs antigen for antibodies to it, it was revealed that antibodies are present in 62% of the tested, which is 280 people.

Table 2. *Data analysis of vaccination cards for children (taking vaccines)*

categories	abs.	%	95% DI
no vaccination card	21	75	47–112

categories	abs.	%	95% DI
received 1 vaccine	15	54	30–87
received 2 vaccine	7	25	10–51
received 3 vaccine	25	89	59–129
received 4 vaccine	212	757	703–806

Note. total N=280

Hepatitis B virus in vaccines accounts for 100% of the total amount of emlangan, 3 and 91.5% of the total number of vaccines taken (Table 2).

Table 3. *Distribution of the presence of antibodies to the HBs antigen and the antigen itself among the studied children.*

gender	Anti-HBs antitelo positive	Anti-HBs antitelo negative
boys	94 (65%)	49 (34%)
girls	74 (55%)	56 (41%)

Note. total N=280

Gender-based data indicate a higher level of immunity among boys compared to girls (Table 3).

Additionally, in the remaining sera without HBs antigen, a quantitative analysis was performed for the presence of antibodies to Hbsag by enzyme immunoassay, which affected 273 samples.

Of the remaining 273 children who participated in our study, all were vaccinated against hepatitis B, and then tested for antibodies to the HBs antigen. Antibodies to HBs were found in 47% of children who received 1 vaccine, in 86% of children who received 2 vaccines, in 81% of children who received 3 vaccines, in 60% of children who received 4 vaccines, and in 57% of children for whom there was no information about vaccination. This indicates that not all children have an immune response to the antigen, but most vaccinated children still have an immune response.

When analyzing the immune status of children participating in the study, it was found that 74% of children under the age of one year, 67% of children aged 1 year, 56% of children aged 2 years, 76% of children aged 3 years, 55% of children aged 4 years, 40% of children aged 5 years old, 62% of children aged 6 years, 50% of children aged 7 years had antibodies to the anti-HBs antigen. From this it can be concluded that most children aged 0 to 3 years have a stable immune status, which begins to weaken from the age of 4, however, there are cases of maintaining immunity without loss of effectiveness.

Table 4.*Detection of anti HBS titration in children's blood serum*

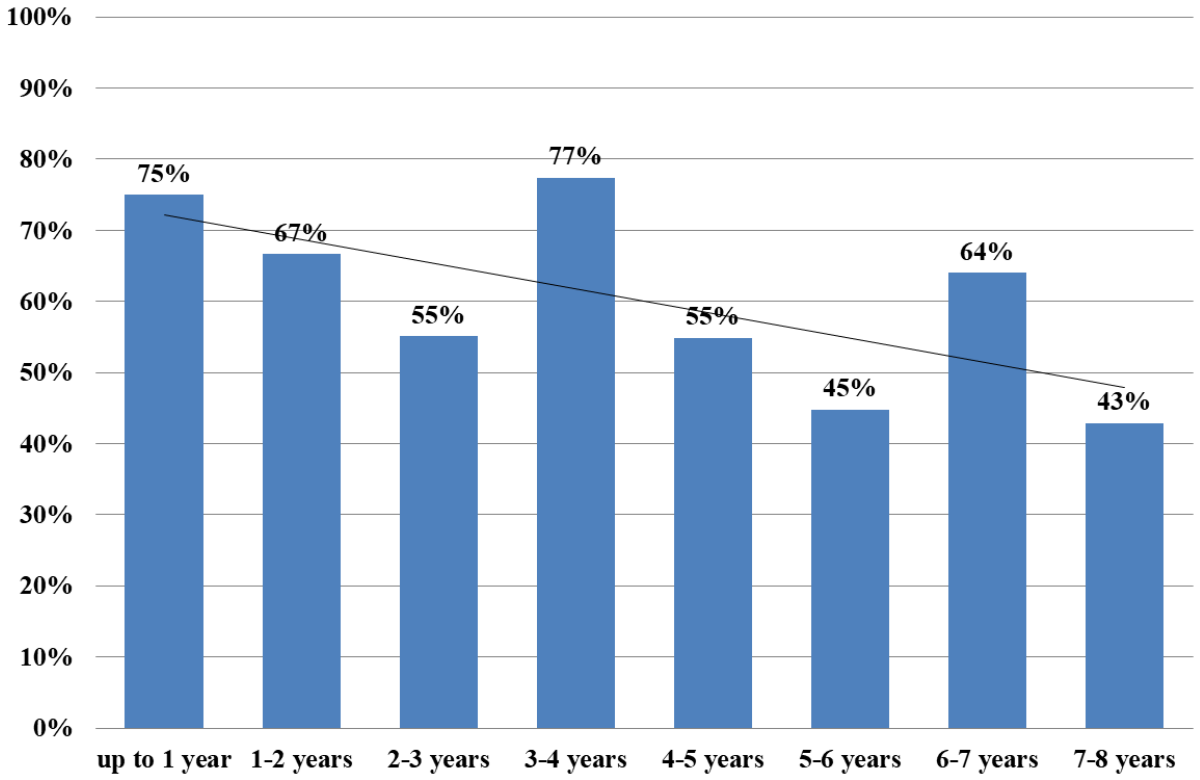
AntiHBS titre	abs	%	95% DI
<10 MMU/ml	105	375	318–435
10–100 MMU/ml	131	468	408–528
100–200 MMU/ml	17	61	36–95
200–300 MMU/ml	12	43	22–74
300–400 MMU/ml	12	43	22–74
400 MMU/ml <	3	11	02–31

Note. total N=280

When evaluating the titers of antibodies to HBs in the blood of 37.5% (105) children, it was found that the titers of antibodies to HBs did not exceed 10 MMU/ml, in children 46.8% (131) titers of 10–100 MMU/ml, in 15.7% (44) children titers above 100 MMU/ml were detected, in including 100–200 MMU/ml in 6.1% (17 out

of 12) children and 4.3% (12) children, antibody titers of 200–300 MMU/ml, 300–400 MMU/ml were detected in the blood/4.3% (12) and 1.1% (3) of children had antibodies in amounts exceeding 400 MMU/ml (Table 4).

Figure 1. *Determination of AntiHBs for children's age (%)*



Note. total N=280

The assumption of the detection of anti-HBs in relation to the age of children (Figure 1.) showed that the percentage of detection of antibodies decreases as children age. In 75% of children under one year of age, we can see that by the age of 7 this figure was 43%.

Discussion

The conducted study did not find positive results for the viral hepatitis B antigen, which is proof of the effectiveness of the immunity acquired as a result of vaccination. It is important to note that over time, this immunity may begin to weaken due to lack of contact with the antigen. This underlines the need for regular monitoring of the condition and, if necessary, additional vaccination to maintain the protective functions of the body.

It is important to understand that the immunity acquired as a result of vaccination is an essential element for protecting the body from viruses and diseases. However, like any system, it needs support and care. Regular monitoring and monitoring of the state of immunity will help to identify possible problems in a timely manner and take the necessary measures to eliminate them.

References

URL: https://cdn.who.int/media/docs/librariesprovider2/euro-health-topics/vaccines-and-immunization/hepatitis_ru.pdf
URL: https://helix.ru/kb/item/40-089
URL: https://ogepatite.ru/b/naprjazhennost-immuniteta.html

Nelson N. P., Easterbrook P. J., Mc Mahon B. J. Epidemiology of hepatitis B virus infection and impact of vaccination on disease. Clin. Liver Dis., 2016. – Vol. 20. – No. 4. – P. 607–628. Doi: 10.1016/j.cld.2016.06.006

Nayagam S., Thursz M., Sicuri E., Conteh L., Wiktor S., Low-Beer D., Hallett T. Requirements for global elimination of hepatitis B: a modelling study. Lancet Infect. Dis., 2016. – Vol. 16. – No. 12. – P. 1399–1408. Doi: 10.1016/S1473–3099(16)30204

Sung H., Ferlay J., Siegeszl R. L., Laversanne M., Soerjomataram I., Jemal A., et.al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 2020; 71(3): 209–49. URL: https://doi.org/10.3322/caac.21660

Wang H., Tian C. W., Wang W. M., Luo X. M. Time-series analysis of tuberculosis from 2005 to 2017 in China. Epidemiol. Infect., 2018. – Vol. 146. – No. 8. – P. 935–939. Doi: 10.1017/S0950268818001115

submitted 14.05.2024;
accepted for publication 28.05.2024;
published 22.07.2024

Contact: matyakubov.mansur@mail.ru

Contents