European Journal of Biomedical and Life Sciences

№ 4 2020

PREMIER Publishing

Vienna
2020

European Journal of Biomedical and Life Sciences

Scientific journal

№ 4 2020

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, German/English/Russian language, peer-reviewed journal. It is published bimonthly with circulation of 1000 copies.

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://www. 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:

ULRICHSWEB™ GLOBAL SERIALS DIRECTORY · НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА eLIBRARY.RU · CYBERLENINKA · Google scholar

GIF® GLOBAL IMPACT FACTOR The journal has the GIF impact factor .562 for 2018.

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

Section 1. Clinical Medicine

https://doi.org/10.29013/ELBLS-20-4-3-5

Alatrash Yehya Ahmad,
6th year student, Faculty of Medicine 2
of SUMPh, "Nicolae Testemitanu",
Chisinau, Republic of Moldova
E-mail: yahia0alatrash@gmail.com

Buruiana Sanda,
Associate Professor of the Department of Hematology
SUMPh, "Nicolae Testemitanu"
Chisinau, Republic of Moldova
E-mail: sandaburuiana69@gmail.com

THERAPEUTIC APPROCHES OF β-THALASSEMIA

Abstract. *β*-thalassemia is an inherited disorder of haemoglobin associated with ineffective erythropoiesis and anemia. During the last 40 years, in addition to the considerable progress made in prevention and treatment of thalassemias, there have also been major advances in their symptomatic management, at least in wealthier countries where appropriate facilities are available. Therapeutic management depends on the severity of the disease.

Keywords: Thalassemia, treatment, erythropoiesis.

1. Background

Thalassemias are characterized by compromised red blood cell survival due to mutations in the genes encoding *α*- and/or *β*-globins. Imbalanced globin chain production results in excess *β* – or *α* -globin precipitation, reduced red blood cell fitness, and ineffective erythropoiesis and hemolysis [1]. *β*-thalassemia is an inherited disorder of haemoglobin associated with ineffective erythropoiesis and anemia [2]. The incidence of thalassemia carriers is high in regions such as Mediterranean, Indian subcontinent, Middle East, South China and Southeast Asia [3]. In the past few decades, migrants from the thalassemia prevalent countries to non-prevalent countries, mainly North America and Central and North Europe, are rapidly increasing in number [3].

During the last 40 years, in addition to the considerable progress made in prevention and treatment of thalassemias, there have also been major advances in their symptomatic management, at least in wealthier countries where appropriate facilities are available [4].

2. Study methods

Electronic searches using Google scholar, PubMed, Cochrane Library were performed for studies published in English and Romanian between 2010–2020.

3. Results

Therapeutic management depends on the severity of the disease. The most severe clinical form

is thalassemia major. The treatment depends on whether the patient is transfusion-dependent patients, frequent and lifelong red blood cell transfusions or is without blood transfusion [5].

Depending on the pathogenetic dysfunctions, we distinguish different treatment methods: blood transfusions, stimulation of Hb F production, correcting dyserythropoiesis, antioxidative treatment, gene therapy, allogeneic hematopoietic stem cell transplantation and prophylaxis of possible complications removal of excess iron, infections and other [5].

Blood transfusion is a symptomatic method of compensating for low erythrocyte counts in the blood. In the 1960 s and 1970 s, the only effective treatment for thalassemia was blood transfusion [6]. Transfusions temporarily relieve anemia, but do not restore normal erythropoiesis and secondarily various complications can be induced such:

- increased rate of transfusion infections (viral B, C hepatitis, human immunodeficiency virus);
- increased rate of allergic reactions to blood components;
- increasing the level of iron in the body with its accumulation in various organs.

Stimulation of Hb F production because high levels of Hb F ameliorate the severity of the disease, mainly by reducing the surplus of α-globin chains [5]. Currently, the only compound in clinical use is hydroxyurea, an S-phase cell cycle inhibitor. However, its mechanism of action on Hb F remains elusive, a subset of patients is resistant, and being myelosuppressive necessitates careful monitoring of patients. Hydroxyurea can attain transfusion independency among patients with certain primary and secondary genetic modifiers from the time of diagnosis [5; 7].

New agents include those that affect chromatin regulators (such as decitabine on DNA methylation and histone deacetylase inhibitors) and others that affect DNA-binding transcription factors [5]. Increased production of γ-globin has been accomplished using lentiviral vectors that express a zinc finger protein. Have been identified BCL11A and ZBTB7A, two potent transcriptional repressors of γ-globin [5]. They act with additional trans-acting epigenetic repressive complexes, lineage-defining factors, and developmental programs the γ-globin genes by working on cis-acting sequences at the globin gene loci. Inhibition of these repressors could reactivate γ-globin production in adult patients [5].

Correcting dyserythropoiesis can be performed with the help of:

– Activin receptor-II trap ligands [2]. Luspatercept, is a first-in-class erythroid maturation agent approved for adult transfusion-dependent patients with β -thalassemia who require regular red blood cells transfusions, binds several TGF- β superfamily ligands to diminish Smad2/3 signaling and enhance late-stage erythropoiesis [2; 5]. Efficacy and safety of Luspatercept in adult patients dependent on transfusion with-β thalassemia requiring regular transfusions of red blood cells has been demonstrated in phase 3 of BELIEVE double-blind randomized, placebo-controlled study [2].

– JAK2 inhibitors: β-thalassemia mice have elevated erythropoietin levels associated with increased JAK2 phosphorylation, resulting in ineffective erythropoiesis and extramedullary hematopoiesis. JAK2 inhibitors could be for non-transfusion-dependent patients with splenomegaly, because these drugs effectively reduce splenomegaly in such mice [5].

– Induction of the heat shock protein 70 (Hsp70) chaperone machinery and is needed for normal termination of erythropoiesis. Factors that control the nucleocytoplasmic trafficking of proteins and RNAs, inhibitors of erythroid progenitors from β-thalassemia major patients, demonstrated induction of HSP70 nuclear localization, and improved terminal erythroid differentiation [5].

Antioxidative treatment: exogenous antioxidants, activation of endogenous antioxidant proteins [5].

Gene therapy. Studies of gene therapy have utilized mainly lentivirus vectors in experimental systems, including cultured CD34 from β-thalassemia

patients and β-thalassemia mouse models. Yet the safety profile of such technologies is still uncertain [5].

Allogeneic hematopoietic stem cell transplantation is a treatment for patients with *β*-thalassemia major, with good risk features have a > 90% chance of a successful outcome but this type of transplantation in high-risk patients is challenging because of graft rejection and transplant-related mortality. Stem cell transplantation is not practical due to a variety of factors like financial costs, donor unavailability and scarcity of transplantation facilities [5; 8].

Removal of excess iron. Chronic transfusion therapy leads to iron overload and, if untreated, usually results in severe organ damage. Removal of excess iron from various tissues, e.g., the liver spleen, heart, and the pituitary, in beta thalassemia patients, has become an essential therapy to prolong life [6; 9]. Due to the early start of iron chelation therapy, a profound knowledge of efficacy and safety in young patients is essential [6; 9]. There are several possibilities for reducing iron in the human body:

– Iron chelators. There are three iron chelators licensed: deferoxamine, deferiprone and deferasirox;

– Modulation of iron absorption (administration of hepcidin);

– Stimulating its expression (inhibition of negative regulators, inhibition of erythroferrone) [6; 9].

4. Conclusion

Based on the results of the literature review we can say that the therapeutic management depends on the severity of the disease, by the development of pharmaceutical technologies and last but not least by the socio-economic possibilities of the country.

References:

1. Kuo K., Layton D. et.al. 7 proof of concept for the oral pyruvate kinase activator mitapivat in adults with non-transfusiondependent thalassemia: interim results from an ongoing, phase 2, open-label, multicenter study. Hema Sphere. Abstract book. 2020; 4(S_1): 109.
2. Taher A., Viprakasit V., Cappellini M. Assessment of longer-term efficace and the fase 3 believe trial of luspatercept to treat anemia in in patients with B-Thalassemia. HemaSphere. Abstract book. 2020; 4(S1): 108.
3. Li C. K. New trend in the epidemiology of thalassemia. Best Practice & Research Clinical Obstetrics & Gynaecology. 2017; 39:16–26.
4. Olivieri N., Brittenham G. Management of the Thalassemia. Cold Spring Harb Perspect Med. 2013; 3(6): a011767.
5. Fibach E., Rachmilewitz E. Pathophysiology and treatment of patients with beta-thalassemia-an update. F 1000 Research. 2017; 6: 2156.
6. Crichton R., Ward R., Hider R. et. al. The efficacy of iron chelators for removing iron from specific brain regions and the pituitary-ironing out the brain. Pharmaceuticals. 2019; 12(3): 138.
7. Hussain S., Munzir S., Ali S. Effectivness of Hydroxyurea therapy: a 15 year experience of Thalassemia management without blood transfusion. HemaSphere. Abstract book. 2020; 4(S1): 720.
8. Yu U., Wang X., Zhang X. Outcomes of haplo-cord transplantation in children with beta thalassemia major. HemaSphere. Abstract book. 2020; 4(S1): 718.
9. Botzenhardt S., Li N., Chan E. et. al. Safety profiles of iron chelators in young patients with haemoglobinopathies. European Journal of Haematology. 2016; 98: 198–217.

https://doi.org/10.29013/ELBLS-20-4-6-8

Asibi (Abu Shtiwi) Anas Dahar,
6th year student, Faculty of Medicine 2
of SUMP, "Nicolae Testemitanu",
Chisinau, Republic of Moldova
E-mail: alasibianas11@gmail.com

Buruiana Sanda,
Associate Professor of the Department of Hematology
SUMP, "Nicolae Testemitanu"
Chisinau, Republic of Moldova
E-mail: sandaburuiana69@gmail.com

MANAGEMENT OF HEMOPHILIA A TREATMENT: PAST AND PRESENT

Abstract. Hemophilia A is the most common severe hereditary hemorrhagic disorder, caused by the deficiency of clotting factor VIII in the blood. The management of Hemophilia A treatment in different historical periods has been depending on level of knowledge about the disease, technological innovations and is constantly improving. Knowing the historical evolution of the treatment applied to patients with Hemophilia A will help us to select an individual therapeutic regimen for each patient.

Keywords: Hemophilia, treatment, clotting factor VIII.

1. Background

Hemophilia, which means love (philia) of blood (hemo), is the most common severe hereditary hemorrhagic disorder [1; 2]. Hemophilia A is usually an inherited condition and caused by the deficiency of clotting factor VIII in the blood. The estimated frequency of hemophilia is around 1 in 10000 live births, and the number of people worldwide living with hemophilia is about 400000. It presents in 1 in 5000 live male births, whereas hemophilia B presents in 1 in 30000 live male births [1; 3]. Over the years, the treatment of Hemophilia has had multiple changes that clearly favored the quality of treatment, the quality of life of the patient with Hemophilia A.

2. Study methods

Electronic searches using Google scholar, PubMed, Cochrane Library were performed for studies published in English and Romanian between 2010–2020.

3. Results

The management of Hemophilia A treatment in different historical periods has been depending on level of knowledge about the disease, technological innovations and others. In the years 1950–1970 the purpose of treatment was only to replace the level of factor VIII in the blood by plasma transfusion, cryoprecipitate [4; 5]. In 1964 Dr. Judith Graham Pool, a scientist and researcher, made the important discovery that the fraction cryoprecipitate from plasma (the precipitate left from thawing plasma contained large amounts of FVIII). This cryoprecipitate could be given to hemophilia A patients in relatively small volumes to help with clotting during bleeding episodes. The shortage in supply led to treatment guidelines including episodic therapy as an alternative for prophylaxis [5]. Treatment to replace factor VIII with blood products has been shown to have many side effects:

– increased rate of transfusion infections (viral B, C hepatitis, human immunodeficiency virus);

– increased rate of allergic reactions to blood components;

– plasma or cryoprecipitate was performed only in hospital conditions, so the patient had to travel to the hospital;

– difficult venous access in both, children and adults, which conditions low adherence to treatment;

– early invalidation of the patient due to frequent relapses of hemorrhage in the joint;

– decrease the quality of life of patients with Hemophilia A with economic impact.

The modern treatment of hemophilia A is considered to have started in the 1970 s, with the production of lyophilized plasma concentrates of coagulation factor VIII. This technological innovation greatly improved the quality and expectancy of life of people with hemophilia A as it enabled the widespread adoption of home replacement therapy with the early control bleeding episodes [5]. However, the treatment remained to be applied only in case of development of hemorrhagic complications.

The discovery in 1977 of the synthetic agent desmopressin provided a new, inexpensive and safe treatment for many patients with mild hemophilia A, which reduced the exposure to non-virus inactivated plasma-derived products.

The most important advance in this field was based on the rapid progress in DNA technology (following the cloning in 1982 of FVIII genes), which allowed the industrial production of recombinant FVIII [3; 5]. The treatment of hemophilia A may involve:

– home replacement therapy with the early control bleeding episodes;

– qualitative prophylactic treatment;

– management of bleeding episodes;

– treatment and rehabilitation of hemophilia synovitis;

– decreased rate of transfusion infections (viral B, C hepatitis, human immunodeficiency virus);

– increased the quality of life of patients with Hemophilia A;

– hemophilia treatment Centre network expanded.

But at the same time, negative effects were determined, such as:

– difficult venous access it remained a big problem, especially in home condition. A family member had to be taught to administer the intravenous factor VIII;

– frequent intravenous injections, which reduced adherence to treatment;

– short half-life of the synthetic factor;

– increased the risk of antibody formation to factor VIII.

This moment motivated the study of alternative treatment methods, such as: the creation of factor VIII preparations with a longer half-life, non-factoring agents, genetic treatment, the possibility of subcutaneous administration [6].

The treatment with non-factoring agents is given depending on whether or not antibodies have formed. In the case of Hemophilia A with inhibitors can be used: induction of immunological tolerance, Emicizumab, recombinant coagulation factor VIIa (rFVIIa). In the case of Hemophilia A without inhibitors can be used: Emicizumab, substitution treatment with coagulation factor [7; 8]. It is recommended to increase the level of factor VIII by more than 12–15% [9].

Gene therapy means a technique of introducing a nucleic acid into a patient's cells in order to treat certain diseases [5]. There are methods that use viral vectors (biological nanoparticles), videlicet modified viruses and non-viral methods that are based on naked DNA complexes. Because hemophilia involves a genetic defect, viral vectors are used to deliver unmodified copies of genes into the patient's body. Non-viral methods are considered superior because they allow the production of proteins on a large scale and require a lower immune response from the host. Gene therapy in hemophilia A was evaluated in a study by BioMarine company.

The implementation of these treatment methods has a positive impact which is manifested by:

- Decreases the risk of possible bleeding;
- Preserving the structure and functionality of joints in children;
- Stopping the destruction of the joint in children and adults in whom arthropathy began to develop;
- Improving the quality of life [5; 8].

4. Conclusion

Based on the results of the literature review we can say that the modern treatment of patients with Hemophilia A should be applied not only during bleeding, but also prophylactically for the prevention of possible hemorrhages. The goal of treatment is to avoid any bleeding, which will increase the quality of life of these patients.

References:

1. Weyand A., Steven W. New therapies for hemophilia. Blood. 2019; 133(5): 389–398.
2. Castaman G., Matino D. Hemophilia A and B: molecular and clinical similarities and differences. Haematologica. 2019; 104(9): URL: https://doi.org/10.3324/haematol.2019.221093
3. Peyvandi F., Garagiola I., Young G. The past and future of haemophilia: diagnosis, treatments, and it complications. – Lancet. 2016; PubMed: PMID: 26897598.
4. Guidlinex for the management of haemophilia 2012.
5. Franchini M. The modern treatment of haemophilia: a narrative review. Blood Transfusion. 2013; 11(2): 178–182.
6. MASAC. Treatment Recommendations 2020. URL: https:/www.hemophilia.org/Researches Healthcare-Providers/Medical and Scientific
7. Kitazova T et. al. Factor VIIIa-mimetic cofactor activity of a bispecific antibody to factors IX/IXa and X/Xa, Emicizumab, depends on its ability to bridge the antigens. Thromb Haemost. 2017; 117(07): 1348–1357.
8. Young G. A multicenter open-label, phase 3 study of emicizumab prophilaxis in children with Haemophilia A with inhibitors. Blood. 2019; 134 (24): 2127–2138.
9. Jimenez-Yuste V et al. Achieving and maintaining an optimal throught level for prophilaxis in hemophilia: the past, the present and the future. Blood Transfusion. 2014; 12(3): 314–319.

https://doi.org/10.29013/ELBLS-20-4-9-14

Mursalova Zenfira Shukurgizi,
Senior Researcher,
Scientific Research Institute of Pediatrics
E-mail: zenamursalova@gmail.com

PREDICTORS OF BRONCHOPULMONARY DYSPLASIA IN INFANTS WITH LOW BIRTH WEIGHT

Abstract: As a result of this study, it has been shown that preeclampsia, oophoritis, frequent acute respiratory viral infections in the mothers of the examined children, congenital pneumonia and severe anemia in children create conditions for the development of BPD. Assessment of significant informative diagnostic and prognostic factors allows timely detection of children from the risk group associated with the development of BPD.

Keywords: bronchopulmonary dysplasia, infants, low birth weight.

Мурсалова Земфира Шукур гызы,
Старший научный сотрудник,
Научно-Исследовательский Институт Педиатрии
E-mail: zenamursalova@gmail.com

ПРЕДИКТОРЫ БРОНХОЛЕГОЧНОЙ ДИСПЛАЗИИ У ДЕТЕЙ ГРУДНОГО ВОЗРАСТА, РОДИВШИХСЯ С НИЗКОЙ МАССОЙ ТЕЛА

Аннотация. В результате настоящего исследования показано, что преэклампсия, оофорит, частые ОРВИ у матерей обследованных детей, врожденная пневмония и тяжелая анемияу детей создает условия для развития БЛД. Оценка значимых информативных диагностических и прогностических факторов, позволяет своевременно обнаружить детей из группы риска, связанной с развитием БЛД.

Ключевые слова: бронхолегочная дисплазия, дети грудного возраста, низкая масса тела при рождении.

На сегодняшний день бронхолегочная дисплазии является одной из важных проблем в неонатологии и современной педиатрии [1,2,3]. Наиболее часто встречающимся осложнением у недоношенных новорожденных детей, в том числе родившихся с низкой и экстремально низкой массой тела, поступивших в отделение реанимации и интенсивной терапии с различными перинатальными патологиями является хроническое обструктивное заболевания легких – бронхолегочная дисплазия (БЛД) [4; 5]. Развитие БЛД у грудных детей связано с рядом неблагоприятных последствий, таких как полиорганная дисфункция [6].

Бронхолегочная дисплазия, возникшая у детей грудного возраста, это хроническое заболевание легких преимущественно недоношенных детей, характеризующееся повреждением всех структурных компонентов легкого и формирующееся

в процессе комбинированного воздействия первичного респираторного заболевания и интенсивной терапии дыхательных расстройств на незрелые легкие, с основными проявлениями в виде дыхательной недостаточности [5; 7].

В последнее время, в отличии от классической формы БЛД у глубоконедоношенных детей после полученной сурфактантной терапии при респираторном дистрес синдроме чаще регистрируется новая форма БЛД [2; 8; 9]. Развитию новой формы БЛД способствуют как антенатальные факторы (преэклампсия, тяжелая фетоплацентарная недостаточность, задержка развития плода), так и относительная гипероксия, задерживающая развитие легких на раннем этапе [10].

Длительное ИВЛ, реинтубация, тяжелая перинатальная асфиксия, тяжелые гипоксические ишемические повреждения ЦНС, сохранение персистентных коммуникаций, повышение легочного артериального давления и др. способствует ускоренному развитию БДЛ.

Результаты, проведенного рандомизированного ретроспективного исследования выявили, что определение факторов риска БЛД не всегда показывает реализацию заболевания, поэтому сложно предсказать результат заболевания у конкретного пациента. Поэтому следует обратить внимание на характер респираторной терапии во время реанимационно-интенсивного лечения, которое имеет большое значение в формировании группы риска в развитии БЛД. Исследования, посвященные влиянию антенатальных, постнатальных факторов на развитие бронхолегочной дисплазии у детей грудного возраста, родившихся с низкой массой тела актуальны и требуют дальнейшего изучения.

Цель исследования: определение информативных предикторов бронхолегочной дисплазии у детей грудного возраста, родившихся с низкой массой тела.

Материалы и методы исследования:

Дизайн исследования состоял из ретроспективного и проспективного обследования недоношенных детей с низкой массой тела при рождении в отделении реанимации и интенсивной терапии в Научно-Исследовательского Института Педиатрии имени К. Фараджевой. К исследованию были привлечены 128 детей с гестационным возрастом 28–36 недель (средний возраст 31.8 ± 2.1 недель), с массой тела при рождении 2360 г. (средняя масса 1588±340.0 г), переведенные с родильного дома на второй этап выхаживания. Среди них 72 мальчика, 56 девочек, в том числе 6 близнецов.

Обследованные дети были разделены на две группы. В первую группу вошли 32 ребенка (25.0%) с БЛД, во вторую 96 детей (25.0%) без признаков БЛД. На втором этапе был составлен прогностический алгоритм на основе антенатальных и неонатальных факторов риска. Параметр исключения: дети с массой больше 2500 г, возраст менее 1 месяца, пороки легких и сердца (кроме случаев – открытое овальное окно, открытый артериальный проток)пороки развития ЦНС, ВУИ со специфической этиологией, сепсис, синдром аспирации. Диагноз классической БЛД был поставлен на основании наличия в анамнезе РДС, ИВЛ с «жесткими параметрами» более 3х дней с рентгенологическими характерными признаками (вздутие легких, фиброз, буллы). Диагноз новой формы БЛД ставился детям с гестационным возрастом менее 32 недель, у которых применялись препараты сурфактанта для профилактики РДС, а респираторная поддержка была щадящей. Рентгенологически характерно гомогенное затемнение легочной ткани без ее вздутия. Для подтверждения диагноза БЛД проводилось клиническое обследование детей и инструментальные исследования (компьютерная томография, рентгенография).

Согласно последовательному статистическому исследованию А. Вальдина была выявлена частота встречаемости каждого признака и прогностический критерии БЛД. Полученные показатели обрабатывались по «Статистика-6».Различие в сравниваемых группах проводилось по критерию Фишера, критерий x^2.

Результаты исследования и их обсуждение

Все обследованные дети были распределены относительно веса при рождении следующим образом: дети с экстремальной низкой массой (меньше 1000 г)-10 (7,81%), 36 детей (28,1%) с очень низкой массой тела(1000–1500 г.) –66 детей (51,6%), с низкой массой тела (1500–2000 г), 16 детей (12,5%) с массой более 2000 г.

В первой группе средний гестационный возраст составил 29,8 ± 1,2 недель и колебался в пределах 28–33 недель. Масса тела при рождении составила 1208,4 ± 284,5 г. и была распределена следующим образом: до 1000 г. 4 ребенка (12,5%), в пределах 1000–1500 г.-14 (43,7%), в пределах 1500–2000 г.–12 (37,5%) и больше 2000 г.-2 (6,25%). В первой группе мальчиков было 20 (62,5%), девочек –12 (32,5%).

Во второй группе мальчиков было –42 (43,7%), девочек –54 (56,3%). Средний гестационный возраст составил 32,1 ± 2,1недели и колебался в пределах 28–36 недель. Масса тела при рождении составила 1700,5±322,5 г. и была распределена следующим образом: до 1000 г. 6 детей (6,25%), в пределах 1000–1500 г.–24 (25, 0%), в пределах 1500–2000 г.– 54 (56,2%) и больше 2000 г.-12 (12,5%) детей.

Таким образом, проведенный сравнительный анализ в первой и во второй группах показал, что показатели гестационого возраста и массы тела при рождении статистически различались ($p < 0.01$).

При анализе анамнестических показателей была выявлена частота их встречаемости. Сравнительная характеристика таких случаев как бесплодие, спонтанный аборт, угроза настоящей беременности, инфекционно-хронические заболевания у матери (урогенитальные, экстрагенитальные), соматические заболевания и др. Из таблицы 1.видно, что в первой группе у матерей при беременности ОРВИ (x^2= 6,29; $p < 0.05$), преэклампсия ($x^2 = 7,01$; $p < 0.01$) и оофорит ($x^2 = 9,09$; $p < 0.01$) статистически значимы по сравнению со второй группой.

Таблица 1.– Акушерско-гинекологический анамнез и частота встречаемости заболеваний матерей обследованных детей

Показатели	**1-гуппа n-32**	**2-группа n=96**	**Всего**	**X^2**
1	***2***	***3***	***4***	***5***
Бесплодие в анамнезе.	2(6,25)	9(9,37)	11	0,65
Прерывание беременности более 2–3 раз.	4(12,5)	15(15,6)	19	0,72
Угроза прерывания текущей беременности.	4(12,5)	12 (12,5)	16	0,23
Повторное ОРВИ во время беременности.	18(56,2)	27(28,1)	45	6,29*
Хроническая фетоплацентарная недостаточность.	13(40,6)	33(34,3)	46	0,87
Преэклампсия.	15(46,8)	24(25,0)	39	7,01*
Соматические заболевания:				
анемия.	12(37,5)	27(28,1)	39	0,96
сердечно-сосудистые заболевания.	4(12,5)	18(18,7)	22	1,08
заболевания дыхательных органов.	2(6,25)	7(7,31)	9	0,32
заболевания желудочно-кишечного тракта.	8(25,0)	33(34,3)	41	0,85
заболевания почек и мочевыводящих путей.	2 (6,25)	9(9,37)	11	0,71
эндокринные заболевания.	1(3,12)	6(6,25)	7	0,44
заболевания ЛОР органов.	5(15,6)	18(18,7)	23	0,19

1	*2*	*3*	*4*	*5*
Гинекологические заболевания у матери:				
кольпит.	6(18,7)	12 (12,5)	18	0,65
оофорит.	10(31,2)	6(6,25)	16	9,09*
эндометрит.	1(3,12)	3(3,12)	4	0,56
эрозия шейки матки.	9(28,1)	24(25,0)	33	1,66
миома шейки матки.	1(3,12)	2(2,08)	3	0,81
киста яичников.	1(3,12)	4(4,25)	5	0,52

Примечание: в скобках относительные (в процентах) показатели x–p < 0.01(x^2 -критерий)

В исследовании также проведено изучение частоты встречаемости коморбидных (сочетанных) заболеваний у обследованных детей. Из (таблицы 2) видно, что в первой группе у детей случаи апноэ (x^2 = 12,0), открытый артериальный проток (x^2 = =10,6; p < 0.01) и анемия недоношенных (x^2 = 11,7; p < 0.01) статистически значимы по сравнению со второй группой.

Таблица 2.– Частота встречаемости коморбидных сочетанных заболеваний у обследованных детей

Заболевания	**1-гуппа n-32**	**2-группа n=96**	**Всего**	**x^2**
Гипоксически-ишемические поражения ЦНС (ПВЛ, киста).	15(46,9)	48(50,0)	53	0,48
Гипоксически-геморрагические поражения ЦНС.	17(53,1)	48(50,0)	65	0,27
Инфекционные поражения ЦНС.	4(12,5)	11(11,4)	15	0,51
Токсическо-метаболические поражения ЦНС.	3(9,37)	6(6,25)	9	0,92
Ведущий синдром при поражениях ЦНС:				
судорожный синдром.	4(12,5)	9(9,37)	13	0,62
гипертензионный синдром.	3(9,37)	7(7,35)	10	0,39
гидроцефальный синдром. вегето-висцеральный синдром.	4(12,5)	10(10,4)	14	0,59
апноэ.	6(18,6)	9(9,37)	15	12,12*
Коньюгационная гипербилирубинемия.	12(37,5)	30(31,2)	42	0,69
Открытое овальное окно.	20(62,5)	42(43,3)	62	0,65 *
Открытый артериальный проток.	6(18,6)	6(6,25)	12	10,6*
Анемия недоношенных.	24(75,0)	54(56,2)	78	11,7*

Примечание: ПВЛ – перивентрикулярная лейкомаляция; ПВК – перивентрикулярное кровоизлияние; ВЧК – внутрижелудочковые кровоизлияние

Для прогнозирования развития БЛД для выявления высокой группы риска применялся последовательный метод А. Вальда. В результате анализа выявлены индивидуальные особенности матери и ребенка. На основе полученных данных рассчитан прогностический и информативный коэффициенты. На основе корреляционного анализа были определены более характерные признаки, имеющие большое значение при развитии БЛД.

Был проведен сбор прогностических коэффициентов для прогностической информации и их обработка. Сбор прогностической информации был проведен до определенного уровня (порог) (+13 или –13).Если количество прогностических коэффициентов больше 13 – то это означало развитие патологического процесса (неблагоприятное течение), с +13 до –13 неопределенный ответ, больше –13 – отсутствие патологического процесса, то есть развитие БЛД (благоприятное течение). В реальности несоответствие прогноза принимается пределах 5%. Его повышение объясняется двумя причинами: 1)принятие во внимание всех диагностических признаков при обработке; 2) дополнительные факторы, которые не были приняты во внимание. Ясно, что если врач сможет учесть все факторы, то прогноз будет точным.

В настоящем исследовании были выбраны статистически значимые ($p < 0.01$) и обладающие корреляционными связями клинические признаки, имеющие важное прогностическое значение при развитии БЛД, также был разработан диагностический алгоритм.

Таблица 3.– Диагностический алгоритм для прогнозирования развития БЛД у грудных детей

Признаки	**Диапазон признаков или варианты**	**Информативность**	
		ДК	**ИК**
Гестационный возраст.	<32 недель	+4,5	0,98
	32–37 недель	–3,0	0,56
масса тела.	<1500 г	+4,2	0,92
	1500–2000 г	–3,0	0,54
РДС.	7–10 баллов	+3,5	0,70
		–2,0	0,41
nCPAP.	<7 баллов	+4,6	0,85
		–3,0	0,56
Длительная ИВЛ (меньше 5 дней)	Да	+3,5	0,72
	Нет	–2,5	0,48
Необходимость в кислороде после недели жизни.	Да	+5,6	1,66
	нет	–3,2	0,84
Внутриутробная пневмония.	Да	+4,8	1,02
	нет	–3,4	0,74
Открытый артериальный проток.	Да	+4,2	1,08
	нет	–3,2	0,81
Тяжелая форма анемии.	Да	+4,6	0,98
	нет	–3,2	0,66
Апноэ.	Да	+4,0	1,09
	нет	–3,6	0,72

Примечание: ДК – диагностический коэффициент (>+ 2.5). ИК – информативный коэффициент (> 0.25). РДС – респираторный дистресс синдром. CPAP – режим искусственной вентиляции лёгких постоянным положительным давлением. ИВЛ – искусственная вентиляция лёгких

Как видно из (таблицы 3) к высоким информативным факторам развития БЛД можно отнести 21% необходимость в кислороде (после первой недели жизни) (ИК=1,66), внутриутробную пневмонию (ИК=1.02), апноэ (ИК=1,0), тяжелая форма анемии (ИК=0,98),

гестационный возраст менее 32 недель (ИК=0,98).

Таким образом, проведенный ретроспективный анализ показал, что при беременности у матерей, обследованных детей наличие ОРЗ, преэклампсии и оофорита, апноэ, анемии, открытый артериальный проток, внутриутробная пневмония и тяжелая анемия у детей создает условия для развития БЛД. Оценка значимых информативных диагностических и прогностических факторов, позволяет своевременно обнаружить детей из группы риска, связанной с развитием БЛД.

Выводы:

1.Антенатальные факторы риска развития БЛД у детей с низкой массой тела при рождении это преэклампсия ($x^2 = 7{,}01$; $p < 0.01$), повторные ОРВИ ($x^2 = 6{,}20$; $p < 0.01$) и оофорит ($x^2 = 9{,}09$; $p < 0.01$) у матерей при беременности.

2. К неонатальным предикторам БЛД относятся масса тела при рождении ниже 1500 г., гестационный возраст менее 32 недель, тяжелый РДС, CPAP и длительная ИВЛ, необходимость в кислороде после недели рождения, открытый артериальный проток, внутриутробная пневмония, тяжелая форма анемии.

3. К высоким информативным факторам развития БЛД у грудных детей относится – необходимость в более чем 21% в кислороде (после первой недели жизни)(ИК=1,66), внутриутробная пневмония (ИК=1.02), апноэ (ИК=1,0), открытый артериальный проток (ИК=1,08).

Список литературы:

1. Научно-практическая программа «Бронхолёгочная дисплазия»,– М.: оригинал-макет 1. 2012.– 88 с.
2. Овсянников Д. Ю., Антонов А. Г., Ионов О. В. и др. Проект протокола по диагностике, профилактике и лечению бронхолёгочной дисплазии // Неонатология, 2014.– № 1.– С. 161–175.
3. Сенаторова А. С., Логвинова А. Л., Черненко Л. К., Мурадов А. Р. Бронхолёгочная дисплазия // Здоровье Украины, 2011.– № 1(16).– С. 2–7.
4. Панов П. В., Панова Л. Д., Ярукова Е. В., Ахмадеева Е. Н. Прогностические факторы риска формирования бронхолегочной дисплазии у недоношенных младенцев // Практическая медицина, 2016. № 31953.– С. 45–53.
5. Лебедева О. В., Чикина Т. А. Прогнозирование течения респираторного дистресс-синдрома у глубоко недоношенных новорожденных // Доктор. Ру. 2014.– № 3(91).– С. 7–14.
6. Бойцова Е. В., Запевалова Е. Ю., Овсянников Д. Ю. Респираторные, неврологические и структурно-функциональные последствия бронхолёгочной дисплазии у детей и взрослых // Неонатология, 2014.– № 1 (3).– С. 71–79.
7. Bhandari A., Bhandari V. Bronochopulmonary dusplasia: an uptate –İndian J. Pediatr. 2007.– 74.– P. 73–77.
8. Greenough A. Bronchopulmonoru dusplasia: Long term fallow up. Pediatr. Resp. Rev. 2006.– 7 p.– P. 189–191.
9. Cutr E., Chiassion D. Chronik Lung disease after premature birth. N. Enql. J. Med. 2008.– 358.– P. 743–745.
10. Theband B., Lacare Masmonteil T. If your placenta doesn't have it either. The "Vascular Hypothesis" of bronchopulmonary dysplasia strats in utero İ. J. Pediatr. 2010.– № 156.– P. 521–523.

https://doi.org/10.29013/ELBLS-20-4-15-28

Ke Ophelia,
Cate School Class of 2022,
Advised by Dr. Pingzhang Wang of Peking University
and Ivy Mind Analytics
E-mail: ophelia_ke@cate.org

INTEGRATED ANALYSIS OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNP) SITES AND MUTATIONS IN THE CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR (CFTR) GENE

Abstract. CF (Cystic Fibrosis) is a genetic health condition that affects a person's lungs and digestive system, which affects more than 70.000 people worldwide. It is characterized by a faulty protein (CFTR) that affects the body's cells, tissues, and glands which produce mucus and sweat. The research was done to determine and develop an integrated analysis of the SNP (Single Nucleotide Polymorphisms) sites, or a type of genetic variation representing a difference in a nucleotide, found in the CFTR gene.

Objectives: The purpose of this study is to identify SNP sites and mutations in the CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) gene that might cause Cystic Fibrosis and various strains of cancer. It revealed potential SNPs, which can help medical professionals set early diagnosis and risk evaluation for CF patients. Other information that was focused on also includes the structure of the CFTR gene, the distribution of its exome variant functions, the relationship with types of cancer and so on. If these results are studied and analyzed to a further extent, they may reveal a new method for the diagnosis, treatment, and hopefully, a cure for CF.

Method: Using the genetic data collected by various institutions such as the National Center for Biotechnology Information (NCBI), genetic information in both afflicted and healthy individuals was downloaded. Then, by using the Macintosh operating system (Terminal), SNP sites were extracted into a VCF file. With this information, integrated and statistical analysis was utilized to pinpoint how they affect the phenotypic variabilities of CF through the use of the online tool, wANNOVAR. This information was then projected onto visual means through the use of software systems such as R and RStudio and Microsoft Excel. Online means like the Genome Browser were also utilized.

Results: The integrated analysis identified key information on this genetic disorder such as the distribution of SNP functional classes, the frequency of CF occurrences, the associated types of cancers, the structure of the CFTR gene as well as the gene product, SIFT Score as well as a plethora of other information across the downloaded and extracted exome dataset.

Conclusion: Through an accurate and thorough analysis on single nucleotide polymorphisms (SNPs) and mutations in the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene, the effects, influence, and types of SNP sites and somatic mutations were identified. This can further

our understanding of health conditions associated with the gene and its product (CFTR protein) and enable medical communities to take the necessary action to overcome these disorders.

Keywords: Catalogue of Somatic Mutations in Cancer (COSMIC), CFTR Modulator Therapy, ClinVar, Cystic Fibrosis (CF), Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), Exome, Gene Therapy, Genome, National Center for Biotechnology Information (NCBI), Single Nucleotide Polymorphism (SNP), Single Nucleotide Polymorphism Database (dbSNP), SNP Functional Classes, Sorting Intolerant from Tolerant (SIFT Score and Prediction), Single Nucleotide Variant (SNV)

I. Introduction

Cystic Fibrosis (CF) is a rare and inherited genetic disorder whose symptoms often include chronic cough, lung infections, and shortness of breath. It is caused by defects in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Situated on the seventh chromosome, the gene consists of twenty-seven exons of DNA and codes for 1,480 amino acids. Its product is the CFTR protein, which regulates the chloride ion content of epithelial cells that line the nasal cavity, lungs, and stomach by acting as a channel across the membrane [1; 2; 3].

The CFTR protein acts as a channel carrying chloride ions into and out of human body cells, which aids in the movement of water in tissues. This is crucial for the production of mucus, a substance that lubricates and protects the lining of the respiratory, digestive, and reproductive systems. The CFTR protein also controls the functions of other channels, such as the ones which move sodium ions across cell membranes. These are necessary for organs like the lungs and pancreas to work. When these chloride ions cannot leave the cell, water is kept through osmosis, which causes the production of more viscous fluids [4].

The symptoms of CF depend on which organs are affected and the severity of the condition. The most serious and common complications in regard to cystic fibrosis are those of pulmonary or respiratory problems, which may include serious lung infections. Patients diagnosed with CF often also have problems maintaining good nutrition, as they find it difficult to absorb the nutrients from food, delaying growth. [5] Until late into the twentieth century, few people diagnosed with CF lived beyond childhood. While improvements in medical care have succeeded in increasing life expectancy, there still exists no cure for it.

There exist more than 1.200 discovered faults on the CFTR gene. Of this number, the most frequent mutation remains the result of the deletion of a single amino acid at position 508 on the CFTR protein. It is also referred to as ΔF508 and accounts for approximately seventy percent of CF cases. Other mutations to the CFTR gene cause changes to the protein's structure, stability, or production, ultimately inhibiting the successful regulation of chloride ions in epithelial cells [6].

Single nucleotide polymorphisms, more commonly known as SNPs, are the most common type of genetic variation among people, each representing a difference in a nucleotide, or a single DNA building block. These variations are most frequently found in the DNA between different genes. Nowadays, the scientific community uses SNPs as biological markers because they help pinpoint which genes are associated with the disease. However, when SNPs occur within a gene or a regulatory region near the gene, they can actually be the cause of the affliction by affecting the function of that gene [7].

II. Procedure

A. Materials

This study was mainly based on bioinformatics analysis and involves the usage and integration of publicly available datasets, tools, software, as well as other online resources. The various tools used to complete this study include Macintosh operating system (MacOS); Microsoft Excel; dbSNP database,

an online database for single-nucleotide variations; wANNOVAR, which was used to annotate functional consequences of genetic variation from high-throughput sequencing data. Other online tools that were used were UCSC Genome Browser; National Cancer Institute GDC Data Portal; and cBioPortal.

B. General Overview

From the online genetic database, dbSNP, data on Cystic Fibrosis patients and healthy controls from around the world were downloaded in the form of a GZ file. After decompressing the file with the Macintosh Operating System (MacOS), all of the SNP sites on the CFTR gene were extracted. Then, individual analysis was performed on this information with the web-based tool wANNOVAR, resulting in an analysis of the genome and exome of the CFTR gene. With this information gathered, a more integrated study was performed, furthered by the extensive use of visual representations as well as detailed explanations.

C. Online Gene Database (Single Nucleotide Polymorphism Database)

The Single Nucleotide Polymorphism Database (dbSNP) is a free public web-based archive that records genetic variation within and across different species. It was developed by the National Center for Biotechnology Information (NCBI) in collaboration with the National Human Genome Research Institute (NHGRI). It contains human single nucleotide variations (SNV), microsatellites, small- scale insertions and deletions, along with information on publication, population frequency, molecular consequences, as well as genomic and RefSeq mapping information for both common variations and clinical mutations.

D. Extraction of SNP Sites on MacOS

After decompressing the file with the Macintosh Operating System (MacOS), all of the SNP sites on the CFTR gene were extracted using the command: grep CFTR00-All.vcf >CFTR.result.vcf.

E. Annotation of Genetic Variation on wANNOVAR

By default, it performs "individual analysis" on the VCF file to help find the genes which cause the disease as well as various other online sources. The resulting files are split into information on the exome and genome. It can be read by Microsoft Excel and includes a plethora of data on the CFTR gene. [8][9]

F. Using Terminal to Reveal SNP Exonic Functions

By using the Terminal function in the Macintosh Operating System (MacOS), the unique parts were classified and sorted from the 2,481 individual data into 11 exonic variants: frameshift deletion, frameshift insertion, frameshift substitution, non-frameshift deletion, non-frameshift insertion, non-frameshift substitution, nonsynonymous SNV, start-loss, stopgain, stop-loss, and synonymous SNV. After that, the distribution was represented visually as a pie chart.

G. Using Terminal to Reveal SNP Genomic Functions

Due to significant time restraints, most data analysis was restricted to information on the exome, but there is still Similar to the procedure above, genomic functions were sorted from the 33,355 individual data into 9 genomic variants: UTR3 (3 Prime Untranslated Region), UTR5 (5 Prime Untranslated Region), downstream, exonic, exonic splicing, intergenic, intronic, splicing, and upstream. The data was then represented visually as a pie chart.

H. UCSC Genome Browser: Human GRCh37/hg19

By entering a specific range of the CFTR gene onto the UCSC Genome Browser: Human GRCh37/hg19, a detailed picture of the gene's structure was created and downloaded.

I. Frequency of CF Occurrence (ClinVar_DIS)

Through the integrated analysis wANNOVAR performed, information from ClinVar, a database which aggregates information about genomic variation and its relationship to human health, was accessed. By bridging this crucial gap, the frequency of CF occurrence in relations to SNP sites on the CFTR gene can be concluded with a chart.

J. SIFT Score Prediction

SIFT, short for Sorting Intolerant From Tolerant, is a system that can give accurate speculation as to whether or not an amino acid substitution will affect protein function. This way, users, particularly those in the medical community, can prioritize substitutions for further study. The SIFT score ranges from 0.0 (deleterious/harmful) to 1.0 (benign/tolerated). By looking into this, the connection between the portion of the gene and its effect was drawn.

K. COSMIC_ID, COSMIC_DIS:

COSMIC, or the Catalogue of Somatic Mutations in Cancer, is the world's forefront online database for the investigation on the impact of somatic mutations in human cancer. After directing and pinpointing the database on the CFTR gene, figures and statistics were downloaded for the benefit of this project.

L. NCI GDC Data Portal and cBioPortal

To venture further into the topic of mutations of the CFTR gene and its connection to human cancer, the NCI GDC Data Portal, a data-driven platform and cBioPortal for Cancer Genomics, a software was accessed. Both provided alteration frequency as well as frequent somatic mutations.

M. UniProt Knowledgebase

The UniProtKB (abbrev. for Knowledgebase) is the world's central resource on the functional information on proteins, the information of which is derived from the current research literature. With this, information on the CFTR gene was gathered and clarified for the purpose of providing a more thorough analysis.

III. Results

N. Distribution of Exome Variant Functions

Table 1. – A portion of the resulting CSV file (displayed on Microsoft Excel)

Start	End	Ref	Alt	ExonicFunc.refGene
117120149	117120149	A	G	startloss
117120150	117120150	T	A	startloss
117120150	117120150	T	C	startloss
117120151	117120151	G	A	startloss
117120151	117120151	G	T	startloss
117120152	117120152	C	T	stopgain
117120158	117120158	T	G	nonsynonymous SNV
117120159	117120159	C	A	stopgain
117120159	117120159	C	T	nonsynonymous SNV
117120160	117120160	G	T	synonymous SNV

*That shows the location of the SNP site along with the exome variant associated with it. The highlighted column (**ExonicFunc.refGene**) includes the following: frameshift deletion, frameshift insertion, frameshift substitution, non-frameshift deletion, non-frameshift insertion, non-frameshift substitution, start- loss, stop-gain, stop-loss, nonsynonymous SNV, and synonymous SNV*

As referenced above in (Table 1), wANNOVAR sorted the types of SNPs on the CFTR gene into eleven types. SNP sites can fall within the coding sequences, non-coding regions of the gene, or in the intergenic zones. Sites that are located within the coding regions are either synonymous and nonsynonymous SNPs. Synonymous mutations are fairly common, but since they do not affect the amino acid sequence of a protein, they are not noticed (**non-frameshift insertion**, **deletion**, or **substitution**)

On the other hand, in a nonsynonymous mutation, this is not the case. There is commonly an insertion or deletion of one nucleotide in the coding sequence during the process of transcription. The single missing or added nucleotide causes a frameshift mutation (**frameshift insertion, deletion,** or

substitution) which proceeds to throw off the entire reading frame of the amino acid sequence, ultimately mixing up the codons. These comprise of nonsense and missense mutations.

A missense mutation is a change in one base pair on the DNA that will ultimately result in the substitution of one amino acid for another in the protein product. By doing so, it alters a codon and creates a completely different protein. A nonsense mutation is also a change in one DNA base pair and includes **stop-gain**, **start-loss**, and **start-gain**. Stop-gain refers to a mutation that results in a premature termination codon, signaling the end of translation, while stop-loss is a mutation in the original termination codon, resulting in an abnormal extension of the protein's carboxyl terminus. Start-gain is defined as a point mutation in the transcript's AUG codon, which also serves as an initiation site for the gene product.

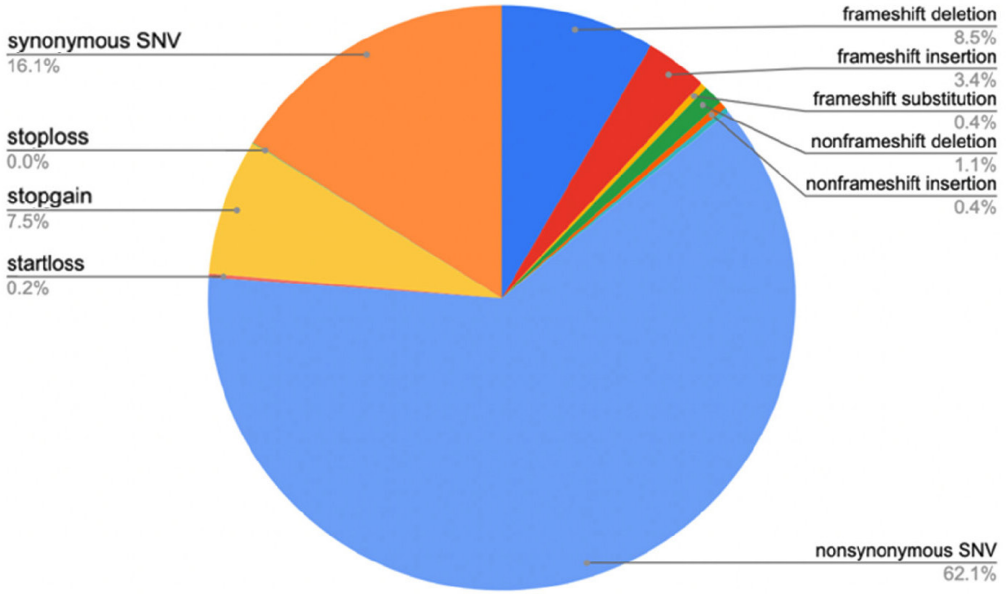


Figure 1. A pie chart depicting the distribution of different SNP exome variant functions on the CFTR gene in humans around the world. They include frameshift deletion, frameshift, insertion, frameshift substitution, nonframeshift deletion, nonframeshift insertion, nonframeshift substitution, nonsynonymous SNV, start-loss, stop-gain, stoploss, and synonymous SNV

Finally, a single-nucleotide variant (**nonsynonymous or synonymous SNV**) is referred to as a variant in a single nucleotide with no limitations on frequency. They are not the same as single- nucleotide polymorphisms due to the fact that when an SNV is detected in a single sample, it can potentially be an SNP. However, this cannot be ascertained given that this variation is only from one organism.

While all the types of SNPs on the CFTR bring about genetic variation in a human population, their effects and frequency certainly differ from each other. By using the Macintosh Operating System (MacOS) to determine the count of each unique type, this distribution can be visually represented in the form of a pie chart, enabling a clearer understanding of these SNP sites.

As shown in (Figure 1), nonsynonymous SNV remains the most common type of SNP on the CFTR gene with an overwhelming majority, while stop-loss exists as the rarest class with only one identified case. Nonetheless, understanding the distribution of the exome variants of SNPs can allow scientists to pinpoint the types of CF that are more common as well as the causes behind them.

O. Distribution of Genome Variant Functions

This research project focused on the exome summary produced by wANNOVAR instead of covering every single detail of the CFTR gene. However,

as shown in (Figure 2), the distribution of genomic variant functions was included to give a general overview of the information that was given in the genome summary. As shown above, the most common form of genomic variant on the CFTR gene is **intronic**, where the variant overlaps with an intron.

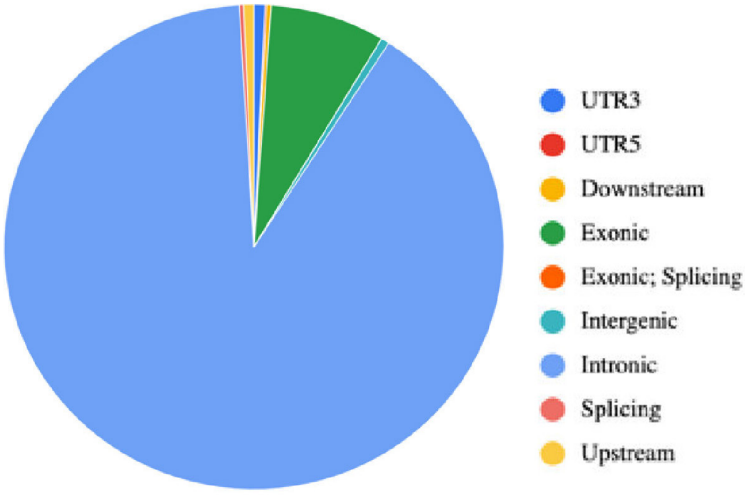


Figure 2. A pie chart depicting the distribution of different SNP genomic variant functions on the CFTR gene in humans around the world. They include intronic, exonic, intergenic, UTR3, UTR5, downstream, upstream, exonic splicing, and splicing

It is closely followed by **exonic**, where the variant overlaps a coding region; **UTR3**, where it overlaps a 3' untranslated region; upstream, where it overlaps 1-kb region upstream of the transcription start site; and **intergenic**, where the variant is within the intergenic region. The rest of the genomic variants are less common in the world.

P. Structure of the CFTR Gene and Protein

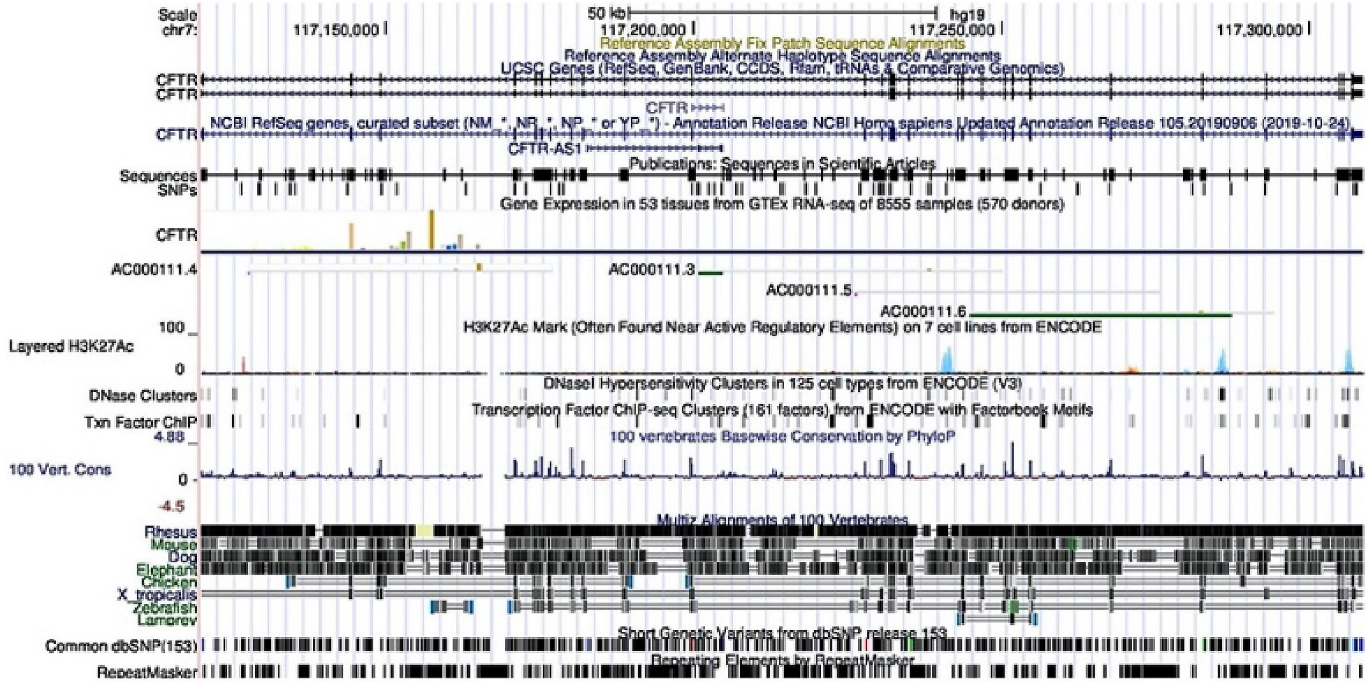


Figure 3. The structure and location of the CFTR gene on Chromosome 7 is shown through the use of the UCSC Genome Browser (Human Feb. 2009; GRC37/hg19 Assembly)

The CFTR (Cystic Fibrosis Transmembrane Conductance Regulator) gene is located in the seventh human chromosome. It provides instructions for making a protein called the cystic fibrosis transmembrane conductance regulator, which consists of 1.480 amino acids. When altered, health conditions such as Congenital Bilateral Absence of the Vas Deferens (CBAVD), Cystic Fibrosis (CF), Hereditary Pancreatitis, and others arise.

The cytogenetic location of this gene, shown in (Figure 3 is q31.2, referring to the long (q) arm of the chromosome at the position 31.2. Its molecular location is also depicted in the image, ranging from the base pairs 117,120,017 to 117,308,718 (188,702 base pairs) [10].

The CFTR gene codes for an ATP binding cassette (or ABC) transporter-class ion channel protein. It conducts the transportation of chloride ions across epithelial cell membranes. This protein also comprises two six-span units, each of which is membrane-bound and attached to a nuclear binding factor for adenosine triphosphate (ATP). Between these two regions, there is an R- domain, consisting of several charged amino acids. This is an entirely unique feature of the CFTR protein within the ABC superfamily [11; 12].

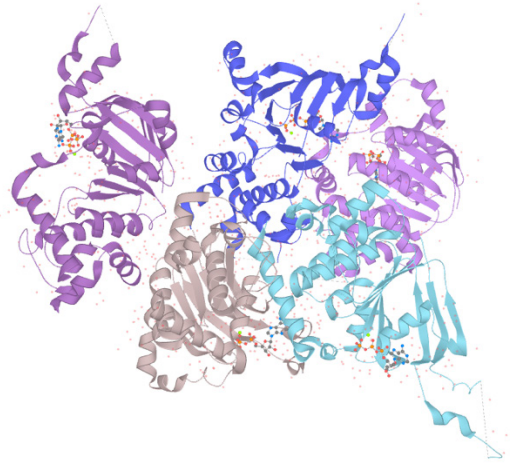


Figure 4. A picture depiecting the shape and structure of the CFTR protein

Q. Probability of SNP Sites in Relations to World Demographics:

Table 2. – The chart that shows the probability of SNP sites in the CFTR gene in relations to the demographics of the world

AAChange.refGene	1000G_ALL	1000G_AFR	1000G_AMR	1000G_EAS	1000G_EUR	1000G_SAS	ExAC_Freq	ExAC_AFR	ExAC_AMR	ExAC_EAS	ExAC_FIN	ExAC_NFE	ExAC_OTH	ExAC_SAS	ESP6500si_ALL	ESP6500si_AA	ESP6500si_EA
CFTR:NM_000492:exon1:c.T4G:p.L15L																	
CFTR:NM_000492:exon1:c.T49C:p.F17L																	
CFTR:NM_000492:exon1:c.51delC:p.F17Lfs*7																	
CFTR:NM_000492:exon1:c.A53G:p.S18G							2.48E-05	0	0	0	0	4.51E-05	0	0			
CFTR:NM_000492:exon2:c.G56A:p.W19X																	
CFTR:NM_000492:exon2:c.G57T:p.W19C																	
CFTR:NM_000492:exon2:c.A58T:p.T20S																	
CFTR:NM_000492:exon2:c.C59G:p.T20S																	
CFTR:NM_000492:exon2:c.C60T:p.T20T	0.0002	0.0008					8.29E-06	9.66E-05	0	0	0	0	0	0			
CFTR:NM_000492:exon2:c.G62T:p.R21I							2.48E-05	0	0	0	0.0003	1.51E-05	0	0			
CFTR:NM_000492:exon2:c.A63G:p.R21R																	
CFTR:NM_000492:exon2:c.A66G:p.P22P							1.66E-05	0.0002	0	0	0	0	0	0			
CFTR:NM_000492:exon2:c.G72A:p.L24L																	
CFTR:NM_000492:exon2:c.G72C:p.L24F																	
CFTR:NM_000492:exon2:c.G74A:p.R25K																	
CFTR:NM_000492:exon2:c.A76G:p.K26E							1.66E-05	0	0	0	0	3.01E-05	0	0			
CFTR:NM_000492:exon2:c.A78C:p.K26N																	
CFTR:NM_000492:exon2:c.79delG:p.G27Dfs*63																	
CFTR:NM_000492:exon2:c.G79A:p.G27R																	
CFTR:NM_000492:exon2:c.G79C:p.G27R																	
CFTR:NM_000492:exon2:c.G79T:p.G27X							8.27E-06	0	0	0	0	1.51E-05	0	0			
CFTR:NM_000492:exon2:c.G80A:p.G27E																	
CFTR:NM_000492:exon2:c.C84T:p.Y28Y																	
CFTR:NM_000492:exon2:c.G86C:p.R29T							8.27E-06	0	0	0.0001	0	0	0	0			
CFTR:NM_000492:exon2:c.C88T:p.Q30X																	
CFTR:NM_000492:exon2:c.C91A:p.R31S																	
CFTR:NM_000492:exon2:c.C91T:p.R31C	0.0014	0.0008		0.002	0.003	0.001	0.0017	0.0004	0.0003	0.0065	0	0.002	0	0.0003	0.0008	0.0005	0.0009
CFTR:NM_000492:exon2:c.G92A:p.R31H							4.14E-05	0	0.0003	0.0001	0	0	0	6.02E-05	0.0001		0.0003
CFTR:NM_000492:exon2:c.G92T:p.R31L							1.65E-05	9.63E-05	0	0	0	1.51E-05	0	0			
CFTR:NM_000492:exon2:c.C93A:p.R31R							2.48E-05	0	0.0003	0	0	0	0	0			
CFTR:NM_000492:exon2:c.C94A:p.L32M							8.27E-06	0	0	0	0	0	0	6.02E-05			

The source of the region highlighted yellow is 1000 Genomes; the region highlighted pink is ExAC Browser; the region highlighted green is the Exome Variant Server (ALL: All; AFR: African; AMR: American; EAS: Eastern; EUR: European; SAS: South Asian; FIN: Finnish; NFE: Non-Finnish European; OTH: Other)

Table 2 depicts the result of wANNOVAR's filter-based annotation. It gathers information on the cataloging of genetic variation among different ethnicities, races, and nationalities from around the world to display them all in the above fashion. Its purpose is to establish the frequency of variants in whole-genome data. The 1000 Genomes (Represented by 1000G) dataset provides allele frequencies in six populations that are whole-genome variants. The Exome Aggregation Consortium (ExAC) is represented by a group of investigators who collect and systematize exome-sequencing data from a variety of large-scale projects. The Exome Sequencing Project (ESP) is an exome-sequencing project that is funded by the National Heart, Lung, and Blood Institute (NHLBI). It identifies genetic variants in exonic regions from over 6.000 individuals, including healthy controls as well as those with different health conditions.

Right now, there are more than 10 million known Americans who are carriers of one mutation of the CFTR gene, which amounts to a total of 30.000 CF patients. The chances of being a carrier of one CFTR mutation or being afflicted with CF, which is caused by two CFTR disease-causing mutations, depends on race and ethnicity. And although it is not shown in Figure 8, the most affected group includes Caucasians of northern European ancestry (British, Scandinavians, French, certain Eastern Europeans). On the other hand, the disease is considerably more infrequent in other ethnicities, affecting about 1 in 1 in 100.000 Asian-Americans and 17.000 African-Americans.

According to a study by the John Hopkins University, the risk of particular ethnicities carrying the faulty CFTR gene is 1 in 29 for Caucasians; 1 in 46 for Hispanics; 1 in 65 for African Americans; and 1 in 90 for Asians. Thus, it is clear that given information and resources on the genetic variation in different ethnic populations, Caucasians remain the most affected group.

R. Health Conditions Caused by CFTR Defects (ClinVar_DIS):

As shown in the pie chart below, more than half of the faults in the CFTR gene have no effect or are not specified. However, the health condition most clearly associated with this is Cystic Fibrosis. Patients with CF experience issues with their respiratory, digestive, and reproductive systems. Although it is not shown in Figure 5, most men with CF also have a congenital bilateral absence of the vas deferens (CBAVD), a condition in which the Vas Deferens or the tubes that carry sperm are clogged with mucus, effectively sterilizing most patients. Other health conditions caused by the faulty CFTR gene include hereditary and idiopathic pancreatitis, as well as sweat chloride elevation without CF.

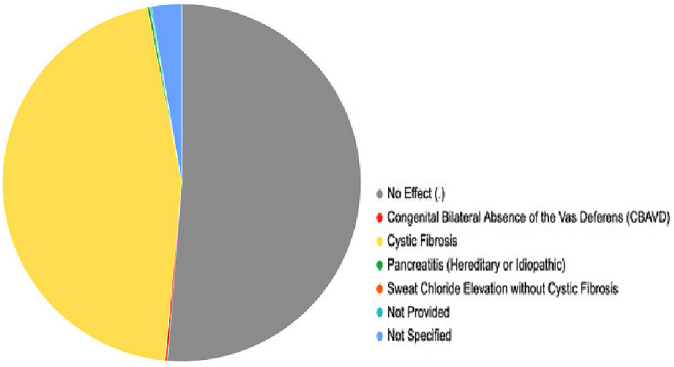


Figure 5. Frequensy of CF Occurrence (Clin Var_DIS). A pie chart that depicts the frequency of CF occurrence in SNP sites on the CFTR gene, along with a plethora of other health conditions, which include Congenital Bilateral Absence of the Vas Deferens (CBAVD), Pancreatitis (Both hereditary and idiopathic), and Sweat Chloride Elevation without Cystic Fibrosis

S. SIFT Score Prediction

Single nucleotide polymorphism (SNP) studies and random mutagenesis projects that determine amino acid substitutions in protein-coding regions, as each substitution can potentially affect the function of the protein. Other than that, the SIFT score was also scrutinized during the research process. It predicts whether an amino acid substitution will af-

fect protein function, making it an invaluable source for the classification of benign and deleterious effects.

A SIFT score predicts whether an amino acid substitution affects protein function. The SIFT score ranges from 0.0 (deleterious) to 1.0 (tolerated): In the range from 0.0 to 0.05, variants are considered deleterious, where those with scores closer to 0.0 are more confidently predicted to be deleterious. In the range from 0.05 to 1.0, variants are predicted to be tolerated (benign), where those with scores very close to 1.0 are more confidently predicted to be tolerated.

Table 3. – A chart showing the SIFT score associated with the given health condition, revealing whether or not it is pathogenic

ClinVar_SIG	ClinVar_DIS	SIFT_ score	SIFT_converted rankscore	SIFT_ pred
Pathogenic	Cystic_fibrosis	0	0.912	D
not provided\ x2cLikcly patho-genic	Cystic_fibrosis\x2cCystic_fi-brosis	0	0.912	D
not provided\ x2cLikely patho-genic	Cystic_fibrosis\x2cCystic_fi-brosis	0	0.912	D
not provided\x2c-not provided	Cystic_fibrosis\x2cCystic_fi-brosis	0	0.912	D
not provided\x2c-not provided	Cystic_fibrosis\x2cCyslic_fi-brosis	0	0.912	D
not provided	Cystic_fibrosis			
		0.017	0.512	D
Pathogenic	Cystic_fibrosis			
Pathogenic	Cystic_fibrosis	0.004	0.654	D
other	Cystic_fibrosis	0	0.912	D
		0.019	0.501	D
Pathogcnic | Patho-genic	Cystic_fibrosis | Hereditary_ pancreatitis	0.005	0.632	D

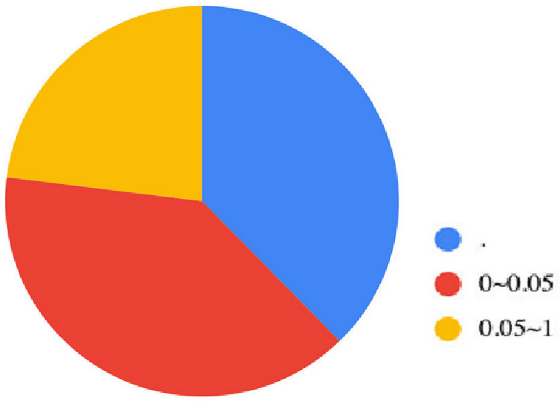


Figure 6. A pie chart showing the distribution of different SIFT score ranges associated with the CFTR gene

Referring to (Table 3), SNP sites whose SIFT scores are closer to 0.0 are more likely to be considered "pathogenic" and are associated with Cystic Fibrosis, CBAVD, and pancreatitis, while those with SIFT scores closer to 1.0 are considered "not provided." According to Figure 6, the majority of individuals in the dataset have a SIFT Score of between 0 to 0.05, meaning that it is considered pathogenic or deleterious. By summarizing information such as that, every fault and blemish in the CFTR gene can be accounted for accordingly, which will allow for reliable and accurate predictions.

T. CFTR Gene and Human Cancer

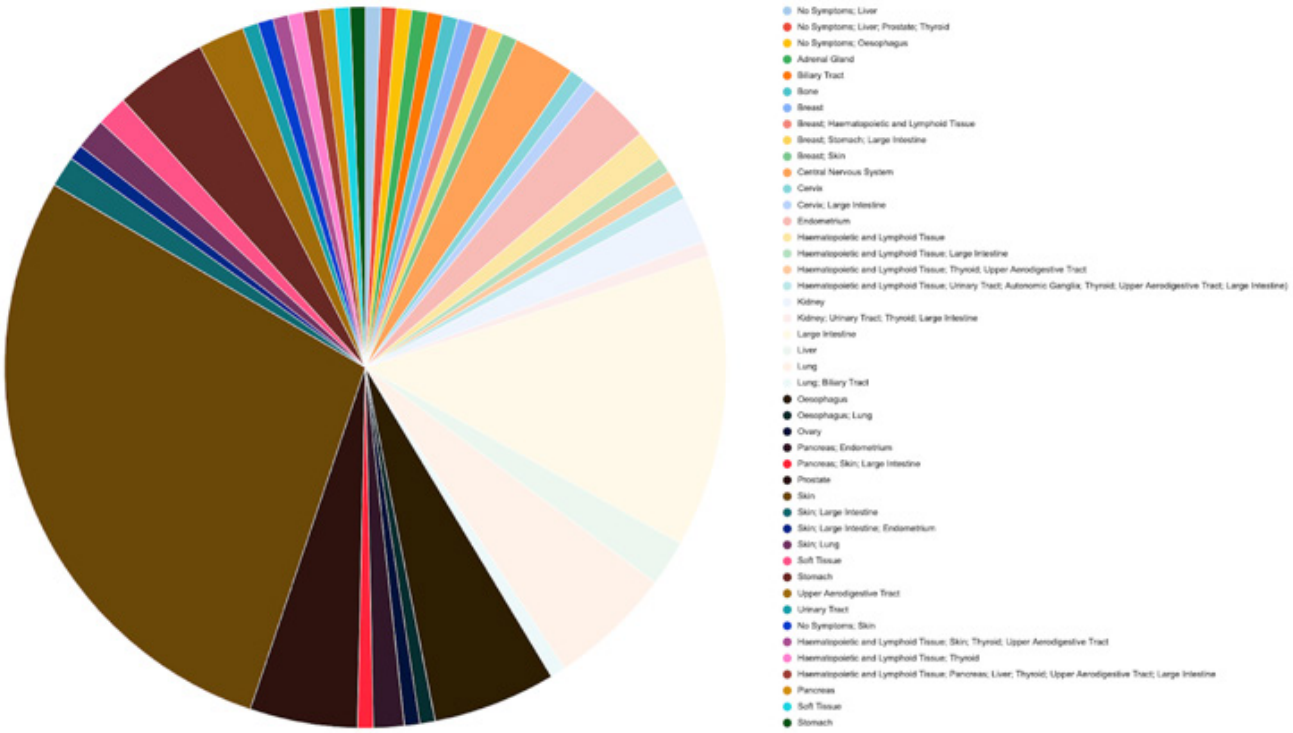


Figure 7. Types of Cancers Associated with SNPs in tre CTFR Gene. A pie chart showing the types of cancers associated with SNP sites in the CFTR gene, including those that cause multiple human cancers (two or more). "No Symptoms" was not included in the chart above

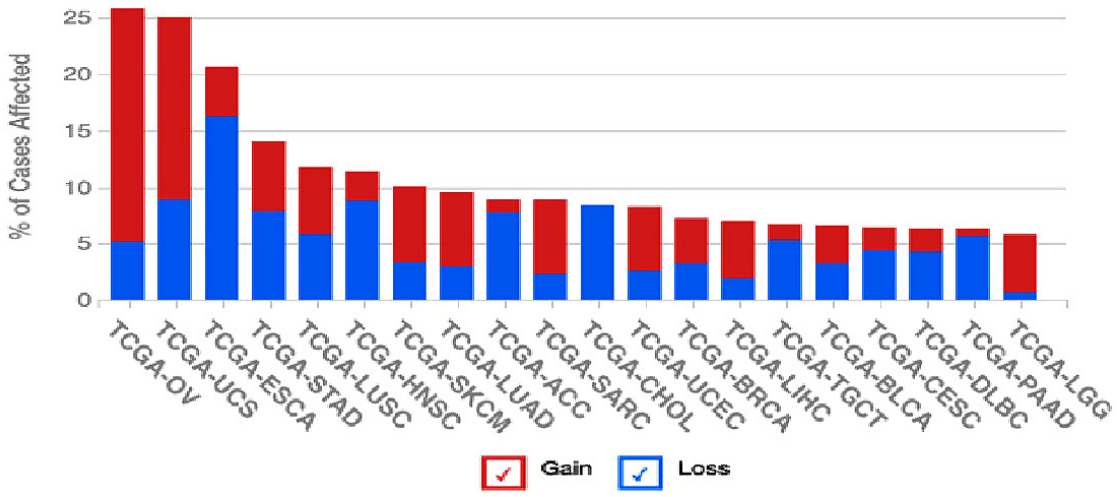
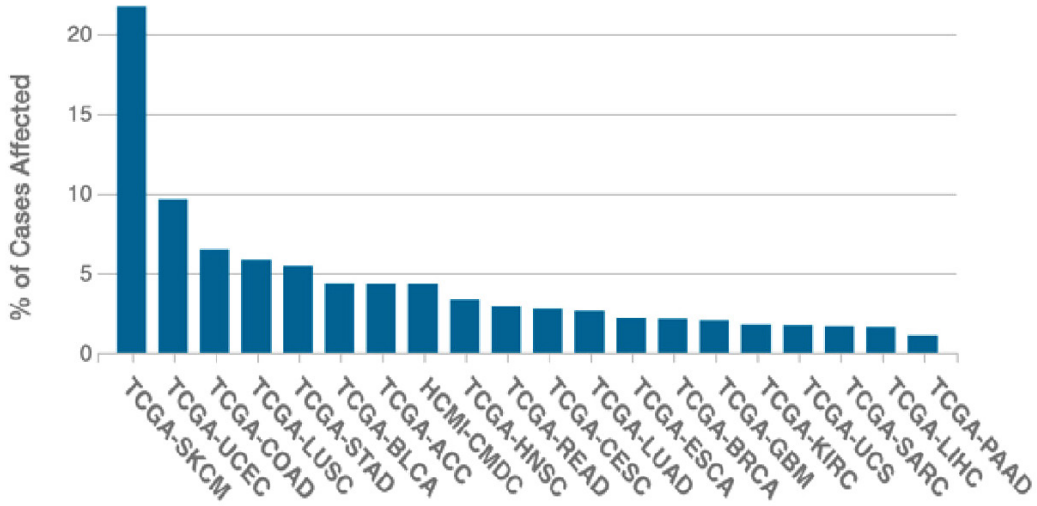


Figure 8. CNV Distribution (Left)

Although defective CFTR is commonly associated with CF, a common genetic disorder in the Caucasian population, there is accumulating evidence that suggests the role of CFTR faults in various cancers, particularly gastroenterological cancers such as pancreatic cancer and colon cancer (of the large intestine). Figure 7 (Above) does not include "No Symptoms" due to the overwhelming majority. But common human cancers associated with the CFTR gene include skin cancer, while the rest are similarly frequent.



Figures 9. Cancer Distribution (Right)

Figures 8 (Left) and 9 (Right): Two bar graphs created with data from the National Cancer Institute GDC Data Portal. (Left) **CNV (Copy Number Variation) Distribution** (OV: Ovarian; UCS: Uterine; ESCA: Esophagea; STAD: Stomach Adenocarcinoma; LUSC: Lung Squamous Cell Carcinoma; HNSC: Head-Neck Squamous Cell Carcinoma; SKCM: Skin Cutaneous Melanoma; LUAD: Lung Adenocarcinoma; ACC: Adrenocortical Carcinoma; SARC: Sarcoma; CHOL: Cholangiocarcinoma; UCEC: Uterine Corpus Endometrial Carcinoma; BRCA: Breast Invasive Carcinoma; LIHC: Liver Hepatocellular Carcinoma; TGCT: Testicular Germ Cell Tumors; BLCA: Bladder Urothelial Carcinoma; CESC: Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma; DLBC: Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; PAAD: Pancreatic Adenocarcinoma; LGG: Low Grade Glioma) (Right) **Cancer Distribution** (SKCM: Skin Cutaneous Melanoma; UCEC: Uterine Corpus Endometrial Carcinoma; COAD: Colon Adenocarcinoma; LUSC: Lung Squamous Cell Carcinoma; STAD: Stomach Adenocarcinoma; BLCA: Bladder Urothelial Carcinoma; ACC: Adrenocortical Carcinoma; HNSC: Head-Neck Squamous Cell Carcinoma; READ: Rectum Adenocarcinoma; CESC: Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma; LUAD: Lung Adenocarcinoma; ESCA: Esophagea; BRCA: Breast Invasive Carcinoma; GBM: Glioblastoma; KIRC: Kidney Renal Clear Cell Carcinoma; UCS: Uterine; SARC: Sarcoma; LIHC: Liver Hepatocellular Carcinoma; PAAD: Pancreatic Adenocarcinoma).

Copy number variation (CNV), defined as large-scale gains and losses of DNA fragments, forms another one of the major classes of genetic variation. After the Human Genome Project, it became clear that the human genome goes through gains and losses of DNA material. The extent to which CNVs are attributed to certain human afflictions remains unknown. However, it is an established fact that certain cancers are associated with heightened copy numbers of specific genes. According to (Figure 7), ovarian cancer, uterine cancer, and esophageal cancer each take up 25.81%,

25%, and 20.65%, respectively of total CNV distribution. In (Figure 8), the distribution of human cancer types caused by SNPs in the CFTR gene. It reveals that Skin Cutaneous Melanoma (SKCM) is the prevalent form of cancer, covering 21.75% of affected cases.

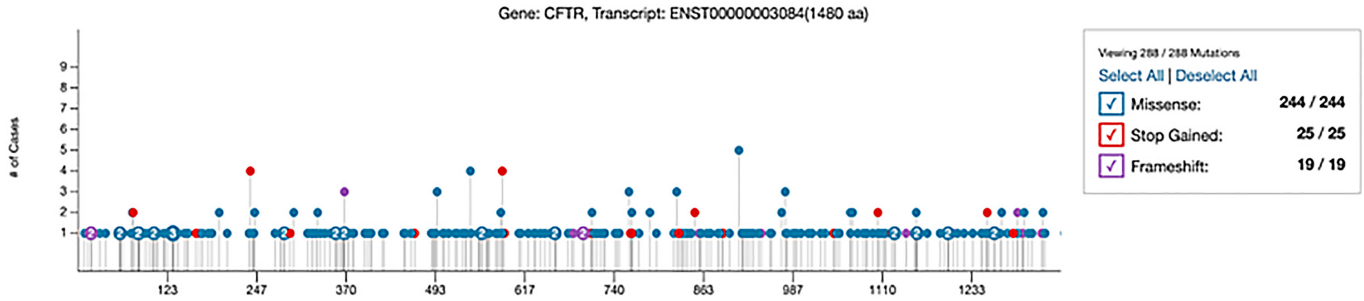


Figure 10. Transcript of the CFTR Protein provided by the National Cancer Institute GDC Data Portal showing the number of cases for each type of mutation.

When referring to (Figure 9), the points at which particular strains of cancer are caused are brought up, along with the number of cases and type of mutation it is associated with. According to the protein transcript, chr 7: g.117642527C>T appears the most. It is a synonymous mutation that involves the substitution of a specific base, making up approximately 0.62% of affected cases in CFTR (7 affected cases across the GDC). Other common cases include chr7: g.117603609C>T (0.44%), chr7: g.117530985G>A (0.35%), and chr7: g.117590426G>T (0.35%), all of which are substitution mutations. By accumulating and analyzing data on the specific type, effect, and frequency of somatic mutations, common causes can be identified for a more personalized and thorough health treatment.

Discussion

Ever since the discovery of the CFTR gene more than thirty years ago, the scientific and medical communities have been trying to find ways to alter and ultimately correct the mutations in the gene, specifically those that cause CF. Although progress was noticeably slower in the beginning, scientific breakthroughs in the past decade have accelerated advances in gene therapy, CFTR modulator therapies, along with other treatments. Through integrated analysis such as this very research project, the medical community can correctly classify specific instances, functional types, as well as detrimental effects of different SNPs. This will potentially aid in the development of treatment and therapies for disorders, cancers, and diseases associated with the gene.

Gene therapy is the process by which the correct version of the CFTR gene is to be positioned in a person's cells. Although the faulted copies of the gene still remain, the correct copy now allows cells to produce functioning CFTR proteins.

There are three main types of gene therapy: **Integrating**, **Non-integrating**, and **RNA Therapy**. First, in integrating gene therapy, a portion of DNA with the correct version of the CFTR gene is delivered to a patient's cells, which then stays within their genome.

Similarly, in non-integrating gene therapy, an accurate version of the CFTR gene is delivered to a patient's cells. However, unlike integrating gene therapy, this DNA remains separate from the person's genome. This way, the cell can still utilize the new copy to make normal CFTR proteins.

Both therapies previously mentioned involve the "donation" of DNA copies with the correct CFTR gene to a patient's cell, which allows it to make its own RNA copies through transcription. But recently, another approach has been advancing to the forefront of this field. It involves directly giving the cell these RNA copies. This is known as RNA therapy.

Other than the bright future for gene therapy, other possible paths for the treatment of CFTR- related diseases include CFTR modulator therapies.

They are designed to correct the faulty protein produced by the CFTR gene. Due to the fact that different mutations bring about different faults in the resulting protein, these treatments will only work for patients affected by specific mutations.

There are four CFTR modulators for people with specific mutations, including Ivacaftor (Kalydeco®), lumacaftor/ivacaftor (Orkambi®), tezacaftor/ivacaftor (Symdeko®), elexacaftor/tezacaftor/ivacaftor (TrikaftaTM). With the rise of easily acceptable datasets, user- friendly software programs, and a new generation of aspiring researchers, more potential CFTR modulators will be in circulation to address the underlying cause of the disease in people with other CF mutations, including this very research project.

Declaration

I hereby declare that the papers submitted are the research work and research results obtained under the guidance of my instructor/supervisor Dr. Pingzhang Wang of Ivy Mind Analytics. As far as I am aware, the paper does not contain research results that have been published or written by others, except for the content specifically listed in the reference and the acknowledgments. If there is anything wrong, I am willing to bear all related responsibilities.

Conclusion

Understanding, analyzing, and pinpointing some of the SNP sites along with their effects, type, and influence on the CFTR gene and protein can allow us to understand how it is correlated with diseases and disorders such as CF, CBAVD, pancreatitis, and cancer. By doing so, it can open up the medical community to more possibilities for prediction, prevention, diagnosis, and treatment. Such possibilities would be more specific to the needs of each individual, allowing for a rise in a combined accuracy and prediction of related health treatments.

Acknowledgment

I want to thank Dr. Pingzhang Wang for offering me his guidance and expertise while working on the S. T. Yau Biology Award. His classes in biostatistics and genetic biology has inspired the topic of my project. I would also like to thank Ms. Betty Wang for the opportunity to learn about this fascinating field through Dr. Wang's online classes.

References:

1. About Cystic Fibrosis. CF Foundation. URL: https://www.cff.org/What-is-CF/About-Cystic- Fibrosis
2. Cystic Fibrosis. URL: https://www.nhlbi.nih.gov/health-topics/cystic-fibrosis
3. Basics of the CFTR Protein. CF Foundation. URL: https://www.cff.org/Research/Research-Into- the-Disease/Restore-CFTR-Function/Basics-of-the-CFTR-Protein
4. National Institutes of Health. (2020, August 17). Cystic fibrosis – Genetics Home Reference – NIH. U. S. National Library of Medicine. URL: https://ghr.nlm.nih.gov/condition/cystic-fibrosis.c
5. The Embryo Project Encyclopedia. Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Gene | The Embryo Project Encyclopedia. URL: https://embryo.asu.edu/pages/cystic- fibrosis-transmembrane-conductance-regulator-cftr-gene
6. National Institutes of Health. What are single nucleotide polymorphisms (SNPs)? – Genetics Home Reference – NIH. U. S. National Library of Medicine. URL: https://ghr.nlm.nih.gov/primer/genomicresearch/snp.
7. Wang K., Li M., Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research, 38: e164, 2010.
8. Chang X., Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. J Med Genet. 2012. Jul;49 (7): 433–6.
9. National Institutes of Health. (2020, August 17). CFTR gene – Genetics Home Reference – NIH. U. S. National Library of Medicine. URL: https://ghr.nlm.nih.gov/gene/CFTR

10. Ng P. C. & Henikoff S. (2003, July 1). SIFT: Predicting amino acid changes that affect protein function. Nucleic acids research. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC168916
11. U. S. Department of Health and Human Services. Cystic Fibrosis. National Heart Lung and Blood Institute. URL: https://www.nhlbi.nih.gov/health-topics/cystic-fibrosis
12. SIFT score. Ionreporter.thermofisher.com. URL: https://ionreporter.thermofisher.com/ionreporter/help/G UID-2097F236-C8A2–4E67–862D-0FB5875979AC.html

Section 2. Medical science

https://doi.org/10.29013/ELBLS-20-4-29-36

Melnyk A. V.,
MD, PhD, Professor of Department of Biological and General chemistry,
National Pirogov Memorial Medical University (Vinnytsya)
Zaichko N. V.,
MD, PhD, Professor, Head of Department of Biological and General chemistry,
National Pirogov Memorial Medical University (Vinnytsya)
Palamarchuk I. V.,
Assistant of Department of Biological and General chemistry
National Pirogov Memorial Medical University (Vinnytsya)
Strutynska O. B.,
Assistant of Department of Biological and General chemistry
National Pirogov Memorial Medical University (Vinnytsya)
E-mail: anderneting@gmail.com

EFFECT OF H_2S METABOLISM MODULATORS ON THE LEVEL OF GALECTIN-3 IN AORTA, HEART AND KIDNEYS OF RATS WITH STREPTOZOTOCIN-INDUCED DIABETES MELLITUS

Abstract. Diabetes mellitus (DM) is often comorbid with developing angiopathy, cardiomyopathy and nephropathy. Immune-inflammatory activation of the endothelium, cardiac and renal fibrosis following chronic hyperglycemia can be linked to impaired production of hydrogen sulphide (H_2S) and galectin-3. The causal relation between these factors remains disputable. Aim of the research was to evaluate the connection between changes in H_2S system and the level of galectin-3 in the blood and tissues (aorta, heart and kidneys) of rats with streptozotocin-induced diabetes mellitus. It is established that streptozotocin-induced diabetes mellitus is associated with the increase in galectin-3, the decrease in H_2S, the decrease in the activity and expression of cystathionine-γ-liase (CSE) in blood, aorta, heart and kidney. The decline in the activity of H_2S / CSE system is accompanied by a significant increase in galectin-3 in blood and tissues which can accelerate the development of diabetes-associated organ dysfunction and angiopathy.

Keywords: galectin-3, hydrogen sulfide, diabetes mellitus.

Introduction. Diabetes mellitus (DM) is a global medical and social problem due to its high prevalence, early development of complications, high disability and mortality. According to the World Health Organization, the prevalence of diabetes in different countries vary from 4% to 8% [1; 2]. Among

severe and common complications of diabetes there are diabetic angiopathy, cardiomyopathy and nephropathy. The various pathobiochemical disorders (glycosylation of proteins, oxidative stress, inflammation, mitochondrial dysfunction) are involved in the pathogenesis of heart, vascular and renal disease in diabetes, among which the hyperproduction of profibrogenic mediators takes an important place [3; 4]. The literature analysis showed that galectin-3 is one of the sensitive biochemical marker of the development of endothelial dysfunction, myocardial and renal fibrosis in patients with diabetes [5; 6; 7].

Recently, the attention of scientists has been focused on the study of a biologically active molecule – hydrogen sulfide (H_2S). This molecule possesses the properties of an antiplatelet agent, anticoagulant, vasodilator, regulator of insulin secretion and glucose metabolism in the liver [3; 8; 9]. H_2S is an important nephro- and cardioprotector due to its antioxidant and anti-inflammatory activity and ability to stabilize cell membranes [3; 8; 9; 10]. It was shown that the use of H_2S donors in a streptozotocin (STZ) diabetic model has a cardio- and renoprotective effect [3]. Disorders of hydrogen sulfide (H_2S) and galectin-3 production may be potential factors in immunoinflammatory activation of the endothelium, myocardial and renal fibrosis in the setting of chronic hyperglycemia, yet, the relationship between these factors remains debatable.

Aim of the research was to evaluate the connection between changes in H_2S system and the level of galectin-3 in the blood and tissues (aorta, heart and kidneys) of rats with streptozotocin-induced diabetes mellitus.

Materials and methods. The experiments were conducted on 40 white non-linear male rats, initial body mass within 200–250 g, which were obtained from the vivarium of the National Pirogov Memorial Medical University (Vinnytsya). DM was modelled in three animal groups (n = 10) by single intraperitoneal (IP) injection of streptozotocin (Sigma, USA) which was freshly dissolved in 0.1 M citrate buffer (pH 4.5) at 40 mg/kg of rat mass. Control rats received IP equivalent volumes of 0,1 M citrate buffer (0.1 ml /100 g). Substances were injected after the animals had been deprived food for 24 hours. Two animal groups (Group 3 and 4) from day 3 to day 28 after streptozotocin injection were injected modulators of the H_2S system IP once a day as freshly prepared water solution at 0.1 ml per 100 g of rat mass. A CSE inhibitor D, L-propargylglycine (PPG) (Sigma, USA) was injected at 50 mg/kg, and H_2S donor NaHS (Sigma, USA) – at 3 mg/kg. The doses, routes and duration of H_2S modulators' delivery were taken from the literature and did not cause animal mortality [11; 12; 13]. Rats of the first group (control) and the second group after STZ-diabetes initiation were injected IP once a day with 0.15 M NaCl (0.1 ml per 100 g rat mass). The pathology signs were polydipsia, polyuria, and body mass loss. The rats with glycaemia levels >15 mmol/l were selected for the study.

All stages of experiments were carried out according to general ethical principles approved by European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Purposes (Strasbourg, 1986). Research protocols were approved by the Committee on Bioethics of National Pirogov Memorial Medical University (Vinnytsya).

Whole venous blood was obtained by decapitation and collected into sterile plastic Vacuette tubes (Greiner Bio-One, Austria) without anti-coagulant and with EDTA, when necessary. Plasma was obtained by centrifugation of the whole blood at 1500 g for 25 minutes at 18–22° C, plasma aliquots were collected into sterile plastic Eppendorf microtubes and stored at –20° C until further use.

To determine H_2S level in organs we used post-nuclear homogenates prepared in the following way: the myocardium and kidneys were washed with cold 1.15% KCl solution, cut up with scissors, homogenized in 0,01 M NaOH at the ratio of 1:5 (m/v) at 3000 rpm (teflon-glass). To 1 ml of homogenate we added 250 microliter 50% TCA, centrifuged for 15

minutes at 1200 g, then aliquots were collected into Eppendorf microtubes and the H_2S level was immediately determined in the supernatant.

For other biochemical studies, myocardium and kidney homogenates were prepared by the following procedure: tissue samples were homogenized in 0.25 M sucrose, 0.01 M Tris (pH 7.4) at the ratio of 1:5 (m/v) at 3000 rpm (teflon-glass), centrifuged for 30 min at 600 g at 4° C, then aliquots were collected into Eppendorf microtubes and stored at –20°C.

Levels of glucose in peripheral blood were measured using electronic glucometer Accu-Chek Active (Rouche Group, Germany). Levels of galectin-3 in blood plasma, postnuclear myocardium homogenates and kidney were determined by immunoassay using the Rat Galectin 3 (GAL-3) ELISA Kit (MyBiosource, Cat№ MBS2600708) according to the instruction. Standard solution concentrations were GAL-3 0.156; 0.312; 0.625; 1.25; 2.5; 5; 10 ng /ml. The detection was carried out using the STAT-FAX 303 analyser (USA) at 450 nm (differential filter 630 nm).

Levels of H_2S in aorta, myocardium and kidneys were determined spectrophotometrically by the reaction with N, N-dimethyl-para-phenylenediamine sulfate in the presence of $FeCl_3$ [14]. All manipulations were performed in hermetically sealed plastic tubes to prevent losses of H_2S. Sulfide ion content in the sample was calculated using a calibrated graph. The standards were water solutions of $Na_2S{\cdot}9H_2O$ in the range of 31.2–3120 μM. Optical density was measured at 670 nm in a cuvette with optical path of 1,0 cm using Apel PD-303 spectrophotometer (Japan).

Activity of the H_2S-synthesizing enzyme cystathionine-γ-liase (CSE, EC4.4.1.1) in the postnuclear supernatant of the homogenates of aorta, myocardium and kidneys was determined by the increase in sulfide anion as described here [15]. Activity of the CSE in the reaction of cysteine desulfuration was determined in incubation medium containing in final concentrations pyridoxal phosphate 1.34 mM, L-cysteine 6.0 mM, Tris-HCl buffer 0.08 M (pH 8.5). To 0.5 ml incubation medium were added postnuclear homogenates of organs (protein content 1–2 mg). The samples were incubated at 37° C for 60 min in sterile hermetically sealed plastic Eppendorf tubes (to prevent H_2S losses). The reaction was stopped by cooling the tubes on ice, then there was added 0.5 ml 1% zinc acetate solution to bind the produced H_2S. The control samples were treated similarly, for the exception that the investigated material was added to the medium only after incubation and cooling. The amount of H_2S was determined by the methylene blue production by a standard method [16]. To the samples were added 0.5 ml 20 mM N, N-dimethyl-p-phenylenediamine in 7.2 M HCl, 0.4 ml 30 mM $FeCl_3$ in 1.2 M HCl, incubated for 20 min at 18–22° C, then added 1 ml 20% TCA, centrifuged for 10 min at 3000 rpm. The optical density of the supernatant was measured at 670 nm in a cuvette with optical path of 1.0 cm using Apel PD-303 spectrophotometer (Japan). Sulfide anion content in the sample was calculated using a calibrated graph.

The expression of the CSE gene in aorta, myocardium and kidneys were determined by Real-Time PCR. Total RNA was isolated from tissues using AmpliSens RIBO-zol-B kit (AmpliSens, Russia). To obtain cDNA a reverse transcription kit (Sintol, Russia) was used, that included 2.5 x reaction mix, 15 U/ml primer of oligo$(dT)_{15}$, 50 U/μl reverse transcriptase MMLV-RT, 5 U/μl RNAase inhibitor and deionized water free from RNases. 1–2 μg total RNA was added to the mix.

The determination of the CSE gene expression was done in the presence of SYBR Green I dye, using detection amplifier DT-Light (DNK-Tekhnologia, Russ) in the reactionary mix: 10 x buffer for amplification with SYBR Green I dye; 25 mM $MgCl_2$; 2.5 mM deoxynucleoside triphosphates; specific primers to *CSE* gene (5'-GCTGAGAGCCTGGGAGGATA-3', 5'-TCACTGATCCCGAGGGTAGCT-3') and 5 U/μl SynTag DNA-polymerase. 5 μl DNA sample were added to the mix. The primers to the β-actin gene: 5'-ACCCGCGAGTACAACCTTCTT-3' and 5'-TATCGTCATCCATGGCGAACT-3' were used

as the referent gene. Amplification regime was: 94° C, 3 min, 40 cycles: – 94° C, 15 s; – 64° C, 40 s. For data analysis, we used the Ct method: the relative level of mRNA CSE / β-actin was estimated as $2^{-\Delta Ct}$, where $\Delta Ct = Ct_{CSE} - Ct_{\beta\text{-actin}}$, and Ct_{CSE} is the threshold cycle of cDNA amplification of the target gene CSE; $Ct_{\beta\text{-actin}}$ – threshold cycle of cDNA amplification of the referent gene for β-actin.

Raw data were treated using universal statistical programs MS Excel, SPSS22 for Windows, STATISTICA 6,0 (license № AXXR910A374605FA). To evaluate between-group difference we used Student's parametric t-criterion, when the data were not normally distributed – Mann-Whitney U-test. The normality was evaluated using the Shapiro-Wilk test. The correlation was evaluated after Pearson. Statistical significance was assumed at $p<0.05$. Results are given as mean and standard deviation (M±SD).

Results. It was shown that at STZ-diabetes there is an increase in galectin-3 in blood, heart and kidneys by 3.5, 5 and 5.5 times, respectively ($p<0.05$), compared to control (Fig. 1). Injection of propargylglycine to STZ-diabetic rats caused further increase in galectin-3 in blood, heart and kidneys of rats by 55, 77.6 and 80%, respectively ($p < 0.05$), compared to untreated animals with DM. Whereas, introduction of NaHS decreased galectin-3 in blood, heart and kidneys of STS-diabetic rats: its level was lower by 22, 58.2 and 63.2%, respectively ($p < 0.05$) than in untreated animals with DM.

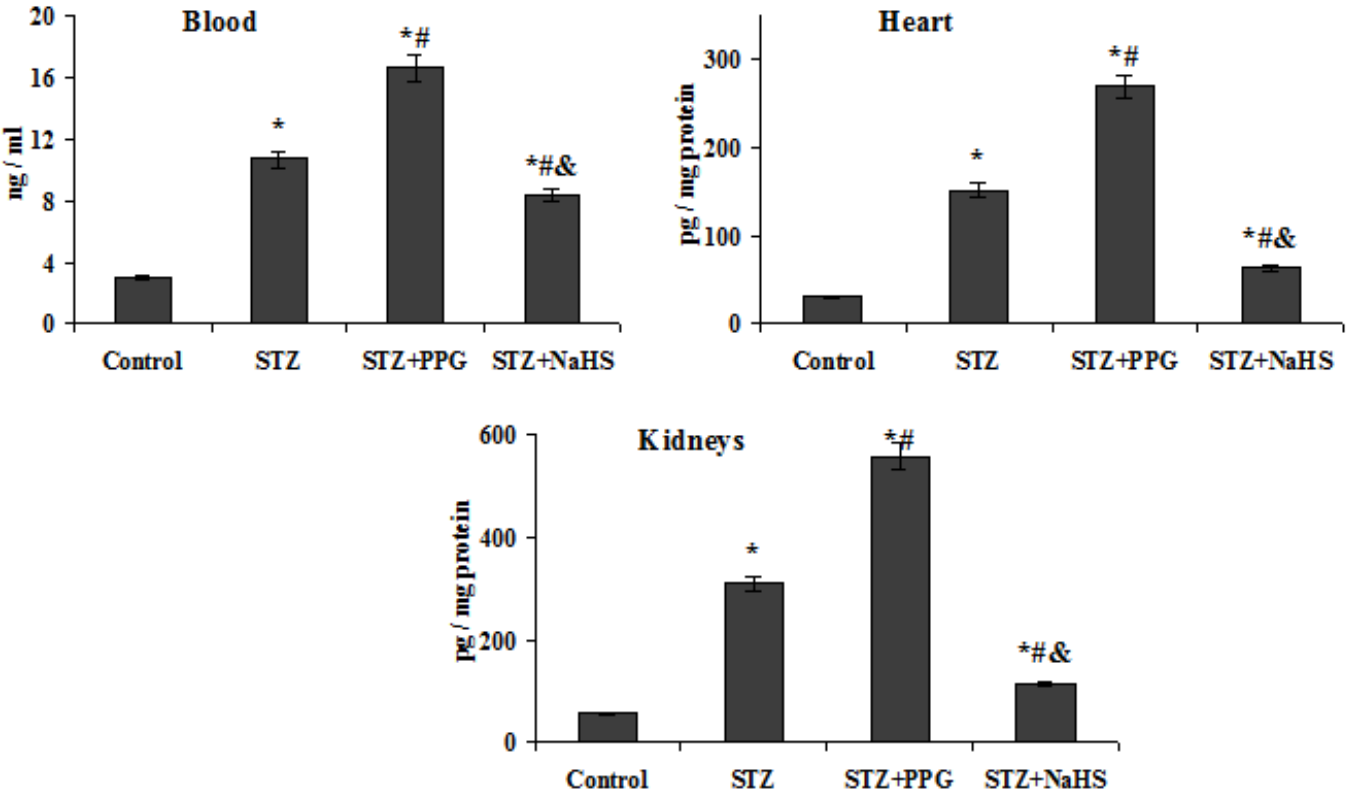


Figure 1. Effect of propargylglycine and NaHS on galectin-3 in blood, heart and kidneys of rats with streptozotocin-induced diabetes (n=10):
* – $p<0.05$ compared to control; # – $p<0.05$ compared to untreated animals with STZ-diabetes;
& – $p<0.05$ compared to animals with STZ-diabetes which were given propargylglycine

Experimental DM causes decrease in H_2S in aorta, heart and kidneys by 40; 36.7 and 38.6%, respectively ($p<0.05$) compared to the control rat group (Fig. 2). The treatment of DM-rats with propargylglycine worsen H_2S deficit in aorta, heart and kidneys – the level was lower by 34.8; 29.4 and

33%, respectively ($p<0.05$) compared to untreated animals. On the other hand, introduction of NaHS to rats was followed by increase in H_2S in aorta, heart and kidneys by 28.1, 23.5 and 26.5%, respectively ($p<0.05$) compared to DM-rats.

According to the correlation analysis, under STZ-diabetic condition in rats, H_2S content in aorta, myocardium and kidneys has statistically significant anti-correlation with the correspondent levels of galectin-3 ($r= -(0,76–0,82)$, $p < 0.05$).

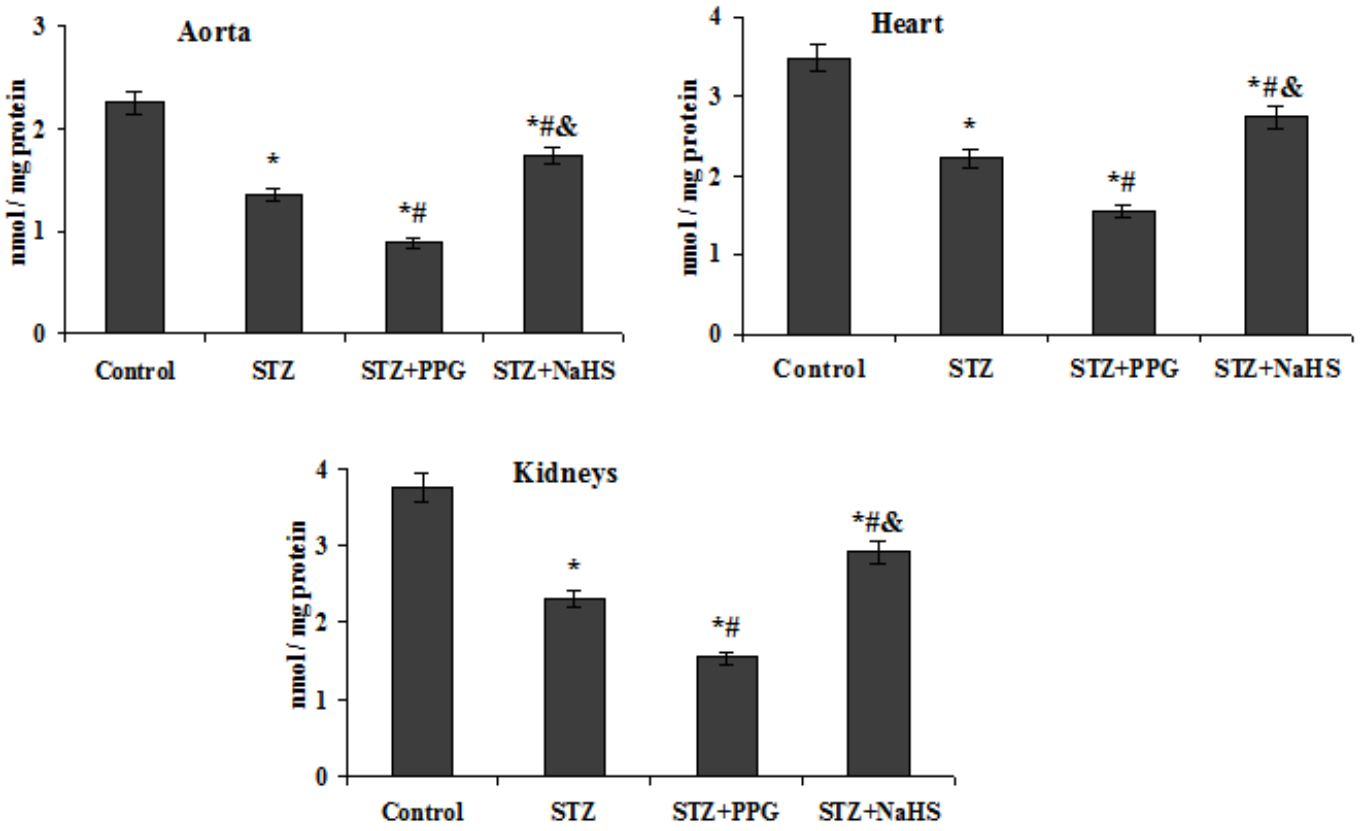


Figure 2. Effect of propargylglycine and NaHS on H_2S in aorta, heart and kidneys of rats with streptozotocin-induced diabetes (n=10):
– $p<0.05$ compared to the control group; # – $p<0.05$ compared to untreated animals with STZ-diabetes; & – $p<0.05$ compared to animals with STZ-diabetes treated with propargylglycine

DM causes decrease in the activity of H_2S synthesis by the mean of cysteine desulfation reaction catalyzed be CSE in aorta, heart and kidneys by 57.7; 55.1 and 56.6% respectively ($p<0.05$) compared to control (Fig. 3). Propargylglycine potentiated the DM's negative effect on the enzymatic synthesis of H_2S: in aorta, heart and kidneys of rats there was registered alleged reduction in CSE activity by 44.4; 35.9 and 42.4%, respectively ($p<0.05$) compared to untreated DM animals. Meanwhile injection of NaHS was followed by the opposite changes; in aorta, heart and kidneys the CSE activity was 2.1, 2 and 2.2-times higher ($p<0.05$) than in untreated animals.

In animals with modelled DM expression of the CSE gene was lower in aorta, heart and kidneys by 72, 60.5 and 70.6% ($p<0.05$), respectively, compared to control (Fig. 4). H_2S metabolism modulators had opposite effects on the parameter. Propargylglycine in DM animals caused yet further depression of this gene. At these conditions the mRNA of CSE/β-actin in aorta, heart and kidneys was lower by 44, 29.4 and 38.5%, respectively ($p<0.05$), compared

to untreated animals with STZ-diabetes. The introduction of NaHS, on the other hand, induced the CSE gene expression: the mRNA of CSE / β-actin in aorta, heart and kidneys of rats was higher by 49, 35.3 and 42.8%, respectively ($p<0.05$), compared to untreated animals.

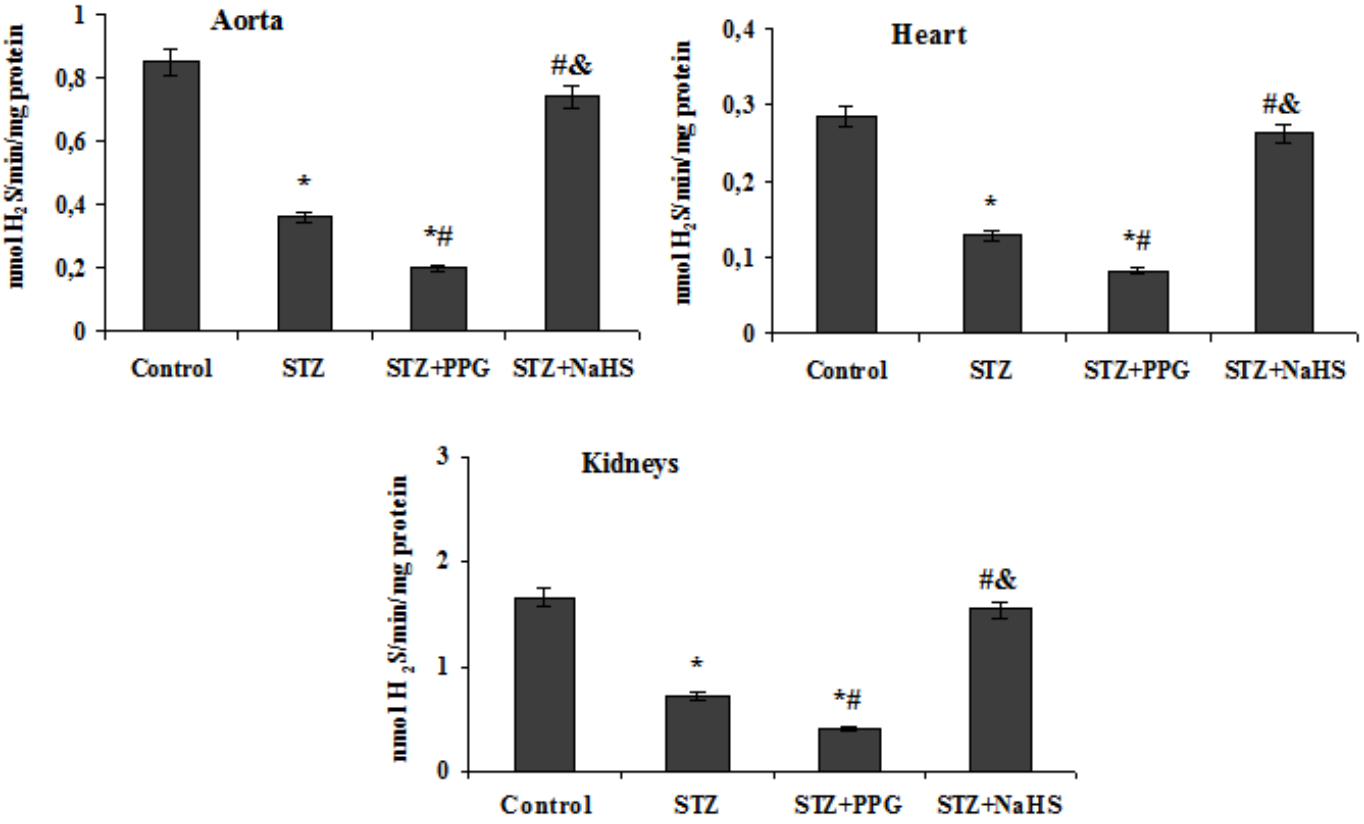


Figure 3. Effect of propargylglycine and NaHS on CSE activity in aorta, heart and kidneys in rats with STZ-induced diabetes (n=10): * – $p<0.05$ compared to the control group; # – $p<0.05$ compared to untreated animals with STZ-diabetes; & – $p<0.05$ compared to animals with STZ-diabetes which were treated with propargylglycine

Discussion. According to the obtained data, diabetes is accompanied by a decrease in the activity and expression of CSE, an increase in utilization of exogenous H_2S and a decrease in H_2S content in the aorta, heart and kidneys of rats. The development of H_2S deficiency in the organs of diabetic rats is closely correlated with the level of hyperglycemia. The obtained results are confirmed by the literature: incubation of endothelial cells of mice in a medium with a high glucose concentration was accompanied by a decrease in H_2S [17].

Disorders of H_2S metabolism in diabetes are among the pathogenetic factors in the development of endothelial dysfunction, cardio- and nephropathy [3; 9]. According to the literature, the negative impact of low concentration of H_2S on the cardiovascular system and kidneys in diabetes is realized through various mechanisms, including the activation of fibrogenesis. Nowadays it is known that the profibrogenic effect of H_2S deficiency is realized through the induction of inflammation, oxidative stress, hyperexpression of transforming growth factor TGF-β1, myogen-activated protein kinase MARK, matrix metalloproteinase MMP-9 [3; 8; 9; 10].

According to our studies, in diabetes the disorders of H_2S metabolism closely correlate with an increase in galectin-3, another important profibrogenic mediator. The use of propargylglycine,

an inhibitor of H_2S synthesis, in diabetes deepens H_2S deficiency in tissues and increases galectin-3, while H_2S donor (NaHS) – alleviates these changes.

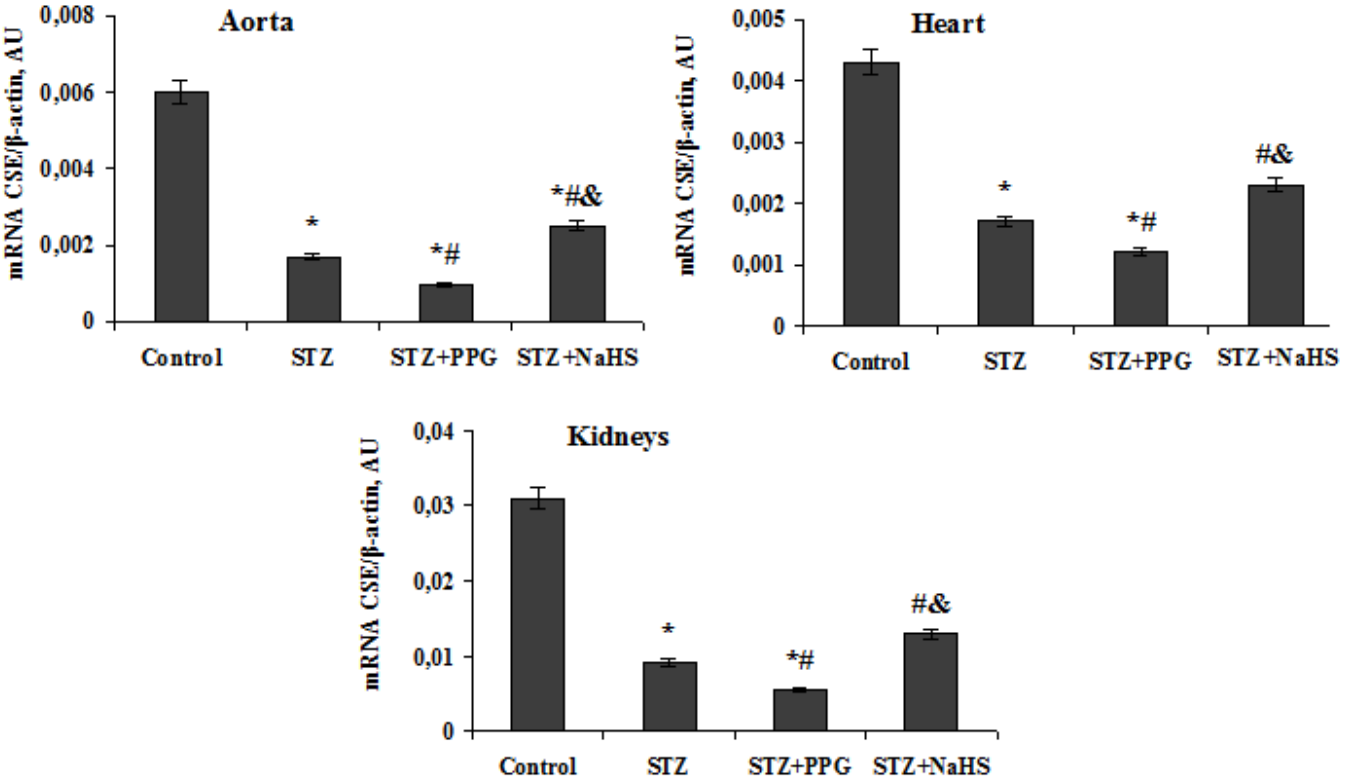


Figure 4. Effect of propargylglycine and NaHS on the expression of the CSE gene in aorta, heart and kidneys of rats with streptozotocin-induced diabetes (n=10):
* – $p<0.05$ compared to control group; # – $p<0.05$ compared to untreated animals with STZ-diabetes; & – $p<0.05$ compared to animals with STZ-diabetes that have received propargylglycine

However, how do the modulators of H_2S effect galectin-3 system, through which biological mechanisms? To answer these questions, it requires further research. Another promising area is the development of new drugs with a stimulating effect on the H_2S system in tissues in order to provide effective correction of the endothelial dysfunction and fibrogenesis in animal organs in diabetes.

Conclusions. Diabetes mellitus, induced by streptozotocin, is associated with an increase in galectin-3 in the blood, aorta, heart and kidneys, a decrease in H_2S, a decrease in the activity and expression of CSE. Inhibition of activity of the H_2S / CSE system is accompanied by a significant increase in galectin-3 in the blood and tissues, which can accelerate the development of diabetes-associated organ dysfunction and angiopathy.

References:

1. World Health Organization [Electronic resource].– Access mode: URL: http://www.euro.who.int.
2. Tsytovskyi M. N. Statistical, clinical and morphological aspects of impact of diabetes on the cardiovascular system. Scientific Bulletin of Uzhhorod University. Series: Medicine. 2017; 1 (55): 168–177. [in Ukranian].

3. Sun H. J., Wu Z. Y., Cao L., Zhu M. Y., Liu T. T., Guo L., Lin Y., Nie X. W., Bian J. S. Hydrogen Sulfide: Recent Progression and Perspectives for the Treatment of Diabetic Nephropathy. Molecules. 2019; 24(15). Doi: 10.3390/molecules24152857.
4. Yilmaz S., Canpolat U., Aydogdu S., Abboud H. E. Diabetic Cardiomyopathy; Summary of 41 Years. Korean Circ J. 2015; 45(4): 266–272.
5. Pugliese G., Iacobini C., Ricci C., Blasetti Fantauzzi C., Menini S. Galectin-3 in diabetic patients. Clin Chem Lab Med. 2014; 52(10):1413–1423.
6. Pejnovic N. N., Pantic J. M., Jovanovic I. P., Radosavljevic G. D., Milovanovic M. Z., Nikolic I. G., Zdravkovic N. S., Djukic A. L., Arsenijevic N. N., Lukic M. L. Galectin-3 deficiency accelerates high-fat diet-induced obesity and amplifies inflammation in adipose tissue and pancreatic islets. Diabetes. 2013; 62(6): 1932–1944.
7. Menini S., Iacobini C., Blasetti Fantauzzi C., Pesce C. M., Pugliese G. Role of Galectin-3 in Obesity and Impaired Glucose Homeostasis. Oxid Med Cell Longev. 2016. Doi: 10.1155/2016/9618092.
8. Zaichko N. V., Melnik A. V., Yoltukhivskyy M. M., Olhovskiy A. S., Palamarchuk I. V. Hydrogen sulfide: modern aspects of metabolism, biological and medical role. Ukr. Biochem. J. 2014; 86(5): 5–25.
9. Szabo C. Roles of hydrogen sulfide in the pathogenesis of diabetes mellitus and its complications. Antioxid Redox Signal. 2012; 17(1): 68–80.
10. Liu M., Li Y., Liang B., Li Z., Jiang Z., Chu C., Yang J. Hydrogen sulfide attenuates myocardial fibrosis in diabetic rats through the JAK/STAT signaling pathway. Int J Mol Med. 2018; 41(4): 1867–1876.
11. Lee A. T., Shah J. J., Li L., Cheng Y., Moore P. K., Khanna S. A nociceptive-intensity-dependent role for hydrogen sulphide in the formalin model of persistent inflammatory pain. Neuroscience. 2008; 152(1): 89–96.
12. Liu H., Bai X. B., Shi S., Cao Y. X. Hydrogen sulfide protects from intestinal ischaemia-reperfusion injury in rats. J Pharm Pharmacol. 2009; 61(2): 207–212.
13. Zhu Y. Z., Wang Z. J., Ho P., Loke Y. Y., Zhu Y. C., Huang S. H., Tan C. S., Whiteman M., Lu J., Moore P. K. Hydrogen sulfide and its possible roles in myocardial ischemia in experimental rats. J Appl Physiol. 2007; 102(1): 261–268.
14. Wiliński B., Wiliński J., Somogyi E., Piotrowska J. Góralska M. Amlodipine affects endogenous hydrogen sulfide tissue concentrations in different mouse organs. Folia Med Cracov. 2011; 51(1–4): 29–35.
15. Zaichko N. V., Pentiuk N. A., Melnik A. V., Shtatko E. I., Andrushko I. I. Production of hydrogen sulfide in organs of rats. Medical chemistry. 2009; 11(4): 7–13. [in Ukranian].
16. Dombkowski R., Russell M., Olson K. Hydrogen sulfide as an endogenous regulator of vascular smooth muscle tone in trout. Am. J. Physiol. Regul. Integr. Comp. Physiol. 2004; 286: 678–685.
17. Kundu S., Pushpakumar S., Khundmiri S. J., Sen U. Hydrogen sulfide mitigates hyperglycemic remodeling via liver kinase B1-adenosine monophosphate-activated protein kinase signaling. Biochim Biophys Acta. 2014; 1843(12): 2816–2826.

Section 3. Medical psychology

https://doi.org/10.29013/ELBLS-20-4-37-39

Buruiana Sanda,
Associate Professor of the Department of Hematology
SUMPh "Nicolae Testemitanu" Chisinau, Republic of Moldova
E-mail: sandaburuiana69@gmail.com

THE IMPORTANCE OF THE PSYHO-SPIRITUAL STATE OF PATIENTS WITH NON-HODGKIN'S LYMPHOMA

Abstract. Non-Hodgkin lymphomas (NHL) are a diverse group of neoplastic disorders, the incidence of which has shown a significant increase in rates over the year, worldwide, as well as in the Republic of Moldova. Is one of the most common hematological malignancies. Due to the efficient treatment methods, the rate of complete remissions increased and the patient's lifespan. The long time of monitoring involves special attention to the quality of life, especially the psychological and spiritual well-being of the patient with malignant lymphoma. Psycho-spiritual well-being is a subjective experience that incorporates both emotional health and meaning-in-life concerns. The psychological well-being was highlighted in the results of the GWB (Global Well Being) study group. The information obtained from the patient, from this questionnaire, although it is subjective, counts when making decisions in medical conduct and in the complex, subsequent recovery.

Keywords: lymphoma, well-being, quality of life.

1. Background

Non-Hodgkin lymphomas (NHL) are a diverse group of neoplastic disorders, the incidence of which has shown a significant increase in rates over the year, worldwide, as well as in the Republic of Moldova [1]. Due to the efficient treatment methods, the rate of complete remissions increased and the patient's life expectancy and lifespan [2; 3]. The response to treatment and patient survival did not reflect data about psychological and emotional well-being of the patient with NHL [3; 4]. Psycho-spiritual well-being is a subjective experience that incorporates both emotional health and meaning-in-life concerns [5]. Anxiety, depression, disease-related stress, coping strategies and optimism they are risk factors, which negatively influence the quality of life, survival and dysfunction of patients with NHL [6]. NHL patients experience high levels of psychological distress during intensive and long-term diagnosis and treatment [7; 8]. Secondary depression of the patient with oncological disease, is manifested by persistent depressive mood. Patients have a feeling of insecurity, regret that they got sick and no longer have the opportunity to do the things, the activities they could have done before [9]. The general clinical psychic manifestations of depression are: malaise, pessimism, decreased libido, lack of appetite, inactivity, inability to concentrate, frequent thoughts about death, suicidal tendencies [9]. The general somatic clinical manifestations of depression are: insomnia, irritability,

weight loss [9; 10]. The anxiety of cancer patients is manifested by: fear, fear of death, insecurity of the disease, fear of not being able to control pain, fear of disabilities that create dependence on others, fear of changing appearance [9; 10].

The quality of life of a lymphoma patient can be influenced by the choice of the type of treatment, but also by the implementation of rehabilitation measures [8; 11].

2. Study methods

The study included 66 patients with the diagnosis of NHL: 37 patients with aggressive NHL and 29 patients with indolent NHL, who have met the criteria for inclusion in the study. The psycho-spiritual well-being of patients with NHL was assessed on the basis of the patient's independent completion of the GWB (general psychological well-being index) questionnaire. Patients answered 18 questions with 6 answer options, which reflect 6 basic parameters: anxiety, depression, general health, self, vitality and wellbeing. Each answer has 6 options (0–5), both negative and positive. This questionnaire reflects only the patient's subjective feelings about his own suffering and psychological well-being. The average time required to complete the questionnaire was approximately 7–10 minutes.

3. Results

Patients with NHL included in the study, treated in the Hematology Department of PMSI IO (Public Medico-Sanitary Institution Oncological Institute, Chisinau, Republic of Moldova), had a mean age of 58.3 ± 1.3 years (32–78 years). Women accounted for 53% (35 patients) and men – 47% (31 patients).

The results of the analysis of the psychological suffering of the NHLs patients, depending on the age, find a severe psychological suffering at the stage of establishing the diagnosis, regardless of age. Medium stress and good psychological well-being were more common in patients over 50 years of age (11% and 14%, respectively).

Following the analysis of the obtained results, we can mention the predominance of severe psychological distress in all types of NHL with a predominance in aggressive lymphomas (64%). Good psychological condition predominated in patients with indolent NHL (26%), as opposed to patients with aggressive NHL (17%). This could be explained by the fact that indolent NHL is characterized by slow progressive progression, prolonged lack of signs of general intoxication that would affect the quality of life, but aggressive NHL has a rapid evolution, but with a high sensitivity to chemotherapy with lasting complete remissions.

Severe psychological distress was determined by the predominance of anxiety, depression, psychological tension especially in women (73%). This could be explained by the fact that in today's society, women's roles often include family obligations, caregiving for children and/or elderly parent and work responsibilities as well as other roles. They may feel a sense of failure in not being able to meet expectations for themselves and others.

I found that patients with NHL are subjected to high levels of psychological distress during diagnosis and treatment regardless of the degree of spread of the tumor process. The diagnosis of NHL conditions psycho-spiritual changes, due to discomfort, pain, aesthetic changes, interruption of professional activity, lifestyle changes, both the patient and his family. In patients with localized process (I and II stages) good psychological state was manifested in 50% of cases, and in patients with generalized process (stage IV), in 64% of cases, severe psychological suffering was highlighted.

4. Conclusion

Severe psycho-emotional suffering is characteristic of all patients, regardless of age, with malignant lymphoma at the stage of establishing the diagnosis, especially in aggressive lymphomas (64%). Regardless of the morphological type of NHL (aggressive or indolent) in 73% of cases, women developed severe psychological stress. Good psychological well-being was manifested in 50% of patients with stage I of NHL, and in 64% patients in stage IV severe psychological suffering was highlighted. Assessing patient's

psycho-spiritual changes will allow us to individually select specific and nonspecific treatment behaviors. The basic goal is that we treat not only cancer, but a patient with cancer.

References:

1. Buruiana S. Actualități în studierea limfoamelor non-Hodgkin indolente. Anale științifice ale Universității de Stat de Medicină și Farmacie "Nicolae Testemițanu". 2013.– Vol. 3.– P. 470–477.
2. Monul V. The quality of life of patients with advanced cancer undergoing outpatient chemotherapy. INFO-MED. 2016; 28(2):223–227.
3. Buruiana S., Robu M., Mazur-Nicorici L., Tomacinschii V., Mazur M. Assessing the quality of life in patients with non-Hodgkin's Lymphoma is a burden or an advantage? Archives of the Balkan Medical Union. 2020; 55(3):418–424.
4. Papadopulus D. The role of well-being, spirituality and religiosity for successful aging in later life: a brief review. Advances in aging research. 2020; 9(2). DOI: 10.4236 / aar.2020.92003.
5. Lin H., Bauer-Wu S. Psycho-spiritual well-being in patients with advanced cancer: an integrative review of the literature. Journal of Clinical Nursing. 2003; 44 (1): 69–80.
6. David A., Nastase S. Quality of life in patients with oncological conditions and depressive-anxiety co-morbid pathology. Quality of Life. 2012; 2(1): 45–62.
7. Wang Z., Li L., Shi M. et. al. Exploring correlations between positive psychological resources and symptoms of psychological distress among hematological cancer patients: a cross-sectional study. Psychology Health. 2016; 21(5): 571–582.
8. Попов Т., Чулкова В. Отношение к лечению и типы реагирования на болезнь у онкологических пациентов при системных методах терапии. Психологические науки. Наука и образование. 2017; 108–111.
9. Csaba D. Psychosocial aspects of tumor diseases. University dam of Cluj. 2015; 1–47.
10. Pulgar A., Alcala A., Reyes del Paso G. Psychosocial predictors of quality of life in hematological cancer. Behavioral Medicine. 2015; 41(1). URL: https://doi.org/10.1080/08964289.2013.833083
11. Mansano-Schlosser T., Ceolim M. Quality of life of cancer patients during the chemotherapy period. Texto contexto-enferm. 2013; 21(3). URL: https://www.researchgate.net/publication/338164108

Section 4. Life sciences

https://doi.org/10.29013/ELBLS-20-4-40-46

Jiaao Bao,
George School
1690 Newtown-Langhorne Rd, Newtown, PA 18940
E-mail: baoj@georgeschool.org
Dr. Jinan Liu,
PhD, Director of Outcome Research, Merck & Co
E-mail: jliu4@tulane.edu
770 Sumneytown Pike, West Point, PA 19486

PREDICTING ADOLESCENT PHYSICAL ACTIVITY: DEVELOPMENT AND VALIDATION OF TWO PREDICTIVE MODELS

Abstract. Physical Activity plays an imperative role in adolescents' body and mental development. Regular physical activities can help reduce the risk of developing diseases like type II diabetes, while being physical inactive will lead to increased risk of cardiovascular disease. For both parents and teachers, they need to have a robust instrument to evaluate adolescents' physical activity condition. In this report, response data of 9.045 high school students of 14 to 17 years old from the 2017 Youth Risk Behavior Surveillance Survey are analyzed. Several pre-processing techniques such as missing value exclusion, and min-max scaling are applied to prepare the data set for model-building. Then a list of selected variables including physical attributes, demographic variables, and sleeping habits are used to develop and validate two predictive models for predicting the probability of being physical active. The predictive models are further validated by an overall evaluation of the model, statistical tests of individual predictors, and an assessment of relative importance of the independent variables. The predictive models demonstrate good and similar performance. The AUC of the models are 0.724 and 0.732, respectively. The results indicate that holding more physical education (PE) class should be the most effective way to improve adolescents' physical activity level.

Keywords: Physical activity, Youth Risk Behavior Surveillance Survey, predictive model, logistic regressions.

1. Introduction

Physical activities play an unignorable role in people's daily life and have both short- and long-term health benefits. Regular physical activities can improve health and reduce the risk of developing diseases like type II diabetes, cancer and cardiovascular disease. For adolescents, regular exercise can help them improve cardiorespiratory fitness, build strong bones and muscles, control weight, reduce symptoms of anxiety and depression, and reduce

the risk of developing health conditions such as obesity. The consequences for physical inactivity, according to the Centers for Disease Control and Prevention (CDC), includes energy imbalance and the increased risk of factors for cardiovascular disease. Therefore, for the health of the students and children, it is important for schools, parents, and the government to have some understanding about the physical condition of the adolescents.

The main hypothesis of this study is that the likelihood that a high school student conducts regular physical activities is related to one or more factors such as his/her race, sex, age, weight, sleeping habit, dietary habit, smoking, use of alcohol, use of drug, etc. The main purpose of this study is to develop a predictive model to detect adolescent physical activity condition. In this study, two predictive models – logistic regression and artificial neural network are built, and their respective performance are measured. With the models, schools can collect survey data and score the students' probability of performing regular activities. For students with higher probability of inactivity, appropriate measures can be taken in early stage to improve their physical condition. The predictive model can be used to help foster physically and mentally healthy adolescents.

2. Method

2.1 Data

Using a three-stage cluster sample design, the Youth Risk Behavior Surveillance System (herein after referred to as YRBS dataset) is an epidemiologic surveillance system established by the CDC to monitor the prevalence of youth behaviors that most influence health [1] for 9th through 12th grade students. YRBS is a cross-sectional study and focuses on priority health-risk behaviors established during youth that result in the most significant mortality, morbidity, disability, and social problems during both youth and adulthood. These include behaviors that result in unintentional and intentional injuries; tobacco use; alcohol and other drug use; sexual behaviors that result in HIV infection, other sexually transmitted diseases (STDs), and unintended pregnancies; dietary behaviors; and physical activity, plus obesity and asthma.

The dataset of 2017 YRBS is used to identify potential associations of adolescent physical activity and the factors including dietary behavior, drinking behavior, smoking behavior, drug use, sleeping habit, etc. Observations with missing data points are excluded from the analysis. After cleaning, there are 9.045 observations for students between 14 and 17 years old in the YRBS dataset. A list of selected questions is shown in (Table 1).

Table 1. – Description of the selected questions

Item	Question	Function
1	*2*	*3*
1	How old are you?	Independent Variable
2	What is your sex?	Independent Variable
3	In what grade are you?	Independent Variable
4	Are you Hispanic or Latino?	Independent Variable
6	How tall are you without your shoes on?	Independent Variable
7	How much do you weigh without your shoes on?	Independent Variable
30	Have you ever tried cigarette smoking, even one or two puffs?	Independent Variable
40	During your life, on how many days have you had at least one drink of alcohol?	Independent Variable
46	During your life, how many times have you used marijuana?	Independent Variable
69	Which of the following are you trying to do about your weight?	Independent Variable

1	2	3
78	During the past 7 days, on how many days did you eat breakfast?	Independent Variable
79	During the past 7 days, on how many days were you physically active for a total of at least 60 minutes per day?	Dependent Variable
80	On an average school day, how many hours do you watch TV?	Independent Variable
81	On an average school day, how many hours do you play video or computer games or use a computer for something that is not school work?	Independent Variable
82	In an average week when you are in school, on how many days do you go to physical education (PE) classes?	Independent Variable
84	During the past 12 months, how many times did you have a concussion from playing a sport or being physically active?	Independent Variable
88	On an average school night, how many hours of sleep do you get?	Independent Variable

2.2 Statistical Method

A two-stage process is involved in this statistical analysis. At stage I, techniques of missing value exclusion, dichotomizing, and min-max scaling are applied for better training purpose. Then a logistic regression and an artificial neural network model are developed with physical activity as a dependent variable and the variables from selected questions as independent variables. At stage II, several validation metrics are calculated for each model to measure and compare their relative performance.

2.2.1 Pre-processing

The data set is pre-processed in this step to improve both the training speed and accuracy. As most machine learning algorithms are not able to deal with missing values, all the data points with missing entries are excluded from training. Then the dependent variable is dichotomized, where students being physically active for a total of at least 60 minutes per day the week prior to the survey were classified as physical active and the remaining as physical inactive.

Some machine learning algorithms, such as artificial neural networks, require a specific technique called feature scaling which transforms different features into comparable scales for better training speed and accuracy. For each feature, its minimum and maximum value are first computed as x_{min} and x_{max}. Then each data point x_i with respect to that feature is replaced by y_i calculated as:

$$y_i = \frac{\mathrm{x} - x_{min}}{x_{max} - x_{min}}.$$

Finally, for training and test purposes, the YRBS dataset is partitioned into two datasets, the training dataset (70%) for model development, and the test dataset (30%) for model test.

2.2.2 Logistic Regression

Logistic regression is a part of a category of statistical models called generalized linear models, and it allows one to predict a discrete outcome from a set of variables that may be continuous, discrete, dichotomous, or a combination of these. Typically, the dependent variable is dichotomous, and the independent variables are either categorical or continuous. In logistic regression, each feature x_i has its specific weight w_i. The net input y is calculated as follows:

$$\ln\left(\frac{y}{1-y}\right) = w_0 + w_1 x_1 + \ldots + w_m x_m$$

2.2.3 Artificial Neural Network

An artificial neural network is a computational model vaguely inspired by the biological neural networks that constitute animal brains. An artificial neuron that receives a signal then processes it and can signal neurons connected to it. The "signal" at a connection is a real number, and the output of each

neuron is computed by some non-linear function of the sum of its inputs.

A multilayer network is an artificial neural network that consists of one input layer, several hidden layers, and one output layer. The input layer is the first layer, the output layer is the last layer, and any layers between them are hidden layers. The data are passed into the input layer, processed by the hidden layers, and finally transformed into predicted labels in the output layer. In this study, the model has two hidden layers.

2.3 Model Validation

Consider a two-class prediction problem, where the outcomes are labeled either as positive or negative. There are four possible outcomes from a binary classifier. If the outcome from a prediction is positive and the actual value is also positive, then it is called a true positive (TP); however, if the actual value is negative then it is said to be a false positive (FP). Conversely, a true negative (TN) has occurred when both the prediction outcome and the actual value are negative, and false negative (FN) is when the prediction outcome is negative while the actual value is positive. In this way, the true positive rate (TPR) can be calculated as follows:

$$TPR = \frac{TP}{TP + FN}$$

And the false positive rate (FPR) can be calculated as:

$$FPR = \frac{FP}{TN + FP}$$

A receiver operating characteristic curve, or ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The best possible prediction method would yield a point in the upper left corner of the ROC space. A random guess would give a point along a diagonal line from the left bottom to the top right corners. Points above the diagonal represent better than random classification results, while points below the line represent worse than random results. In general, ROC analysis is one tool to select possibly optimal models and to discard suboptimal ones independently from the class distribution. Sometimes, it might be hard to identify which algorithm performs better by directly looking at ROC curves. Area Under Curve (AUC) overcomes this drawback by finding the area under the ROC curve, making it easier to find the optimal model.

3. Results

3.1 Logistic Regression

The results of logistic regression analysis of high school students being physically active are listed in Table 2. From the logistic regression results, it is not hard to find that, taking a 95% confidence level, question 2, 3, 6, 7, 40, 78, 81, 82, 84, 88 are significant predictors of the dependent variable.

Table 2. – Logistic regression results

Predictor (Item Number)	Estimate (β)	Standard Error of Estimate	Wald's $\chi 2$	p
1	*2*	*3*	*4*	*5*
1	0.32	0.29	1.11	0.27
2	0.57	0.08	7.63	< 0.001
3	–0.55	0.21	–2.55	0.01
4	0.088	0.066	1.34	0.18
6	1.38	0.29	4.78	< 0.001
7	–0.80	0.30	–2.69	< 0.001
30	0.10	0.076	1.32	0.19
40	0.63	0.12	5.17	< 0.001
46	–0.13	0.11	–1.21	0.23

1	*2*	*3*	*4*	*5*
69	0.02	0.077	0.27	0.78
78	0.72	0.077	9.33	< 0.001
80	0.15	0.093	0.16	0.87
81	–0.61	0.077	–7.95	< 0.001
82	1.25	0.068	18.4	< 0.001
84	0.76	0.18	4.19	< 0.001
88	0.34	0.13	2.56	0.01

3.2 Artificial Neural Network

The structure of the artificial neural network is shown in (Figure 1). The thickness of the line represents the corresponding weight.

To find the relative importance of independent variables, Garson describes a method that can be used to identify the relative importance of independent variables for a single dependent variable in an artificial neural network [2]. The relative importance of a specific independent variable for the dependent variable can be determined by identifying all weighted connections between the nodes of interest. That is, all weights connecting the specific input node that pass through the hidden layer to the dependent variable are identified. This is repeated for all other independent variables until a list of all weights that are specific to each independent variable is obtained [2]. Figure 2 shows the importance of each question using Garson's algorithm.

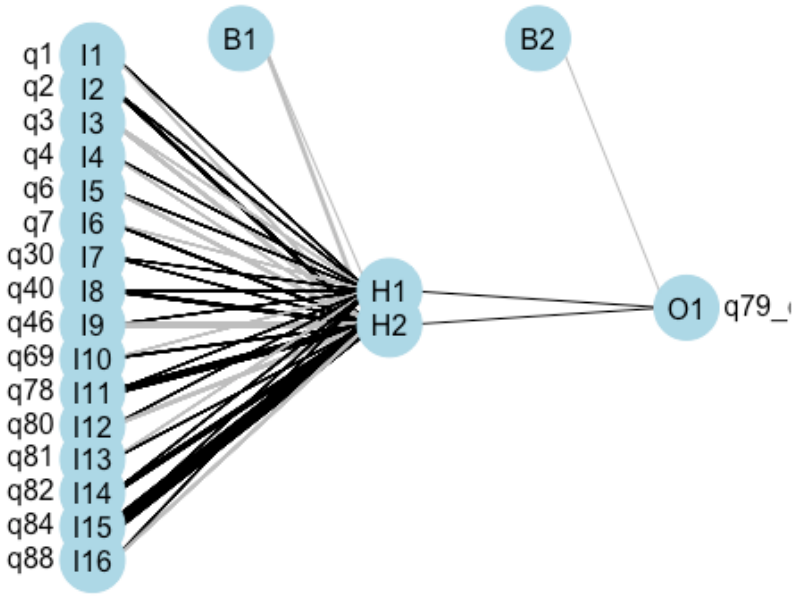


Figure 1. Structure of the artificial neural network

3.3 Model Validation

Figure 3 displays the ROC curve for the two models and Table 3 lists their respective AUC score. Combining both Figure 3 and Table 3, it can be concluded that both models have achieved a rather similar performance, while the artificial neural network being slightly better than the logistic regression. Besides, we can also see that both models have results better than random guessing.

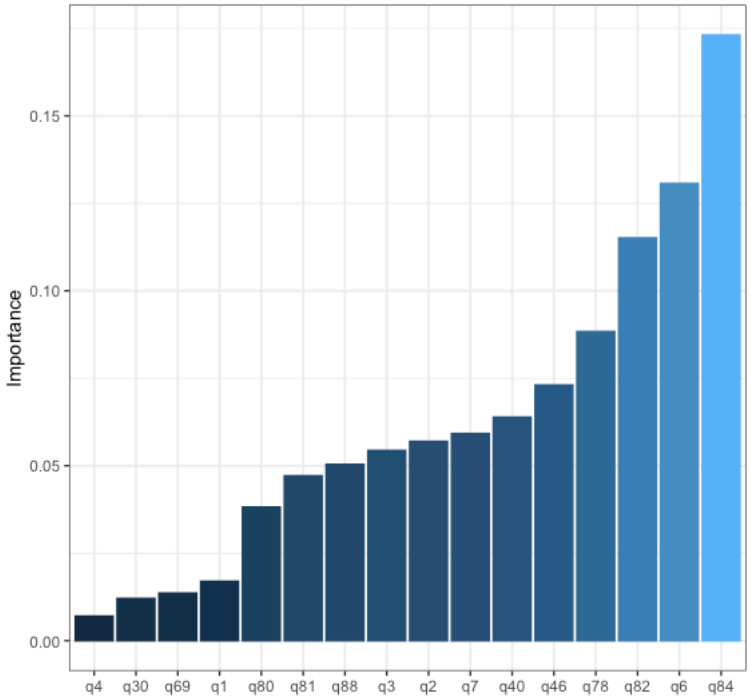


Figure 2. The importance of each question in the artificial neural network

Table 3.– The AUC score for the two models

Algorithm	AUC Score
Logistic Regression	0.724
Artificial Neural Network	0.732

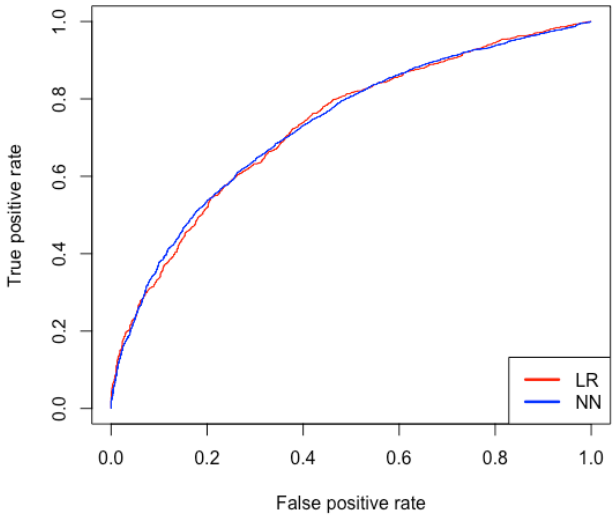


Figure 3. The ROC curve for the logistic regression and artificial neural network

4. Discussion

The intention of this study is to build a predictive model with the best performance and to investigate the factors most related to adolescents' physical activities. Two models – a logistic regression and an artificial neural network – are built, and all of them have achieved a similar performance. Also, using Garson's algorithm, we are able to ascertain that the question number 6, 82, 84 are most related to adolescents' physical activity level. Table 2 corroborates with this result by showing that these questions are also significant predictors of the dependent variable. Combining the results with Table 1, we are able to see that in order to increase adolescents' physical activities, it will be most effective to hold more physical education (PE) classes.

One limitation of the study is that data entries with missing values are excluded from analyzing. This is a timesaving but defective approach. Depending on the number of such data entries, it is possible that we might remove too many sample points, resulting in losing valuable information for the model to learn the relationship among independent variables. For future studies, we may use more advanced techniques such as mean value imputation or k-nearest neighbors (kNN). The mean value imputation

method completes missing values with the mean of the entire feature. This is a simply but effective way to make those entries usable by the learning algorithm. Other techniques include k-nearest neighbors, which replaces missing values with the mean of *k* nearest neighbors of that particular sample. This technique requires more efforts but can generally achieve better performance.

References:

1. National YRBS Data User's Manual. 2009.
2. Garson G. D. Interpreting neural network connection weights. Artificial Intelligence Expert.– 6(4): 1991. 46–51.
3. National YRBS Data Users Manual, 2017. refer to: URL: https://www.cdc.gov/healthyyouth/data/yrbs/index.htm
4. Peng C. J., Lee K. L., Ingersoll G. M. An Introduction to Logistic Regression Analysis and Reporting. The Journal of Educational Research, 96(1), 3–14.
5. Tabachnick B. and Fidell L. Using Multivariate Statistics (4th Ed.). Needham Heights, MA: Allyn & Bacon, 2001.
6. Stat Soft, Electronic Statistics Textbook, refer to: URL: http://www.statsoft.com/textbook/stathome.html
7. Stokes M., Davis C. S. Categorical Data Analysis Using the SAS System, SAS Institute Inc., 1995.

https://doi.org/10.29013/ELBLS-20-4-47-67

Jay Fu,
Watchung Hills Regional High School
E-mail: 221670@whrhs-stu.org

PREDICTING THE DIAGNOSIS OF ALZHEIMER'S DISEASE: DEVELOPMENT AND VALIDATION OF MACHINE LEARNING MODELS

Abstract. Patients with Alzheimer's disease progressively lose their memory and thinking skills and, eventually, the ability to carry out simple daily tasks. It is estimated that more than 5 million Americans are living with Alzheimer's disease. The disease is irreversible, but early detection and treatment can slow down the disease progression. In this research, publicly available MRI data were utilized to build models to predict dementia. Various machine learning models, including Logistic Regression, K-Nearest Neighbor, Support Vector Machine, Random Forest, and Neural Network, were developed. Data were divided into training and testing sets where training sets were used to build the predictive model, and testing sets were used to assess the accuracy of prediction. Key risk factors were identified, and various models were compared to come forward with the best prediction model. Conclusions, limitations, and future research were discussed at the end of the manuscript.

Keywords: Alzheimer's disease, Magnetic Resonance Imaging (MRI), Clinical diagnosis.

1. Introduction

Alzheimer's disease is currently ranked as the fourth leading cause of death in the United States, with approximately 65,800 fatalities attributable to the disease each year [1]. In fact, by early 2017, over 5.5 million people in the United States were diagnosed with Alzheimer's [2]. In the coming years, as more baby boomers reach and pass through old age, the number of Alzheimer's disease cases is expected to grow substantially [3]. In 2050, the number of Alzheimer's patients worldwide is expected to triple from 50 million (in 2018) to 152 million [4]. Furthermore, Alzheimer's disease is the most common form of dementia, which refers to the sustained deterioration of intellectual functions. About 60 to 80 percent of all dementia cases can be attributed to Alzheimer's [2].

Alzheimer's disease is caused by the degeneration and eventual death of a large number of neurons in several areas of the brain. Alzheimer's is a very gradual disease, and it starts with short-term memory loss, followed by the progressive loss of memory and cognitive and intellectual functions. This eventually leads to deterioration of physical functioning and incapacitation [1].

Despite the prevalence of Alzheimer's, especially among the elderly population, diagnosis for Alzheimer's remains a major challenge. Diagnosing Alzheimer's conclusively requires either an autopsy or brain biopsy [3]. However, autopsies can only be done after a patient's death, while brain biopsies are generally regarded as a procedure of last resort since they are costly, difficult to execute, and lengthy. Thus, most patients diagnosed with Alzheimer's are diagnosed through clinical diagnosis [7]. Studies have shown that community doctors in rural areas are only about 50 to 60 percent accurate in clinically diagnosing Alzheimer's [5]. According to researchers from Keenan Research Center for Biomedical Science at St. Michael's Hospital, among more than 1.000 people listed in the

National Alzheimer's Coordinating Center database, only 78% of the patients were accurately diagnosed by doctors [6]. One of the main reasons for inaccuracy of clinical diagnosis for Alzheimer's disease is that there are many other diseases with similar symptoms, like Parkinson's disease, diffuse white matter disease, and alcohol-associated dementia [1].

Although Alzheimer's disease is irreversible, an early, accurate diagnosis can help doctor's devise effective strategies to manage symptoms and plan for long-term care. Since treatment and care can influence how long one survives with Alzheimer's, patients who get diagnosed and treated at an early stage can potentially live longer and experience slower memory deterioration than those who begin later treatment [1]. In addition, an early diagnosis can significantly reduce treatment cost, since patients who begin treatment earlier routinely have access to more affordable options [4].

The goal of this study is to create a machine learning model that can accurately diagnose Alzheimer's based on a patient's demographic and clinical data including magnetic resonance imaging (MRI) data. This model would allow physicians or nurses to diagnose patients accurately, without a brain biopsy. A machine learning model for diagnosing Alzheimer's will help patients at the early stages of Alzheimer's get better care and timely treatment. As a result, an effective model could potentially allow patients to live longer with slower memory impairment.

2. Data

2.1 Overview of Data

The Open Access Series of Imaging Studies (OASIS) data set of Longitudinal MRI in Nondemented and Demented Older Adults consists of 150 distinct individuals, ranging from 60 to 96 years old, as well as data from 373 MRI imaging sessions. These individuals were selected from a larger pool who participated in MRI studies at Washington University. Each individual had at least two visits, separated by at least a year, during which MRI and clinical data were obtained. Based on the Clinical Dementia Rating (CDR) scale, participants were classified as nondemented, demented, or converted (from nondemented to demented).

CDR is a dementia-staging tool that uses six domains to rate subjects for cognitive impairment. The six domains include memory, orientation, judgment and problem solving, role in community affairs, home and hobbies, and personal care. The CDR scale is from 0 to 3. The numbers equate to the following diagnoses: non-dementia (0), very mild/questionable dementia (0.5), mild dementia (1), moderate dementia (2), and severe dementia (3). Individuals who had a CDR of 0 for all visits were categorized as nondemented. Individuals who had a CDR of 0.5 or higher for all visits were categorized as demented. Individuals who had a CDR of 0 on the initial visit and a CDR of at least 0.5 on any subsequent visits were categorized as converted. Of the 150 individuals in the data set, only 14 individuals were categorized as converted.

For each visit, MRI biomarkers, clinical data, and demographic data were collected, including estimated total intracranial volume (eTIV), normalized whole brain volume (nWBV), atlas scaling factor (ASF), Mini Mental State Examination (MMSE) result, age, gender, socioeconomic status, and years of education received. This study uses machine learning models to predict either demented (indicating the patient has Alzheimer's) or nondemented (indicating that the patient does not have Alzheimer's) based on the MRI, clinical, and demographic data obtained on each visit [8].

2.2 MRI Data

MRI produces detailed images of the brain by using radio waves and a strong magnetic field. Although it does not provide a definitive diagnosis, MRI-based measures of the brain have been regarded as valid markers indicating Alzheimer's progression. MRI can also help diagnose diseases that are commonly mistaken for Alzheimer's due similarity in symptoms [9]. Rather than using the raw image data from the MRI scan, the machine learning models in this study use biomarkers from the MRI data, such as eTIV, nWBV, and ASF.

The estimated total intracranial volume (eTIV) is the volume of the cranial cavity (the space inside the

skull) taken from an MRI. It is a standard measure that is used to correct for head size variation across subjects in brain studies, including Alzheimer' s-related studies [10]. In the data set used in this study, eTIV was estimated using the software FreeSurfer and was calculated by dividing a predetermined constant by the atlas scaling factor (ASF), the value that the MRI image is scaled by to align to the MNI305 head atlas, a brain mapping template for MRI scans. Thus, ASF is directly proportional to eTIV [8]. Since they are directly proportional to each other (high collinearity), only eTIV will be used in this study to train and test the machine learning models.

Normalized whole-brain volume (nWBV) refers to the proportion of tissue in the brain volume. In the data set, nWBV was evaluated using the FAST program. First, the MRI image was segmented to classify brain tissue as cerebral spinal fluid, gray matter, or white matter. The segmentation procedure assigned voxels (which are essentially 3D pixels) to the tissue classes using the Markov random field model. The nWBV was then evaluated as the proportion of all voxels categorized as tissue. Previous studies have shown that nWBV atrophies at a much greater rate in patients with early Alzheimer's [13].

2.3 Cognitive Assessment Data

Each patient in this study received a Mini Mental State Examination (MMSE). MMSE is the most widely used assessment for the evaluation of an individual's cognitive state. MMSE is a questionnaire that takes about 5 to 10 minutes to complete, which makes it a fast and easy way to evaluate cognitive state. MMSE is based on a 30-point score. A lower score indicates a greater degree of cognitive impairment. A score between 20 to 24 is associated with mild dementia, 13 to 20 is associated with moderate dementia, and below 12 is associated with severe dementia. MMSE tests six different aspects of cognitive ability applicable to Alzheimer's, including orientation of time and place, short-term memory recall, immediate recall, language, simple math calculation ability, and the ability to create a simple figure. In other words, MMSE tests various, everyday mental skills [11].

2.4 Exploratory Data Analysis

In the OASIS data set of Longitudinal MRI in Nondemented and Demented Older Adults, there are 150 individual participants who are categorized as demented, nondemented, or converted. Since the machine learning models in this study will be designed to predict either demented or nondemented (binary outcome), the first visit of each converted participant will be classified as nondemented and the last visit of each converted patient will be classified as demented. For converted patients, intermediate visits (only 9 visits out of 373) are disregarded from the data.

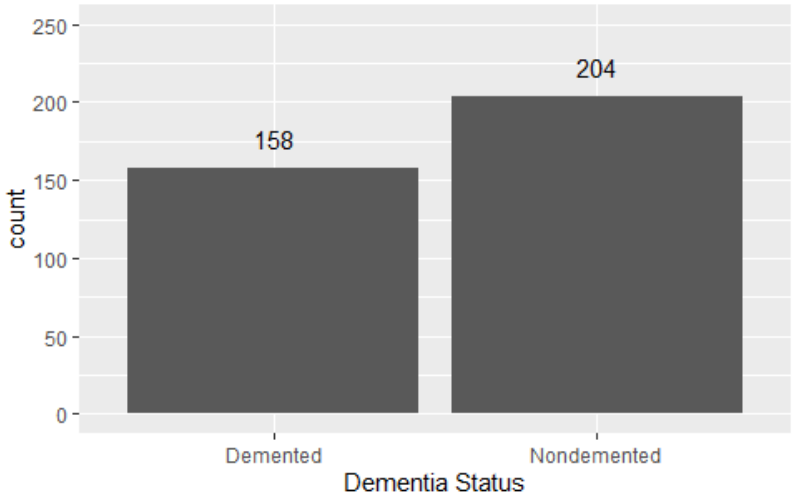


Figure 1. Count of Demented and Nondemented Visit

In addition, there are two subjects with two visits having missing MMSE data. These 2 visits will also be disregarded from the data in this study. Figure 1 provides a histogram representing the number of visits that are categorized as demented and nondemented. There are 158 visits categorized as demented and 204 visits categorized as nondemented.

MMSE is a cognitive assessment that is used to measure an individual's cognitive impairment. Lower scores of MMSE indicate higher degrees of cognitive impairment. Figure 2 clearly shows that demented individuals tend to have lower MMSE scores than nondemented individuals. The median MMSE for nondemented individuals, 29, is approximately 11.5% higher than the median MMSE for demented individuals, 26. Furthermore, the first quartile MMSE for nondemented individuals, 29, is approximately 32% higher than the first quartile MMSE for dementia individuals, 22. Demented subjects have a large variation in MMSE scores compared with nondemented subjects in (Figure 2).

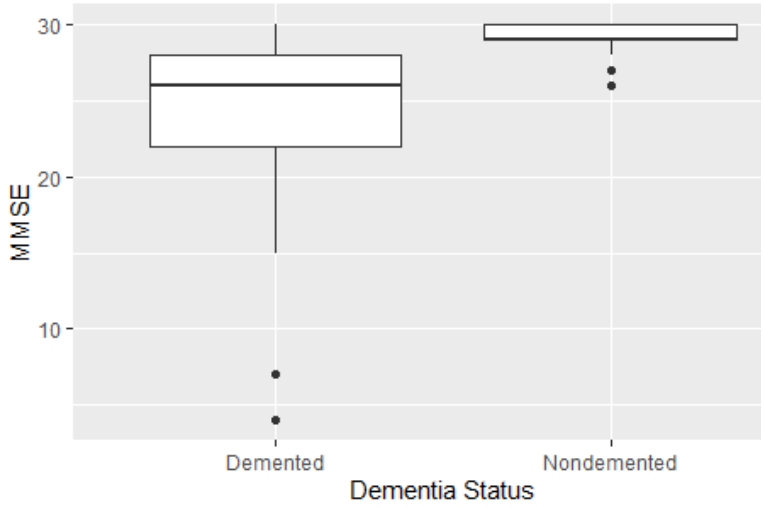


Figure 2. Boxplot of MMSE by Dementia Status

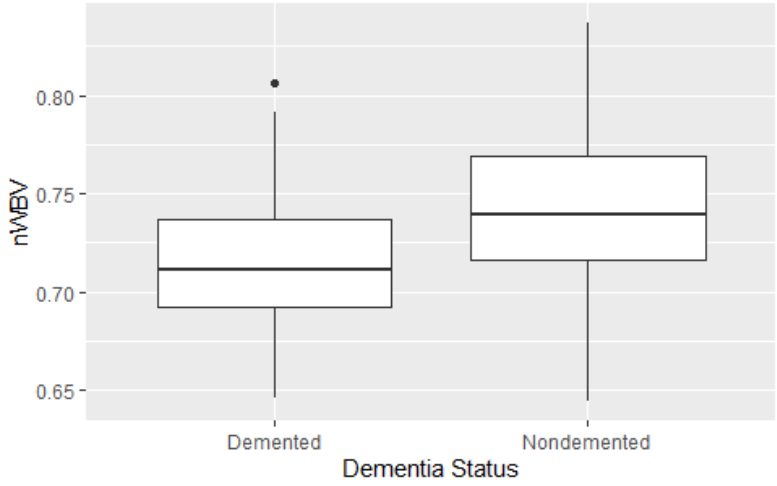


Figure 3. Boxplot of nWBV by Dementia Status

nWBV is calculated as the proportion of tissue in the brain volume. In this data set, nondemented individuals had a median nWBV of 0.7390, which is approximately 3% higher than the median nWBV for demented individuals, 0.711. As seen in Figure

3, nondemented individuals generally had higher nWBV than demented individuals.

In addition to clinical data, the machine learning models in this study also use demographic data to help diagnose Alzheimer's. In the data set used in this study, years of education correlate with the status of dementia. Individuals with more years of educations were less likely to be categorized as demented. As shown in Figure 4, the median years of education for nondemented individuals, 16, was 19% higher than the median years of education for demented individuals, 13.5.

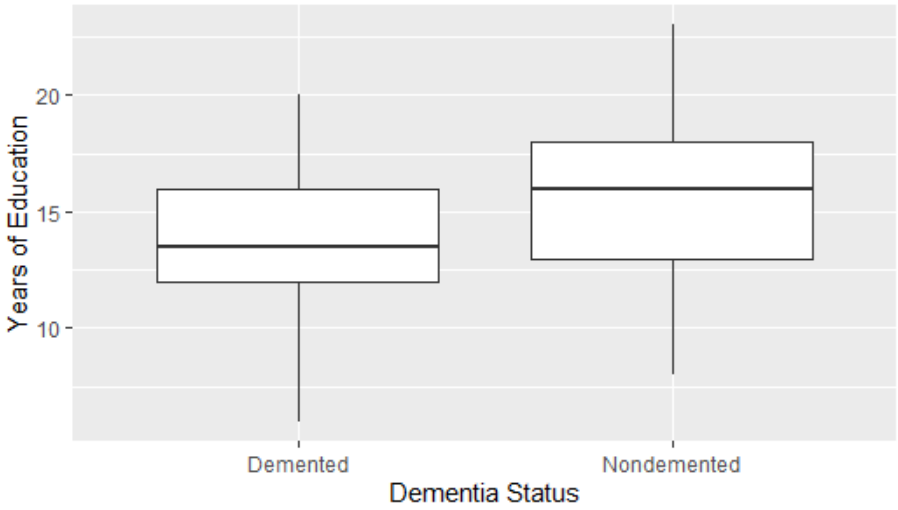


Figure 4. Boxplot of Yeas of Education by dementia Status

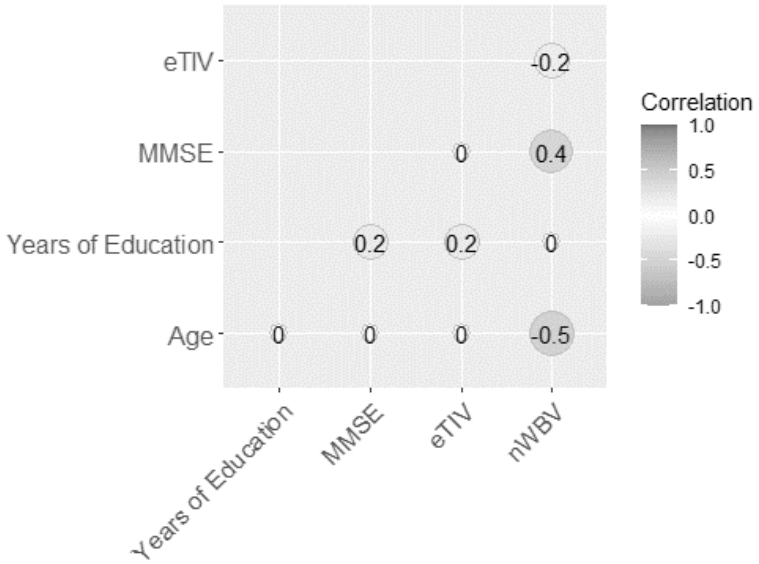


Figure 5. Correlogram of continuous Variables

Among all the numerical continuous variables used in this study, a correlogram was created, as shown in Figure 5. A correlation of 0 indicates that there is no correlation between the variables. Correlations greater than 0 indicate positive correlations, while correlations closer to 1 indicate higher degrees of positive correlation. Correlations less than 0 indicate negative correlations, while correlations closer to –1 indicate higher degrees of negative correlation. In this data set, nWBV and MMSE have a positive correlation of 0.4, and age and nWBV have a negative correlation of –0.5. MMSE and Years of Education,

as well as eTIV and Years of Education both exhibit slight positive correlations of 0.2. Finally, nWBV and eTIV have a slight negative correlation of 0.2. Other relationships between the continuous variables have little correlation in Figure 5.

3. Machine Learning Models

3.1 Logistic Regression

Logistic regression is one of the most widely used machine learning algorithms for solving a classification problem. It is used to predict the probability of a particular outcome, given a set of independent variables, which can be continuous, discrete, dichotomous, or a combination of these. In this study, the dependent variable used in the logistic regression model is the probability of dementia, and the independent variables include years of education, age, gender, eTIV, nWBV, and MMSE. Logistic regression is modeled by the equation below, where P – is the probability of dementia, β_0 is the intercept, β_1 through β_n are the regression coefficients, and x_1 through x_n are the independent variables.

$$\ln\frac{P}{1-P} = \beta_0 + \beta_1 x_1 + \ldots \beta_n x_n \qquad (1)$$

The left side of the equation is referred to as "the logit". The interpretation of the coefficients describes the independent variable's effect on the logit, rather than directly on the probability P. To facilitate interpretation, $e^{\beta n}$, a transformation of the original regression coefficient β_n, can be derived and interpreted as follows:

$$If\ e^{\beta n} > 1,\ P/(1-P) \text{ increases.}$$
$$If\ e^{\beta n} > 1,\ P/(1-P) \text{ decreases.}$$
$$If\ e^{\beta n} > 1,\ P/(1-P) \text{ stays the same.}$$

To build the logistic regression model for this study, 60% of the data were randomly chosen for model development (204 observations), while the remaining 40% of the data (158 observations) were used to test the model. The results of the logistic regression analyses are displayed in (Table 1).

Table 1. – Logistic Regression Analyses of Clinical MRI Data

Predictor	Estimate (β)	Standard Error	P	Odds Ratio (Exp(β))
Intercept	80.1293	15.2464	< 0.001	NA
nWBV	–31.1353	9.8578	0.0016	3.007e–14
eTIV	–0.0038	0.0018	0.0341	0.9962
Age	–0.1200	0.0451	0.0079	0.8869
MMSE	–1.3874	0.2386	< 0.001	0.2497
Years of Education	–0.2594	0.0959	0.0068	0.7715
Gender	1.4993	0.6039	0.013	4.4785

Above, the "Estimate" refers to the coefficient β. P values under 0.05 indicate that the predictor is statistically significant; lower P values indicate higher statistical significance. Thus, in this particular model, every predictors is considered statistically significant. The Odds Ratio is obtained by evaluating the natural exponential of β. According to the logistic analysis results listed in Figure 7, at a significance level of 0.05, the predictive model for Alzheimer's diagnosis is:

Predicted logit of dementia = 80.1293–31.1353 × × nWBV – 0.0038 × eTIV – 0.1200 × Age – 1.3874 × × MMSE – 0.2594 × Years of Education + 1.4993 × ×Gender

The coefficients of the parameters were interpreted as follows. At the significance level of 0.05:

- On average, controlling other variables, for 1 unit increase in nWBV, the odds of being demented is decreased by nearly 100%.
- On average, controlling other variables, for 1 unit increase in eTIV, the odds of being demented is decreased by 0.38%.
- On average, controlling other variables, for 1 year increase in age, the odds of being de-

mented is decreased by 11.31%. (Note that all ages are between 60 and 96.)

- On average, controlling other variables, for 1 unit increase in MMSE, the odds of being demented is decreased by 75.03%.
- On average, controlling other variables, for 1 year increase in Years of Education, the odds of being demented is decreased by 22.85%.
- On average, controlling other variables, female subjects are 347.85% more likely to be demented.

The model indicates that nWBV, eTIV, age, MMSE, years of education, and gender are significant predictors of Alzheimer's. Having a higher nWBV and being female increases the odds of being demented, whereas having a higher eTIV, a higher age, a higher MMSE score, and more years of education decreases the odds of being demented. In this particular model, it seems that an increase in age decreases the odds of being demented, which is contradictory to the general association that being older is correlated with a higher risk of Alzheimer's. This could be just a coincidence of the individuals who participated in the data set.

After the logistic regression model was analyzed, the model was tested with the testing data, which contained 158 observations. Of the 158 observations, 114 observations were of nondemented individuals, and 54 observations were of demented individuals. As shown in the confusion matrix (Table 2), 43 of the 54 observations of demented individuals were accurately predicted, resulting in a sensitivity of 79.63%. 94 of the 114 observations of nondemented individuals were accurately predicted, resulting in a 90.38% specificity. The overall accuracy rate of the logistic regression model is 86.71%.

Table 2. – Confusion Matrix of Logistic Regression Model Test

Actual Diagnosis	Predicted as Demented	Predicted as Nondemented	% Correct
Demented	43	11	79.63%
Nondemented	10	94	90.38%

Overall% Correct: 86.71%

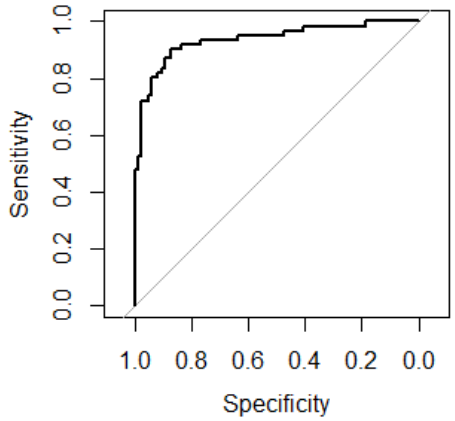


Figure 6. ROC Curve for Logistic Regression Model

In addition to a confusion matrix, an ROC curve plot (Figure 6) was created to visualize the performance of the logistic regression model on the testing data. The X axis shows the specificity or the proportion of observations that are actually demented that are correctly predicted as demented. The Y axis shows the sensitivity, or the proportion of observations that are actually nondemented that are correctly predicted as nondemented.

3.2 K Nearest Neighbor

K Nearest Neighbor (KNN) is a supervised machine learning algorithm that classifies a new data point based on its neighboring data points' features. KNN is a lazy algorithm, which means it memorizes the training data set rather than learning a discriminative function from the training data. It is also a non-parametric model, which means that it doesn't make any assumptions about the data set, thus making it more effective at handling real world data.

In this study, a KNN algorithm is used to determine whether an individual is demented based on independent variables, including the individual's years of education, age, gender, eTIV, nWBV, and MMSE. Each visit in the data set is considered a data point,

and each data point is plotted in *n*-dimensional space (where *n* is the number of independent variables), with the value of each independent variable being the value of a particular coordinate. Each data point is also categorized in a class, either demented or nondemented.

A KNN algorithm determines what class a new data point is by finding what class the majority of the *K* nearest data points are. The proximity between data points is calculated by using the Euclidean distance formula. As shown below, in the Euclidean distance formula, q_1 through q_n are the independent variables for data point *q*. Likewise, p_1 through p_n are the independent variables for data point *p*. The distance between data points *q* and *p* is given by Equation 2 below.

$$d = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \qquad (2)$$

The independent variables in the data set used in this study have different magnitudes. For example, MMSE is calculated as a whole number between 1 and 30, whereas nWBV is a decimal number between 0 and 1. Thus, the data is normalized in order to create a level playing field for all the variables in the data set. Below is the formula used for normalizing values in the data set, where *v* is the original value, *max* is the maximum of the values, *min* is the minimum of the values, and *t* is the normalized value given by Equation 3 below.

$$t = \frac{\upsilon - \min}{\max - \min} \qquad (3)$$

After normalization, all the transformed values were between 0 and 1. 60% of the data set of transformed values (217 observations) was randomly chosen for model development, and the remaining 40% of the data set (145 observations) was used to test the model. In order to optimize the accuracy of the KNN model, all possible values of K (from 1 to 216) were tested. Below is an accuracy plot showing the accuracy of the KNN model for values of K from 1 to 216.

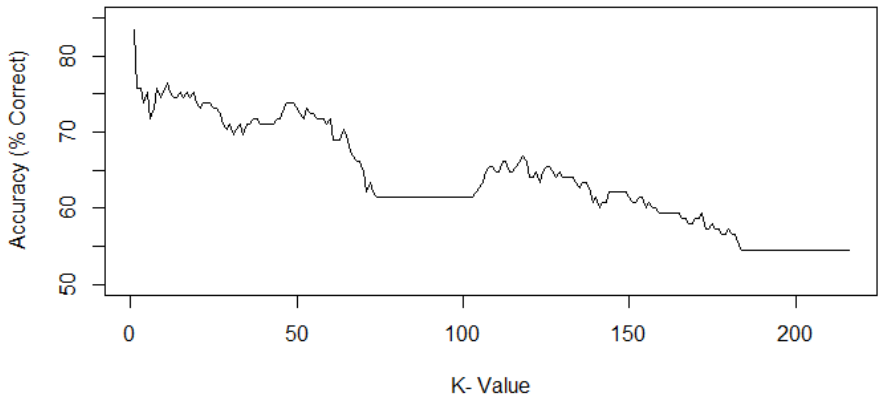


Figure 7. Accuracy Plot of KNN Algorithm

As shown in (Figure 7), the *K* value with the highest accuracy is 1. Thus, obtaining a new data point's class based on the class of its nearest data point yields the highest accuracy rate. A confusion matrix of the testing data tested with a KNN model with *K* value of 1 is shown in (Table 3). Of the 145 observations in the testing data, 87 observations were of nondemented individuals, and 58 observations were of demented individuals. As shown below, 50 of the 58 observations of demented individuals were accurately predicted, resulting in a 86.21% sensitivity. 71 of the 87 observations of nondemented individuals were accurately predicted, resulting in a 91.61% specificity. The overall accuracy rate of the KNN model with *K* value of 1 is 83.45%.

Like the logistic regression model, the performance of the KNN Model with *K* value of 1 can be visualized in a ROC curve plot, in which the x axis shows specificity and the y axis shows sensitivity, as shown in (Figure 8).

Table 3. – Confusion Matrix of KNN Model with K Value of 1

Actual Class	Predicted Demented	Predicted Nondemented	% Correct
Demented	50	8	86.21%
Nondemented	16	71	81.61%

Overall% Correct is 83.45%

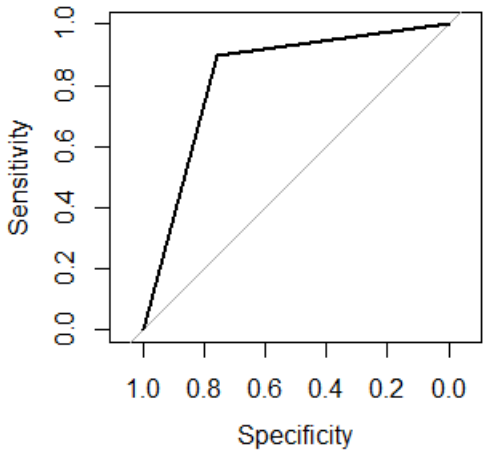


Figure 8. ROC Curve of KNN Model with K Value of 1

3.3 Support Vector Machine

A Support Vector Machine (SVM) is a supervised machine learning algorithm that is commonly used in classification models. In this study, SVM is used to classify individuals as either demented or nondemented based on independent variables, including the individual's years of education, age, gender, eTIV, nWBV, and MMSE. Each observation in the data set is plotted as a point in *n*-dimensional space (where *n* is the number of independent variables), with the value of each variable being the value of a particular coordinate. Each data point is also categorized as a class (which is the dependent variable being predicted), demented or nondemented. Classification is then performed by finding the hyperplane (a subspace whose dimension is $n-1$) that segregates the classes (demented and nondemented) in the best possible way.

If the training data is linearly separable, a hyperplane can be selected that best separates the two classes of data so that the distance (calculated by the Euclidean distance formula) between the hyperplane's two nearest distinct data points (known as the margin) is as large as possible. This is known as "hard-margin classification", in which no data points are allowed inside the margin. This type of classification is too stringent and sensitive to outliers. On the other hand, "soft-margin classification" allows certain data points to be inside the margin. The *c* value is used to tune how many data points are allowed inside the margin. If there are too many data points in the margin, the margin is too simple and does not adequately capture the underlying structure of the data. However, if there are too few or no data points in the margin, the separation may be influenced too greatly by the noise of the training data. Although the separation might be optimal from the training data, it would generalize poorly. As a result, the separation would be suboptimal for unseen data (eg. the testing data). The *c* value is defined as the weight of how much the samples inside the margin contribute to the overall value. With a low *c* value, samples inside the margins are penalized less than with a higher *c*. In other words, a lower *c* value allows for more data points in the margin than a higher *c* value. With a *c* value of 0, samples inside the margin are not penalized at all. An infinite *c* value is essentially hard margin classification, where no data points are allowed in the margin.

Like in previous models, 60% of the data (217 observations) was chosen for model development, and the remaining 40% of the data (145 observations) was used for testing the model. Multiple SVM models were created using the training data, with the following *c* values: 0, 0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.7, and 2. The accuracy of each c value was evaluated using 10 fold cross-sampling validation, in which 90% of the training data was randomly selected and used to train the model, while the remaining 10% of the training data was

used to validate the model. Note that this is not the final accuracy of the model, which will later be evaluated based on the model's performance on the testing data (40% of the original data). (Figure 9) shows the accuracy for each *c* value.

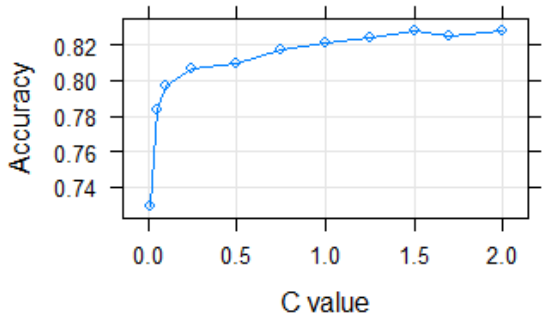


Figure 9. Accuracy of SVM Based on C Value

As shown in Figure 9, the *c* value with the highest accuracy is 1.5. The SVM model with *c* value 1.5 was then tested with the testing data (40% of the original data).

Table 4 below provides a confusion matrix indicating the performance of the model on the testing data. Of the 67 demented individuals, 54 were correctly predicted as demented, resulting in a sensitivity of 80.60%. Of the 78 nondemented individuals, 73 were correctly predicted as nondemented, resulting in a sensitivity of 93.59%. Overall, there was 87.59% accuracy when the SVM model with *c* value of 1.5 was evaluated using the testing data.

Table 4.– Confusion Matrix of SVM Model with *c* value of 1.5

Actual Class	Predicted Demented	Predicted Nondemented	% Correct
Demented	54	13	80.60%
Nondemented	5	73	93.59%

Overall Accuracy: 87.59%

Like previous models, the performance of the SVM model with *c* value of 1 can be visualized in a ROC curve plot. The x axis shows specificity, and the y axis shows sensitivity (Figure 10).

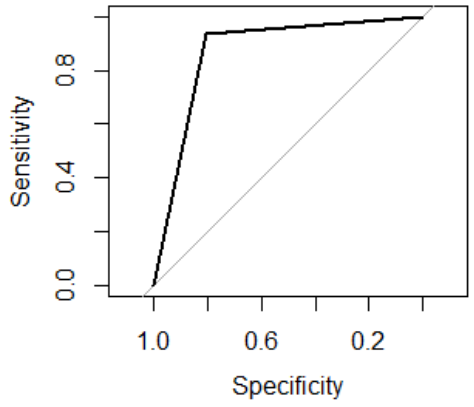


Figure 10. ROC Plot of SVM Model with *c* value of 1.5

3.4 Random Forest

The random forest algorithm is a supervised machine learning algorithm primarily used for classification. In this study, the random forest algorithm is used to classify individuals as demented or nondemented based on indepedent variables, including the individual's years of education, age, gender, eTIV, nWBV, and MMSE. Essentially, the random forest algorithm builds multiple decision trees (called a forest) and combines them to produce an accurate and stable prediction.

A decision tree is a tree-like model of decisions and their possible consequences. Each node in a decision tree represents a "test" on an independent variable (eg. a person's MMSE score is under 27), and each branch represents the result of the test. Each branch connects from the parent node (the node containing the test) to one of its child nodes, which either contains another test or is a leaf node. The leaf nodes (nodes without child nodes) represent the final classification result (in this study, demented or nondemented). In a decision tree, a decision is made by starting from the root node (the node without parent nodes) and descending until a leaf node, which contains the final decision, is reached. A decision tree can also be referred to as a "tree."

The random forest algorithm contains a large number of decision trees that operate as an ensemble. Each individual tree generates a decision containing the class prediction (in this study, demented or nondemented), and the most popular class prediction among all the trees is used as the final prediction of the algorithm.

Before creating the random forest model, 60% of the data (217 observations) was chosen for model development (referred to as training data), and the remaining 40% of the data (145 observations) was used for testing the model.

The first step in creating a random forest model is to create a bootstrapped data set. To create a bootstrapped data set, observations from the training data set were randomly selected. Note that observations in the bootstrapped data set can be repeated if selected multiple times. The second step is to build a decision tree based on the bootstrapped data set. Steps 1 and 2 are then repeated for each decision tree in the random forest model. Note that one third of the training data are left out of the bootstrap samples and are therefore not used to construct the trees. In this study, exactly 2.001 decision trees were built for the random forest model. Each decision tree was constructed using two independent variables, which were selected at random from the total of six independent variables.

Figure 11 shows the Out-of-bag (OOB) error and misclassification error rates for each independent variable based on the number of trees in the random forest model. The OOB error is a way of validating the random forest model. A higher OOB error indicates a higher prediction error. The misclassification error rate refers to the proportion of trees that misclassify a particular class. The OOB error and misclassification error rates are estimated using the one-third of the training data not used in the bootstrapping samples to test the model. The black color represents the OOB error, the red color represents the misclassification error for demented individuals, and the green color represents the misclassification error for nondemented individuals.

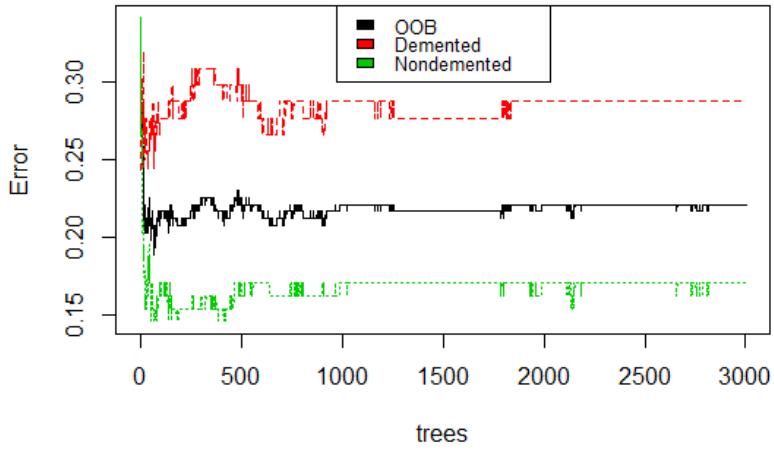


Figure 11. OOB and Misclassification Error Based on Number of Trees in Random Forest Model

As shown in Figure 11, once roughly 1,000 trees have been generated, the OOB and misclassification error rates stay relatively constant. In other words, once there are 1.000 trees in the random forest model, increasing the number of trees in the model does little to reduce the error rates. Figure 12 shows the mean decrease accuracy and mean decrease GINI Impurity (GINI) for each of the independent variables. The mean decrease accuracy estimates the loss in prediction performance when a particular variable is dropped from the training data. A higher mean decrease accuracy indicates greater loss in prediction performance when a variable is omitted.

The mean decrease GINI measures the variable's importance for estimating whether an individual is

demented or nondemented. An independent variable's mean decrease GINI is based on the average decrease of the impurity of nodes that test the independent variable. A node's impurity is the probability of obtaining two different decisions in each of the node's subtrees (trees that are children of the node). For example, if each subtree both have a 50–50 chance of outputting demented or nondemented, this indicates that the node has high impurity. However, if one subtree mostly outputs demented while the other subtree mostly outputs nondemented, the node has low impurity. Variables with lower mean impurity among nodes testing that variable have a higher mean decrease GINI.

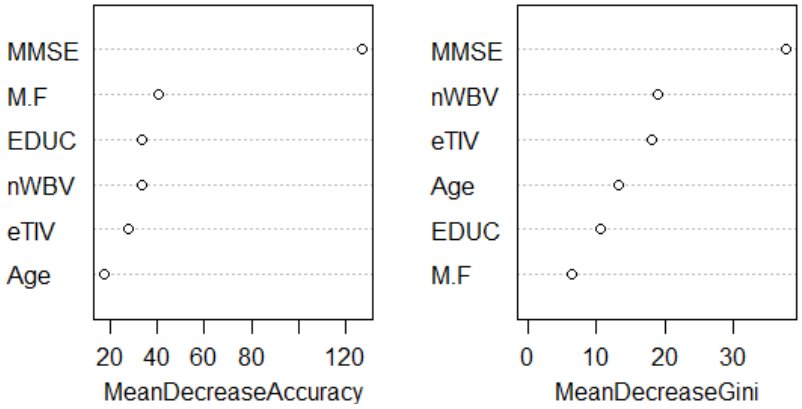


Figure 12. Mean Decrease Accuracy and Mean Decrease GINI of Independent Variables in Random Forest Model

Mean decrease accuracy and mean decrease GINI both measure the importance of each variable to the model. Higher values of mean decrease accuracy and mean decrease GINI both indicate that the variable is more important to the model. As shown in (Figure 12), MMSE was overwhelmingly the most important independent variable to the model in both the mean decrease accuracy and mean decrease GINI plots. Gender (referred to as M.F in Figure 12) has the second-highest mean decrease accuracy but the lowest mean decrease GINI. Years of Education (referred to as EDUC in Figure 12) has the third highest mean decrease accuracy but the second lowest mean decrease GINI. nWBV has the fourth highest mean decrease accuracy and the second highest mean decrease GINI. eTIV has the second lowest mean decrease accuracy but the third highest mean decrease GINI. Lastly, based on mean decrease accuracy, age is the least important variable, but based on mean decrease GINI, it is still more important than years of education and gender. Overall, although the importance of gender, years of education, nWBV, eTIV, and age varies based on mean decrease accuracy and mean decrease GINI, MMSE remains by far the model's most important variable.

Once the random forest model was developed, it was then evaluated using the testing data set. A confusion matrix (see Table 5) was created to analyze the performance of the model. Of the 64 demented individuals, 54 were correctly predicted as demented, resulting in a sensitivity of 84.38%. Of the 81 nondemented individuals, 77 were correctly predicted as nondemented, resulting in a sensitivity of 95.06%. Overall, there was 90.34% accuracy when the random forest model was evaluated with the testing data.

Table 5. – Confusion Matrix of Random Forest Model

Actual Class	Predicted Demented	Predicted Nondemented	% Correct
Demented	54	10	84.38%
Nondemented	4	77	95.06%

Overall Accuracy: 90.34%

Like previous models, the performance of the random forest model can be visualized in a ROC curve plot. The x axis shows specificity, and the y axis shows sensitivity (Figure 13).

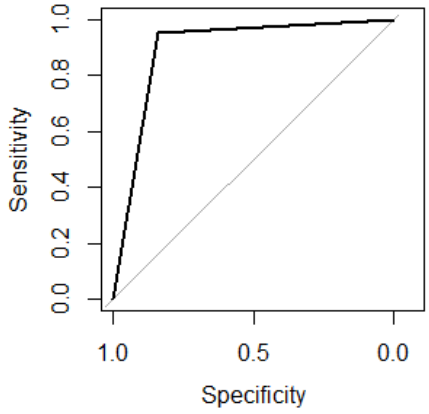


Figure 13. ROC Plot of Random Forest Model

3.5 Neural Network

Finally, the last machine learning method used in this study is a neural network. In this study, neural networks are multi-layer networks of neurons that are used to classify demented or nondemented individuals based on clinical and demographic independent variables. Every neural network has an input layer, at least one hidden layer, and an output layer. The input layer, the first layer, takes inputs based on the existing data. In this study, the input layer takes in the attributes for these independent variables: years of education, age, gender, eTIV, nWBV, and MMSE. The nodes in the hidden layer receive inputs from the input layer nodes, perform some computation, and then provide the output to the output layer. The output layer node contains the prediction of whether the individual is demented or nondemented based on the original input attributes.

Each of the nodes of each layer links to the nodes of the next layer through connections. Each connection has a weight, and each neuron (another term for node) has a bias and an activation function. The output of each node in the hidden layers and output layer can be represented by the following method. Let x_1 through x_i represent the input values of each node in the input layer. Let w_1 through w_i represent the weights of each connection going to the current node. Let m represent the number of connections that go from the activated nodes of the prior layer to the current node (or the number of nodes in the input layer if the prior layer is the input layer). Let b represent the bias of the current node. Below is the equation to obtain the result z for a particular node. Note that z is not the final output of the node.

$$z = \sum_{i=1}^{m} w_i x_i + b \tag{4}$$

After obtaining z, the final output of the node is simply determined by the activation function. In this study, the sigmoid function is used as the activation function. Below is the sigmoid function, where $f(z)$ the final output of the node with result z.

$$f(z) = \frac{1}{1+e^{-z}} \tag{5}$$

The output layer node has an output $f(z)$ that is a value between 0 and 1. If $f(z) \geq 0.5$, the final output of the neural network is 1, indicating that the individual is demented. If $f(z) \geq 0.5$, the final output of the neural network is 0, indicating that the individual is nondemented.

As in previous models, the data was randomly split into a training data set (containing 217 observations) and a testing data set (containing 145 observations). Before creating the neural network, the data was normalized in order to create a level playing field for all variables. Without data normalization, an independent variable may have a large impact on the dependent variable because of its scale, rather than its actual importance. The min-max normalization technique was used to normalize the data in this study, which was described in section 3.2.

Initially, each connection in the neural network was assigned random weights. For each observation, the input values were inputted into the input layer of the neural network and the final output was calculated. This is known as forward propagation. For each observation, after the forward propagation, backward propagation (which is essentially reinforcement learning) is used to modify the weights and biases in order to minimize the cost function.

Backward propagation starts from the output layers and moves through the model until it reaches the starting layer. In backward propagation, the error attributable to each neuron (as determined by the cost function), is calculated, starting from the layer closest to the output all the way back to the starting layer of the model. The cost function used in this study is cross-entropy, as shown below. Note that the cost function is calculated for each neuron. *C* represents the cost, in which the closer *C* is to 0, the closer the neuron's output is to the desired output for the neural network. *X* is the sum of all the input values. *N* is the total number of input nodes. *y* is the desired output for the particular neuron. *a* is the actual final output of the particular neuron.

$$c = -\frac{1}{n} X[y \ln \alpha + (1-y)\ln(1-\alpha)] \qquad (6)$$

Once the error attributable to each neuron is calculated, the bias and weight of the activated neurons are tweaked by the backpropagation process in a way that minimizes the overall cost function for each neuron.

Once forward propagation and backward propagation is completed for each of the observations in the training data, the neural network training is complete. In this study, neural networks of one hidden layer neuron and two hidden layer neurons were created. Below is a diagram representing the first neural network created, which contained only one hidden layer neuron.

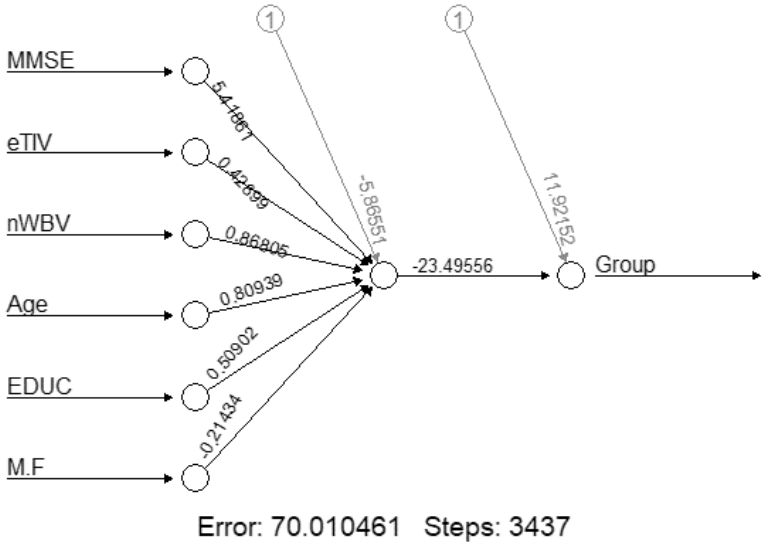


Figure 14. Plot of Neural Network with 1 Hidden Layer Node

Above, the black lines represent the connections between the nodes. A number, representing the weight, is assigned to each connection. The blue numbers indicate the bias of the particular neuron. It took the neural network 3,437 steps to converge, or reach a state in which the neural network has learned to properly predict an individual's dementia state with some margin of error. Below is a plot of the generalized weights with respect to each independent variable for the neural network with 1 hidden layer node.

"EDUC" and "M.F" refer to years of education and gender respectively.

In Figure 15, the distribution of generalized weights suggests that all the independent variables appear to have a nonlinear effect since the variance of their generalized weights is overall greater than one. The second neural network created in this study uses two hidden layer nodes. It uses the same activation function and cost function as the first neural network. Figure 16 shows a diagram representing the weights and biases of the neural network with two hidden layer nodes.

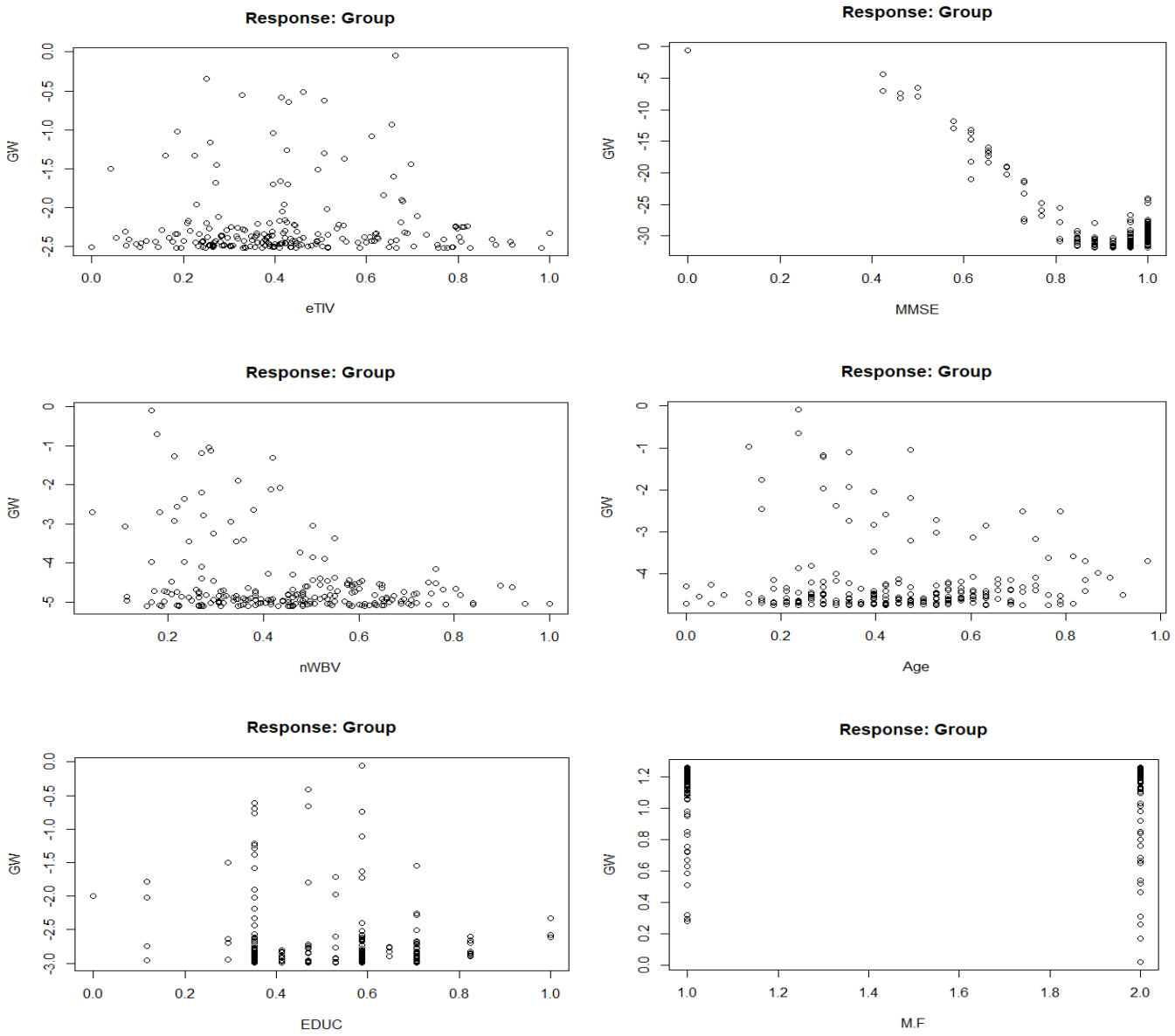


Figure 15. Plot of Generalized Weights with Respect to each Independent Variable for Neural Network with 1 Hidden Layer Node

Like in (Figure 14), the black lines represent the connections between the nodes, whereas the blue numbers indicate the bias of a particular node. Each of the black lines is assigned a numerical value for the weight. It took the neural network 58,078 steps to converge. Thus, the neural network with two hidden nodes required approximately 16.8 times the amount of steps to converge than the neural network with one hidden node, which indicates the training of the neural network with two hidden nodes involved significantly more backpropagation than the neural network with only one hidden node. Below is a plot of the generalized weights with respect to each independent variable for the neural network with two hidden layer nodes.

In Figure 17, the distribution of generalized weights suggests that all the independent variables appear to have a nonlinear effect since the variance of their generalized weights is overall greater than one. Although, at a first glance, MMSE and Age look

like they have a generalized weight of 0, the scale of the plot suggests that the points near 0 are actually above 1. Once the number of steps and distribution of generalized weights was calculated and generated for each of the neural networks, each of the neural networks was tested with the testing data, which contained 145 observations. Confusion matrices were created to evaluate the performance of each of the neural networks when tested using the testing data set.

Of the 145 observations in the testing data, 64 observations were of demented individuals while 81 observations were of nondemented individuals. For the neural network with one hidden layer node, 53 of the 64 observations of demented individuals were accurately predicted, resulting in a sensitivity of 82.82%. 68 of the 81 observations of nondemented individuals were accurately predicted, resulting in a specificity of 83.95%. The overall accuracy rate of the first neural network model is 83.44%. For the neural network with two hidden layer nodes, the training and testing data were split differently; the testing data for the neural network with two hidden nodes had 66 observations of demented individuals and 79 observations of nondemented individuals. 52 of the 66 demented individuals were accurately predicted, resulting in a sensitivity of 78.79%. 73 of the 79 nondemented individuals were accurately predicted, resulting in a specificity of 92.41%. The overall accuracy rate of the second neural network with 86.21%.

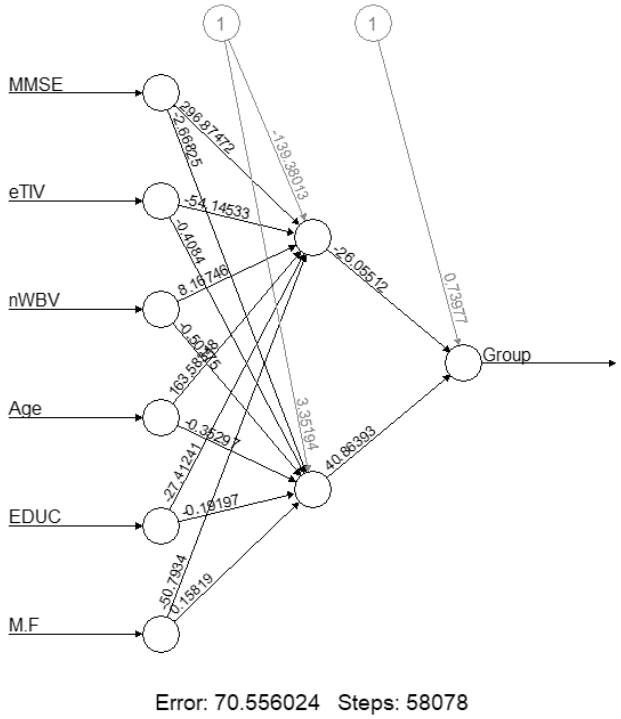


Figure 16. Plot of Neural Network with 2 Hidden Layer Nodes

Table 6. – Confusion Matrix of Neural Network with 1 Hidden Node

Actual Class	Predicted Demented	Predicted Nondemented	% Correct
Demented	53	11	82.82%
Nondemented	13	68	83.95%

Overall Accuracy: 83.44%

Table 7. – Confusion Matrix of Neural Network with 2 Hidden Nodes

Actual Class	Predicted Demented	Predicted Nondemented	% Correct
Demented	52	14	78.79%
Nondemented	6	73	92.41%

Overall Accuracy: 86.21%

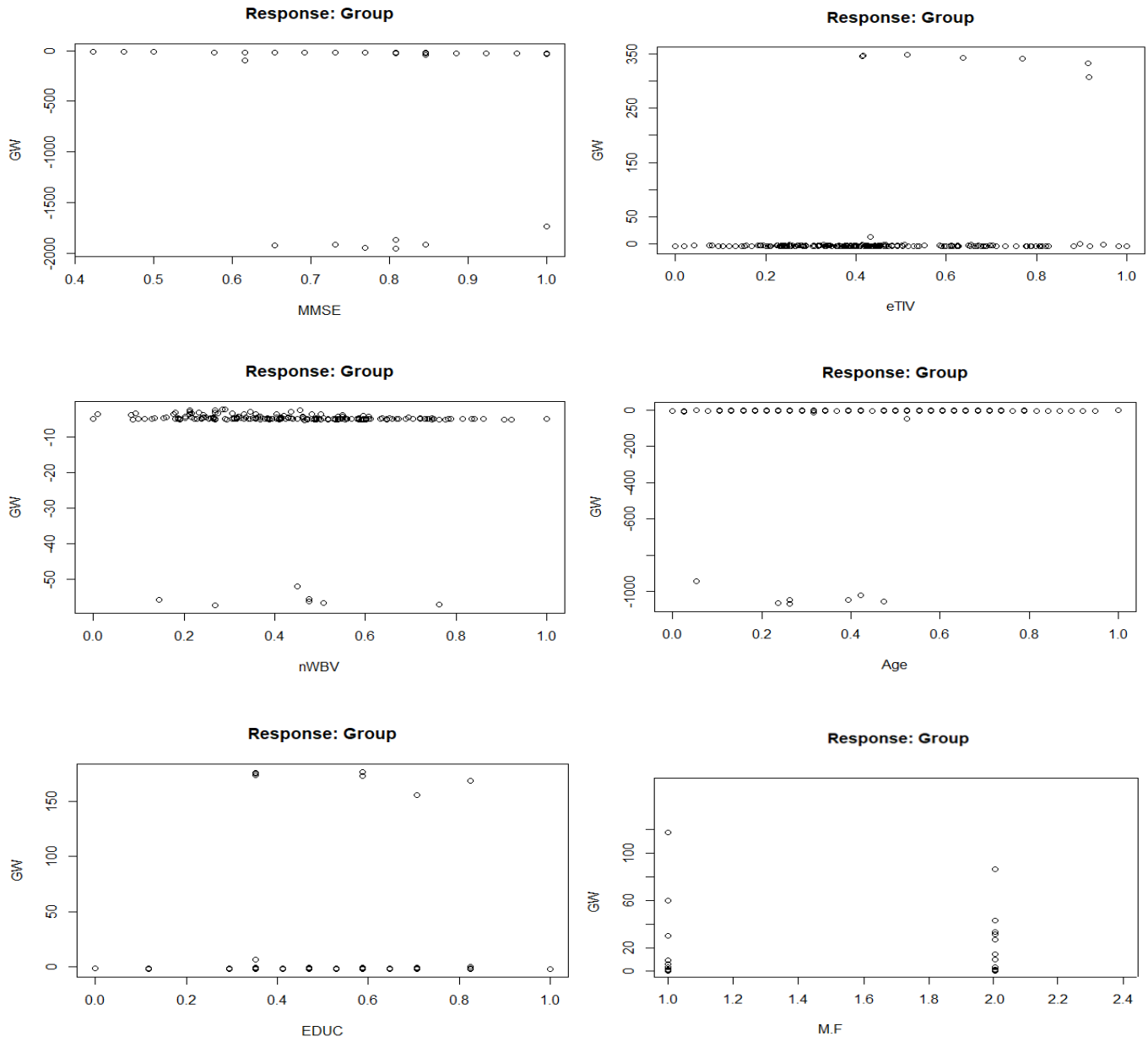


Figure 17. Plot of Generalized Weights with Respect to each Independent Variable for Neural Network with 2 Hidden Layer Nodes

4. Result

In this study, five types of machine learning models were developed to diagnose Alzheimer's based on clinical and demographic variables, including years of education, age, gender, MMSE, eTIV, and nWBV. In Section 3, the results of each machine learning model were evaluated and analyzed using a confusion matrix, and 4 of the 5 models were also evaluated using an ROC curve plot. Below is a plot comparing the highest accuracies of each of five types of machine learning models developed in this study.

As shown in Figure 18, all of the machine learning models are in the range of 83% to 91% accuracy. The random forest model has the highest accuracy, followed by the support vector machine, logistic regression, neural network, and KNN model in order of decreasing accuracy.

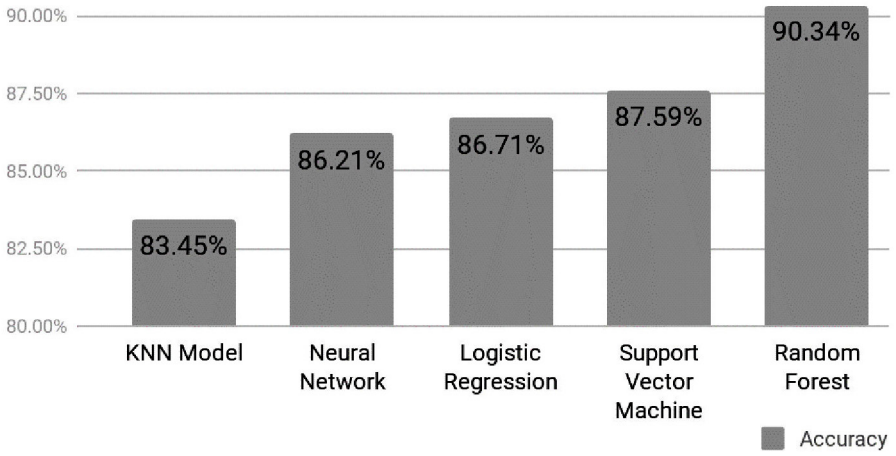


Figure 18. Comparison of Machine Learning model Accuracies

In addition to assess accuracy rate for each model, dataset were divided into the same training and testing sets to assess concordance (agreement on predictions) between various models. The concordance rate between each pair of models was provided in (Table 8).

Table 8.– Concordance Values Between Each Model

	Random Forest	**KNN Model**	**SVM Model**	**Neural Network**
Logistic Regression	91.78%	87.67%	93.15%	97.26%
Random Forest		93.15%	93.15%	91.78%
KNN Model			89.04%	87.67%
SVM Model				90.41%

Overall, the concordance rates are pretty good. The logistic regression and neural network models have the highest concordance value of 97.26%. The rest of the models have fairly high concordance values between 87.67% and 93.15%. Among all the models, the percent in which at least 4 of the 5 models shared the same diagnosis for a testing input was 90.42%.

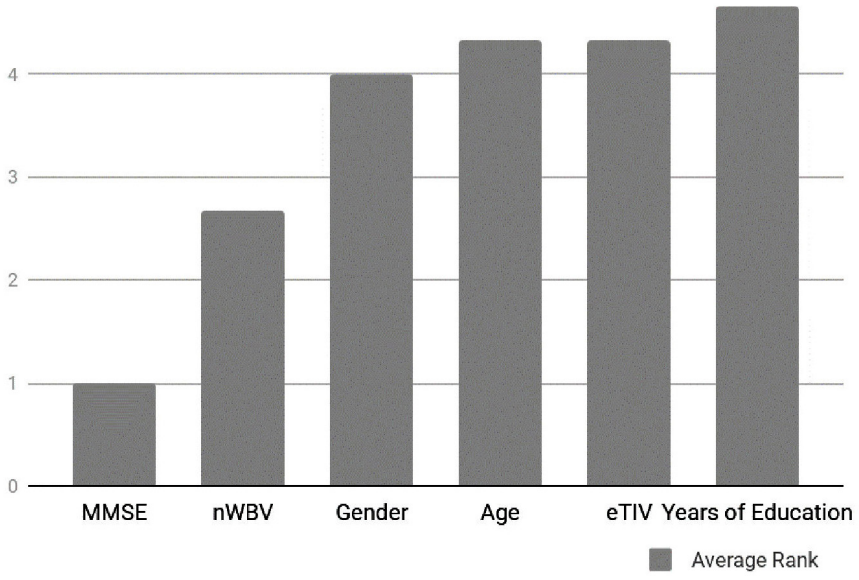


Figure 19. Rank of Importance of each Variable

Another model was created that outputted the result that the majority of the 5 models outputted. This model reached an accuracy rate of 89.04%, which is lower than the accuracy of the random forest model but higher than the accuracy of the other four models.

When creating the logistic regression and random forest models, the importance of variables were evaluated. In logistic regression, the *p* value was used to evaluate a variable's statistical significance, and in the random forest models, mean decrease accuracy and mean decrease GINI were used to evaluate the model's dependent on each variable. Below is a plot indicating the average rank (in terms of importance) among the different ways to evaluate variable importance for each variable. A rank of "1" indicates the greatest importance, whereas a rank of "6" indicates the least importance. A lower numerical value for rank indicates greater importance.

As shown in Figure 19, MMSE is clearly the most important variable by a considerable margin. In decreasing order of importance, MMSE is followed by nWBV, gender, age, eTIV, and years of education. The last four variables all seem to have similar levels of importance, whereas there is a considerable margin between the first three variables.

5. Discussion

The purpose of this research is to create a machine learning model to predict Alzheimer's using publicly available MRI data. There were 373 MRI imaging sessions in the data, with 150 distinct individuals among those sessions. In each session, the individual's CDR, eTIV, nWBV, age, gender, and years of education were collected and used to train the machine learning models. ASF and socioeconomic status were both part of the original data set but were omitted from the study because ASF was directly correlated with eTIV and because roughly 13% of all the imaging sessions did not have a recorded socioeconomic status.

In order to assess sensitivity and robustness of our findings, multiple methods were used to build a model that can predict Alzheimer's. These methods include Logistic Regression, K Nearest Neighbor, Support Vector Machine, Random Forest, and Neural Network. These methods all had fairly consistent accuracy rates, as all the accuracy rates were within an 8% range. MMSE was consistently identified as the most important predictor, while other variables varied in importance.

In another similar study conducted by the French National Institute of Health and Medical Research, researchers presented and evaluated a new method based on SVM to diagnose patients with Alzheimer's using MR images. Using their new method based on SVM, the researchers were able to achieve a 94.5% accuracy in detecting Alzheimer's by bootstrap resampling a dataset containing 16 individuals with AD and 22 controls. In this study, rather than directly using raw MR images, MRI biomarkers along with cognitive and demographic data were used to classify individuals as either AD or non-AD. The SVM method in this study achieved a 87.59% accuracy, which is 7% lower than the accuracy of the new SVM method created by the researchers. The best accuracy in this study, 90.34%, is still roughly 4% lower than the new method created by the researchers. However, it's important to note that the sample size of the data set used by the researchers is roughly 10 times smaller than the data set used in this study [13].

Compared with the clinical diagnosis rate of Alzheimer's among rural doctors in the United States, all the models presented in the study are significantly better. The accuracy rate of diagnosing Alzheimer's among community doctors in rural areas is about 50 to 60 percent [5]. The model with the lowest accuracy presented in this study, KNN model, still does about 40% better than rural doctors in terms of accuracy. The model with the highest accuracy, Random Forest, does about 50% better than rural doctors. Moreover, the clinical diagnosis rate of Alzheimer's is estimated to be 78% among all doctors in The United States [6]. The worst model in the study still does about 7% better in terms of accuracy whereas the best model does about 15% better in terms of accuracy. Thus, the models created in this

study have the potential to aid doctors to create more accurate clinical diagnoses of Alzheimer's.

This study can be improved by using a greater sample size of data and utilizing more MRI biomarkers and other independent variables. In addition, patients can have longer and more frequent follow-up MRI imaging sessions. The models can be improved by also accounting for patients' changes over time, rather than from one-time imaging sessions.

6. Conclusion

In conclusion, a variety of models were utilized to predict Alzheimer's using MRI biomarkers, cognitive assessment data, and demographic data from 371 imaging sessions. All models provide reasonable accuracy ranging from (83% to 91%) in predicting dementia. Among all the variables tested, MMSE was the most influential variable, while other variables also influenced the detection of Alzheimer's at a lesser degree. Using the machine learning models in the study, physicians or nurses can potentially more accurately diagnose Alzheimer's in patients. An accurate diagnosis of Alzheimer's is important because it allows patients to better devise strategies to manage symptoms and plan for long-term care.

References:

1. Cobb B. R., Wells K. R., & Cataldo L. J. Alzheimer's Disease. In K. Key (Ed.), The Gale Encyclopedia of Mental Health (3rd ed., 2012.– Vol. 1.– P. 59–73). Detroit, MI: Gale. Retrieved from: URL: https://link.gale.com/apps/doc/CX4013200025/GPS?u=watchunghrhs& sid=GPS&xid=9f274526
2. Martone R. L. & Piotrowski N. A., Ph D. Alzheimer's disease. Magill's Medical Guide (Online Edition). 2019.
3. Alzheimer's Disease. Funk & Wagnalls New World Encyclopedia, 1; 2018.
4. Yiğit A. & Işik Z. Applying deep learning models to structural MRI for stage prediction of Alzheimer's disease. Turkish Journal Of Electrical Engineering & Computer Sciences, 28(1), 2020.– P. 196–210. Doi:10.3906/elk-1904–172
5. Kolata G. A Blood Test for Alzheimer's? It's Coming, Scientists Report. 2019. August 01. Retrieved from: URL: https://www.nytimes.com/2019/08/01/health/alzheimers-blood-test.html
6. Reinberg S. (2016. July 26). 2 in 10 Alzheimer's Cases May Be Misdiagnosed. Retrieved from: URL: https://www.webmd.com/alzheimers/news/20160726/2-in-10-alzheimers-cases-may-be-misdiagnosed
7. Warren J. D., Schott J. M., Fox N. C., Thom M., Revesz T., Holton J. L., Scaravilli F., Thomas D. G. T., Plant G. T., Rudge P., Rossor M. N. Brain biopsy in dementia, Brain,– Vol. 128.– Issue 9. September, 2005.– P. 2016–2025. URL: https://doi.org/10.1093/brain/awh543
8. Marcus D. S., Fotenos A. F., Csernansky J. G., Morris J. C. & Buckner R. L. Open Access Series of Imaging Studies: Longitudinal MRI Data in non-demented and Demented Older Adults. Journal of Cognitive Neuroscience, 22(12), 2010.– P. 2677–2684. Doi:10.1162/jocn.2009.21407
9. Frisoni G. B., Fox N. C., Jack C. R., Scheltens P. & Thompson P. M. The clinical use of structural MRI in Alzheimer disease. Nature Reviews Neurology, 6(2), 2010.– P. 67–77. Doi:10.1038/nrneurol.2009.215
10. Sargolzaei S., Sargolzaei A., Cabrerizo M., Chen G., Goryawala M., Noei S., Adjouadi M. A practical guideline for intracranial volume estimation in patients with Alzheimers disease. BMC Bioinformatics, 2015.– 16(S7). Doi:10.1186/1471–2105–16-s7-s8
11. Khan T. Biomarkers in Alzheimers disease. – Amsterdam: Academic Press. 2016.
12. Medical Tests. (n.d.). Retrieved from: URL: https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests

13. Magnin B., Mesrob L., Kinkingnéhun S. et al. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51, 2009.– P. 73–83. URL: https://doi.org/10.1007/s00234–008–0463-x

Section 5. General biology

https://doi.org/10.29013/ELBLS-20-4-68-71

Guliyev Mahir Isa,
Senior teacher of Department of Histology,
Cytology and Embryology;
Azerbaijan Medical University
E-mail: mahir-quliyev-65@mail.ru

Israfilova Sabina Aliaga,
assistant of Department of Histology,
Cytology and Embryolog;
Azerbaijan Medical University
E-mail: israfilovi0506@mail.ru

Aliyarbekova Aygun Aliyar,
assistant of Department of Histology,
Cytology and Embryology;
Azerbaijan Medical University.
E-mail: alyarbayova@gmail.com

THE DEFINITION OF EFFECTIVE DOSE OF "SUMAKH FRUIT EXTRACT" FOR CORRECTION THE GENOTOXICITY OF CHEMICALS IN THE PROCESS OF ARTIFICIAL MUTATION IN PLANTS

Abstract: To determine the effective dose of sumakh fruit extract as a modifier during artificial mutagenesis, dry onion seeds were soaked for 3 hours in a test dose of a solution with sumac fruit extract, then either MNG (5 mcg/ml) or NMM (0.02%) was added to this medium. Doses of mutagens were selected by calculating aquatoxicity their mutagenic effect.

From the results of the experiment, it was found that when exposed to MNG and NMM, a dose of sumac fruit extract equal to 0.01 mcg/ml most effectively protects the genome.

Keywords: correction the genotoxicity of chemicals, antimutagens, sumakh fruit extract.

Гулиев Махир Иса оглу,
ведущий преподаватель кафедры
Гистологии, цитологии и эмбриологии;
Азербайджанский медицинский университет
E-mail: mahir-quliyev-65@mail.ru
Исрафилова Сабина Алиага кызы,
ассистент кафедры,
Гистологии, цитологии и эмбриологии,
Азербайджанский медицинский университет
E-mail: israfilovi0506@mail.ru
Алиярбекова Айгюн Алияр кызы
ассистент кафедры
Гистологии, цитологии и эмбриологии,
Азербайджанский медицинский университет
E-mail: alyarbayova@gmail.com

ОПРЕДЕЛЕНИЕ ЭФФЕКТИВНОЙ ДОЗЫ «ЭКСТРАКТА ИЗ ПЛОДОВ СУМАХА» ДЛЯ КОРРЕКЦИИ ГЕНОТОКСИЧНОСТИ ХИМИЧЕСКИХ ВЕЩЕСТВ ПРИ ПРОЦЕССЕ ИСКУССТВЕННОЙ МУТАЦИИ У РАСТЕНИЙ

Аннотация: Для определения эффективной дозы экстракта из плодов сумаха, как модификатора при искусственном мутагенезе сухие семена репчатого лука 3 часа были вымочены в испытуемой дозе раствора с экстрактом из плодов сумаха, далее к данной среде добавлялся либо МННГ (5 мкг/мл), либо НММ (0,02%). Дозы мутагенов были выбраны путем вычисления эквитоксичности их мутагенного воздействия. Из результатов опыта выяснилось, что при воздействии МННГ и НММ доза экстракта из плодов сумаха равная 0,01 мкг/мл наиболее эффективно защищает геном.

Ключевые слова: коррекция генотоксичности химических веществ, противомутагены, экстракт из плодов сумаха.

Начиная со второй половины прошлого века ведется поиск корректоров, способных противостоять вредному воздействию мутагенных и канцерогенных веществ в процессах мутаций синтетического и натурального происхождения. В данном направлении особого внимания заслуживают натуральные вещества растительного происхождения.

В данном исследовании была произведена апробация широкого диапазона дозы «экстракта из плодов сумаха» как модификатора искусственного мутагенеза, сохранившего натуральные вещества растительного происхождения, а так же определение его наиболее эффективной дозы.

Для объективной оценки его эффективности в защите генома, была изучена противомутагенно активная доза экстракта из плодов сумаха при воздействии мутагенных веществ, различающихся по природе, типу, по механизму взаимодействия с наследственными субстратами.

Таблица 1. Определениепротивомутагенной активности экстракта из плодов сумаха в клетках апикальной меристемы ростков семян репчатого лука

Мутагены	Варианты опыта	Доза экстракта, мкг/мл	Все исследуемые клетки	Частота аберрации хромосом		td		P		ФЭП
				n	M ± m	Соответственно контролю	Соответственно мутагену	Соответственно контролю	Соответственно мутагену	
Контроль		0	925	32	3,46 ± 0,60	–	–	–	–	–
Метилнитронитрозо-гуанидинМ метилнитронитрозогуанидин	МННГ	0	911	105	11,53 ± 1,06	9,53	–	<0,001	–	–
	Экстракт сумаха + МННГ	0,001	854	54	6,32 ± 0,83	2,80	3,89	<0,01	<0,001	0,45
		0,01	893	46	5,15 ± 0,74	1,78	4,95	>0,05	<0,001	0,55
		0,1	876	57	6,51 ± 0,83	2,90	3,75	<0,01	<0,001	0,44
		1,0	902	64	7,09 ± 0,85	3,49	3,26	<0,001	<0,01	0,39
		10	878	69	7,86 ± 0,91	4,07	2,64	<0,001	<0,01	0,32
		100	881	75	8,51 ± 0,94	4,55	2,14	<0,001	<0,05	0,26
Нитрозометилмочевина	НММ	0	906	112	12,36 ± 1,09	7,18	–	<0,001	–	–
	Экстракт сумаха + НММ	0,001	901	55	6,10 ± 0,80	2,64	4,64	<0,01	<0,001	0,51
		0,01	887	49	5,52 ± 0,77	2,12	5,14	<0,05	<0,001	0,55
		0,1	911	61	6,69 ± 0,83	3,16	4,14	<0,01	<0,001	0,46
		1,0	896	66	7,37 ± 0,87	3,69	3,56	<0,001	<0,001	0,40
		10	884	73	8,26 ± 0,93	4,36	2,87	<0,001	<0,01	0,33
		100	897	82	9,14 ± 0,96	4,98	2,21	<0,001	<0,05	0,26

Опыты были проведены на свежих семенах репчатого лука. В качестве мутагена использовался МННГ (метилнитрозогуанидин) и НММ (нитрозометилмочевина). На данном этапе опыта испытуемые дозы экстракта добавлялись к объекту до мутагена.

В опытах, проведенных на растительных объектах сухие семена репчатого лука, 3 часа были вымочены в растворе экстракта из плодов сумаха, далее к данной среде добавлялся либо МННГ (5 мкг/мл), либо НММ (0,02%). После того как остатки мутагена промывались семена, до окончания фиксации были инкубированы в воде. Далее был проведен анализ частоты хромосомной аберрации апикальной меристемы клеток ростков.МННГиНММ являясь моно- и полифункциональными алкилирующими соединениями, являются мутагенами прямого типа, т.е. в начальной форменаходятся во взаимодействии с ДНК-мишенью.

При выборемоделей мутагенов, так же учитывался спектр образуемых ими типов первичного повреждения молекул ДНК. Т. о. МННГ и НММ являясь донорами алкильных групп, образуют первичные повреждения нуклеотидного типа.

Исследования экстракта из плодов сумаха в широкого диапазона дозе на семенах репчатого лука показало, что применение экстракта до мутагена понижает генотоксичностьалкилирующих мутагенов (табл. 1). В данной серии опытов выяснилось, что при воздействии МННГ (5мкг/мл) иНММ (0,02%) (дозы были выбраны путем вычисления эквитоксичности их мутагенного воздействия) использование экстракта сумаха в дозе 0,01мкг/мл защищает геном в максимальной степени.

M – частота хромосомной аберрации, *N* – все изученные клетки, *n*- хромосомная абберация, *m* – ошибка, *tg* – погрешность, *P* – уровень значимости.

Способ вычисления:

$$M = \frac{n * 100\%}{N};\ M = \sqrt{\frac{M * (100 - M)}{N}};$$

$$tg = \frac{M2 - M1}{\sqrt{m1^2 + m2^2}}.$$

M – частота мутаций; M_2 – частота мутаций опытного варианта;

M_1 – частота мутаций контрольного варианта; M_1^2 – ошибка варианта контроля; M_2^2 – ошибка варианта опыта.

ФЭП – $\frac{i-c}{i}$. ФЭП – Фактор эффективности противомутагена, i- первичный (предыдущий), с – последующий (определяется при помощи деления разности первичного и модифицированного уровней мутации на первичный показатель).

Список литературы:

1. Агабейли Р.А. Антиоксиданты и антиоксидантные ферменты в регуляции мутационного процесса. -Дис.д.н.б.– Баку, 1991.–256 с.
2. Furukawa H., Vegetables as antimutagenes \\ Bull of Genetics Society of Canada.1994. – V. 25. – № 1. – 16 p.
3. Gichner T., Pospisil F., Volkeova V. et. al. gallic and tannic acids inhibit the mutagenicity of a direct acting mutagen MNNG \\ Biol. Plant. 1986. – V. 28. – № 5. – P. 386–390.
4. Grüter A., Friederich U., Würgler F. E. Antimutagens in mushrooms \\ Mutat. Res. Environ. Mutagenes and Related Subj.1987. – V. 182. – № 5. – 281 p.
5. Grover J.S., Bala S. Antimutagenic activity Terminalia chebula (myroblan) is Salmonella typhimirium \\ Indian J., Exp. Biol. 1992. – V. 30. – № 4. – P. 339–341.

https://doi.org/10.29013/ELBLS-20-4-72-78

William Hou,
High School Student
Ivy Mind Academy, Skillman, New Jersey
E-mail: william.hou9@gmail.com
Betty Wang,
Employment: IvyMind Academy
E-mail: betty.wang@ivymind.org

IDENTIFYING GENE MUTATIONS MOST COMMONLY ASSOCIATED WITH GLIOBLASTOMA

Abstract. Certain gene mutations are often associated with the development and presence of glioblastoma (GBM). The most commonly associated gene mutants with GBM are found to be those of tumor protein p53 (TP53), phosphatase and tensin homolog (PTEN), and epidermal growth factor receptor (EGFR), all of which have a major role in regulating cell proliferation. TP53 had a mutation frequency per kilo base (M/kb) of 28.8, PTEN of 10.85, and EGFR of 7.25. Somatic mutations in specific regions of these genes have been correlated with lower survival rates in GBM patients often due to irregularities in the functions of the genes' proteins. These regions include the DNA-binding domain of TP53, the C2 domain of PTEN, and amino acid 289 in EGFR. Mutations occurring in such regions can alter the genes' proteins; altered proteins with inhibited functions cannot regulate cell proliferation to the same extent, boosting the oncogenesis of GBM and leading to poorer prognoses. EGFR and PTEN are also associated with focal adhesion shown through an enrichment analysis, further suggesting that mutations in these genes lead to inhibited abilities to trigger cell apoptosis, a function that is critical for regulating cell proliferation. The majority of deleterious mutations are missense and truncating, suggesting that these types of mutations have the largest impact on GBM survival rates. Further research into TP53, PTEN, and EGFR may bring new treatments targeting these genes, allowing for a better prognosis in GBM patients and potentially other cancers.

Keywords: glioblastoma (GBM), gene mutation, bioinformatics analysis.

Introduction

GBM is a lethal, aggressive brain tumor, making up a large portion of all brain cancers [1]. While treatments are available for GBM, the prognosis is poor, and the cancer occurs more frequently in older patients at a median age of 64 [2]. Treatments include surgery, radiation, and chemotherapy, but there is currently no cure for glioblastoma [1]. The exact cause of GBM is unknown, as most cases occur irregularly and are not inherited [1]. This study is designated to find genes whose mutations are commonly associated with the development of GBM and encourage further research into these select genes for a better understanding of the tumor. Additionally, the methodology used may be applied to data sets for other diseases to determine genes most commonly associated with those diseases.

The data used for this analysis were taken from XenaBrowser and were originally from The Cancer Genome Atlas (TCGA), a program dedicated to cataloguing genetic mutations associated with various cancers. The data set consists of 314 samples

from GBM patients. Out of the 314 samples taken, only genes with the highest M/kb are used in the analysis because of their higher correlation with GBM.

Methods

The data taken from XenaBrowser are analyzed to find genes with the most frequent mutations [3], and is sorted using Linux commands to find the most frequently appearing genes in the data. These genes and their frequencies are tabulated in Microsoft Excel, with their respective gene lengths and full names. The gene lengths are taken from GenCodeGenes [4], and the genes' full names are matched with the gene symbols. Once the gene lengths and full names are matched, the frequency of each gene is then normalized with its gene length, and the genes are sorted by the highest frequency of mutations per base length (M/kb). Genes with less than 1 M/kb were removed from the analysis due to their relative insignificance and for conciseness, and only genes with an M/kb of 4 or more are tabulated. Data of only genes with more than 1 M/kb are extracted when analyzing the distribution of mutation types. The extracted data are sorted using Linux commands to determine the frequency of each mutation type, and the results are tabulated in Microsoft Excel.

A TCGA Glioblastoma (GBM) study is selected in CBioPortal [5], and is matched with gene symbols TP53, PTEN, and EGFR, which have some of the highest M/kb. A mutation chart is created from the gene symbols inputted, which marks locations of mutations in GBM patients onto each gene.

Detailed analyses of the top genes are conducted using past studies that associate the genes with GBM. An enrichment analysis of genes with the most M/kb is done through DAVID Bioinformatics Resources in order to find associated terms with the genes, linking the genes' mutations to the impairment of their functions [6; 7]. The top genes are submitted as a gene list into DAVID, and a functional annotation chart is created from the list, giving the most associated terms with the genes.

Results

Three of the genes with the most mutations per kilo base (M/kb) are found to be TP53, PTEN, and EGFR. After normalizing the frequency of mutations for every kilo base in each gene, 329 out of the 15,350 genes found in the data set had one or more mutations per kilo base. For these 329 genes, the vast majority of mutations are missense, which changes a single amino acid in the coded protein [8], and makes up 80.2% of all mutations (Table 1).

Table 1. – Frequency of Mutation Types in Top 329 Normalized Genes

Mutation Type	Frequency
Missense	3927
Nonsense	337
3'UTR	115
Splice Site	112
Frame Shift Del	98
RNA	87
Intron	87
5'UTR	55
In Frame Del	28
Frame Shift Ins	24
3'Flank	11
In_Frame Ins	5
5'Flank	5
Translation Start Site	3
Nonstop Mutation	2

However, though only 92 genes were found to have an M/kb of 2 or higher, genes with a lower frequency may still influence GBM. The top 92 genes are sorted, and closer analysis is done on genes TP53, PTEN, and EGFR because of their high value of M/kb (Table 2). Despite that POM121L12 has the third highest M/kb (Table 2), the gene is not analyzed because of its relatively low frequency compared with TP53, PTEN, and EGFR. TP53 had an M/kb of roughly 28.8, the highest from the data provided by TCGA (Table 2), while PTEN and EGFR had an M/kb of 10.85 and 7.25 respectively (Table 2). The majority of TP53 mutations are point mutations that occur in the DNA-binding domain (Fig-

ure 1). A number of PTEN mutations occur in the C2 domain (Figure 2), and a significant amount of EGFR mutations occur on amino acid 289, with 18 mutations out of 67 samples (Figure 3). Additionally, an enrichment analysis through DAVID Bioinformatics Resources found that the top 329 genes with the highest frequency per base length are highly associated with focal adhesion and ECM-receptor interaction, with p-values of 5.75E-10 and 1.08E-06 respectively (Table 3).

Table 2. – Frequency of Gene Mutations per Kilo Base Length

Gene Symbol	Gene Name	Frequency	Length	Frequency per Kilo Base Length (M/kb)
TP53	tumor protein p53	113	3924	28.797145770
PTEN	phosphatase and tensin homolog	109	10048	10.847929936
POM121L12	POM121 transmembrane nucleoporin-like 12	10	1269	7.880220646
EGFR	epidermal growth factor receptor	94	12961	7.252526811
TRIM51	tripartite motif-containing 51	12	1832	6.550218341
TUBA3C	tubulin, alpha 3c	9	1551	5.802707930
KIF2B	kinesin family member 2B	13	2335	5.567451820
DCAF12L2	DDB1 and CUL4 associated factor 12-like 2	9	1673	5.379557681
RB1	retinoblastoma 1	30	6169	4.863024801
TPTE2	transmembrane phosphoinositide 3-phosphatase and tensin homolog 2	11	2288	4.807692308
OR8K3	In multiple Geneids	9	1878	4.792332268
NLRP5	NLR family, pyrin domain containing 5	19	4023	4.722843649
ZFP42	ZFP42 zinc finger protein	12	2651	4.526593738
MAGEC2	melanoma antigen family C2	9	1991	4.520341537
UGT2B28	UDP glucuronosyltransferase 2 family, polypeptide B28	8	1833	4.364429896
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha	40	9411	4.250345341

Table 3. – Associated Terms with Top 329 Normalized Genes for Glioblastoma

Category	Term	Count	%	PValue
1	*2*	*3*	*4*	*5*
KEGG_PATHWAY	hsa04510:Focal adhesion	22	0.04931077	5.75E-10
KEGG_PATHWAY	hsa04512:ECM-receptor interaction	12	0.026896784	1.08E-06
KEGG_PATHWAY	hsa04151:PI3K-Akt signaling pathway	23	0.051552169	1.14E-06
KEGG_PATHWAY	hsa05214:Glioma	10	0.022413986	5.00E-06
KEGG_PATHWAY	hsa05222:Small cell lung cancer	11	0.024655385	6.62E-06
KEGG_PATHWAY	hsa05146:Amoebiasis	12	0.026896784	7.77E-06
KEGG_PATHWAY	hsa05218:Melanoma	10	0.022413986	1.05E-05
GOTERM_BP_DIRECT	GO:0030198~extracellular matrix organization	16	0.035862378	1.87E-06

1	2	3	4	5
GOTERM_BP_DIRECT	GO:0006936~muscle contraction	12	0.026896784	2.66E-06
GOTERM_BP_DIRECT	GO:0007156~homophilic cell adhesion via plasma membrane adhesion molecules	14	0.031379581	4.12E-06
GOTERM_BP_DIRECT	GO:0051209~release of sequestered calcium ion into cytosol	8	0.017931189	6.20E-06

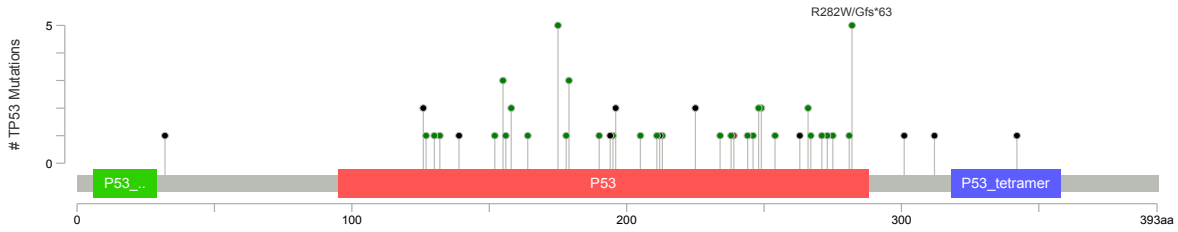


Figure 1. Locations of TP53 Mutations in Glioblastoma Patients (taken from CBioPortal)

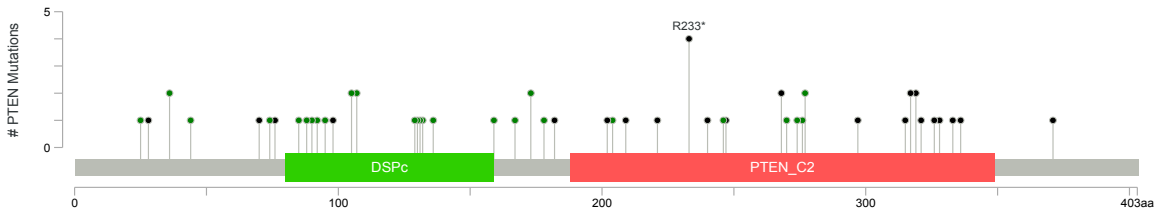


Figure 2. Locations of PTEN Mutations in Glioblastoma Patients (taken from CBioPortal)

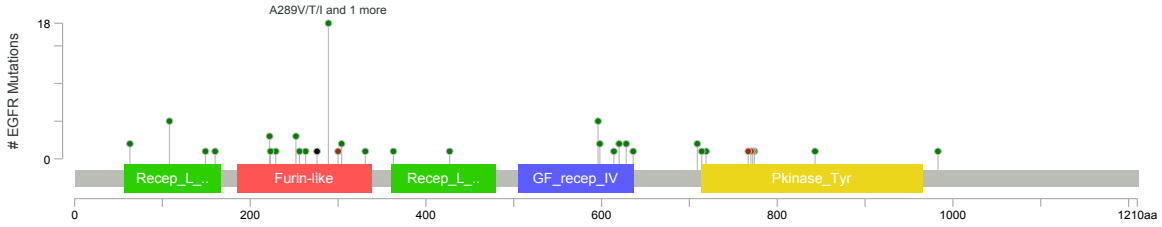


Figure 3. Locations of EGFR Mutations in Glioblastoma Patients (taken from CBioPortal)

TP53 Association with Glioblastoma

The main function of TP53 is to regulate cell division by creating tumor protein 53 (p53), which prevents uncontrollable cell reproduction and is done by triggering cell apoptosis or other genes to repair a cell's DNA. Half of all cancers are found to have somatic mutations in TP53, most of which changes p53 so that the protein is no longer able to regulate cell division, leading to tumor developments. Because of a reduced ability to control cell proliferation, modified p53 proteins are often linked not only to GBM, but a wide variety of other cancers; according to Wang et al. (2013), they may also increase GBM resistance to temozolomide and therefore result in ineffectiveness of certain treatments involving such a drug (Wang et al, 2013).

Out of the 314 GBM samples provided by XenaBrowser [3], 113 TP53 mutations were found (Table 2). Végran et al. (2013) found that 90% of TP53 mutations occur in the TP53 DNA binding to domain, but only missense mutations usually significantly hinder p53 from binding DNA [10].

Similarly, roughly 93% of TP53 mutations provided by CBioPortal occur in the DNA binding domain with all mutations being either missense or truncating (Figure 1). Proper DNA binding is essential for protein p53 to perform its functions, including triggering apoptosis, and impaired DNA binding of p53 was linked to higher cancer rates in mice [11].

PTEN Association with Glioblastoma

PTEN is responsible for a phosphatase enzyme that controls cell division and contributes to the triggering of apoptosis; consequently, the PTEN phosphatase acts as a tumor suppressor. As such, mutations of PTEN often modify the PTEN phosphatase so that the enzyme can no longer perform its main function to the same degree. PTEN inhibits focal adhesions, which in turn controls cell proliferation [12], so mutations of PTEN can change focal adhesions and thus aid tumor progression [13]. Most often, PTEN mutations are missense or truncating (Figure 2). Subsequently, PTEN mutations can lead to shorter survival rates in GBM patients. A meta-analysis found that GBM patients with PTEN mutations generally had a poorer prognosis, though the results are tentative and could have been biased to some extent [14]. Another study, though focused on gliomas in general as opposed to specifically GBM, yielded similar results, adding that PTEN mutations occurred later in glioma progression [15].

In GBM samples, a large amount of truncating and missense PTEN mutations occur in the C_2 domain (Figure 2). Since the C_2 domain is used to inhibit cell migration, the domain may have a role in tumor suppression [16]. Mutations occurring in the C2 domain can likely affect the ability of PTEN to perform its functions and allow faster GBM progression.

EGFR Association with GBM

EGFR is involved in the production of the protein epidermal growth factor receptor, which binds to ligands in order dimerize with other epidermal growth factor receptors, leading to the triggering of cell proliferation. Brennan et al. (2013) found that about 57% of GBMs in one study [17], while Xu et al. (2017) found EGFR overexpression in 60% of primary glioblastomas and 10% of secondary glioblastomas [18]. Although EGFR mutations are highly correlated with the development of GBM, attempts at targeting the gene have not yet been successful [19]. An et al. (2018) yielded similar results, as it found that while EGFR amplification and EGFRvIII appear commonly in GBM, immunotherapies using EGFR inhibitors have been ineffective [20]. Of the GBM data from CBioPortal [5], 18 out of 67 mutations from 543 samples occurred on alanine 289 (Figure 3), and patients generally have a worse prognosis if these mutations occur [21], most likely due to the mutations' negative impacts on protein EGFR's functions.

Conclusion

The purpose of this bioinformatics analysis is to identify several genes that are most frequently associated with the development and occurrence of glioblastoma (GBM). After the analysis based on GBM data from TCGA was conducted, genes TP53, PTEN, and EGFR are found to be highly associated with the oncogenesis of GBM, and may tentatively be applied to other types of cancers, especially gliomas. Mutations in these genes, specifically missense mutations, may result in altered proteins with inhibited functions. PTEN and EGFR are also associated with focal adhesions; a loss of ability to control focal adhesions likely leads to uncontrolled cell proliferation and tumor progression. Successfully targeting these genes for therapeutic treatments may be able to prolong survival in GBM patients and improve the prognosis for this cancer; however, such current attempts have proven to be unsuccessful. The methodology used in this bioinformatics analysis may be applied to other diseases with available data pertaining to gene mutations found in patients, which may assist in identifying key genes related to such diseases.

References:

1. Davis M. E. Glioblastoma: Overview of Disease and Treatment. Clinical journal of oncology nursing, 20(5 Suppl), 2016. – P. 2–8. URL: https://doi.org/10.1188/16.CJON.S1.2-8
2. Thakkar J. P., Dolecek T. A., Horbinski C., et al. Epidemiologic and molecular prognostic review of glioblastoma. Cancer Epidemiol Biomarkers Prev. 2014; 23(10): 1985–1996. Doi:10.1158/1055-9965.EPI-14-0275
3. Goldman M. J., Craft B., Hastie M. et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020.
4. Frankish A., et. al. GENCODE reference annotation for the human and mouse genomes". 2018.
5. Cerami et al., Cancer Discov. 2012 and Gao et al., Sci. Signal. 2013.
6. Huang D. W., Sherman B. T., Lempicki R. A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009; 4(1): 44-57.
7. Huang D. W., Sherman B. T., Lempicki R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37(1): 1–13.
8. Zhou X., Iversen E. S. Jr. & Parmigiani G. Classification of Missense Mutations of Disease Genes. Journal of the American StatisticalAssociation, 2005; 100(469), 51–60. URL: https://doi.org/10.1198/016214504000001817
9. Wang X., Chen J. X., Liu J. P., You C., Liu Y. H., Mao Q. Gain of function of mutant TP53 in glioblastoma: prognosis and response to temozolomide. Ann Surg Oncol. 2014; 21(4): 1337–1344. Doi:10.1245/s10434-013-3380-0
10. Végran F., Rebucci M., Chevrier S., Cadouot M., Boidot R. & Lizard-Nacol S. Only missense mutations affecting the DNA binding domain of p53 influence outcomes in patients with breast carcinoma. PloS one, 8(1), e55103. 2013. URL: https://doi.org/10.1371/journal.pone.0055103
11. Timofeev O., Schlereth K., Wanzel M., et al. p 53 DNA binding cooperativity is essential for apoptosis and tumor suppression in vivo. Cell Rep. 2013; 3(5): 1512–1525. Doi:10.1016/j.celrep.2013.04.008
12. Tamura M., Gu J., Matsumoto K., Aota S., Parsons R., Yamada K. M. Inhibition of cell migration, spreading, and focal adhesions by tumor suppressorPTEN. Science. 1998; 280(5369): 1614–1617. Doi:10.1126/science.280.5369.1614
13. Tamura M., Gu J., Takino T., Yamada K. M. Tumor suppressor PTEN inhibition of cell invasion, migration, and growth: differential involvement of focal adhesion kinase and p130Cas. Cancer Res. 1999; 59(2): 442–449.
14. Han F., Hu R., Yang H., Liu J., Sui J., Xiang X., Wang F., Chu L. & Song S. PTEN gene mutations correlate to poor prognosis in glioma patients: a meta-analysis. OncoTargets and therapy, 2016; 9, 3485–3492. URL: https://doi.org/10.2147/OTT.S99942
15. Yang Y., Shao N., Luo G., et al. Mutations of PTEN gene in gliomas correlate to tumor differentiation and short-term survival rate. Anticancer Res. 2010; 30(3): 981–985.
16. Raftopoulou M., Etienne-Manneville S., Self A., Nicholls S., Hall A. Regulation of cell migration by the C2 domain of the tumor suppressor PTEN. Science. 2004; 303(5661): 1179–1181. Doi:10.1126/science.1092089
17. Brennan C. W., Verhaak R. G., Mc Kenna A., Campos B., Noushmehr H., Salama S. R., Zheng S., Chakravarty D., Sanborn J. Z., Berman S. H., Beroukhim R., Bernard B., Wu C. J., Genovese G., Shmulevich I., Barnholtz-

Sloan J., Zou L., Vegesna R., Shukla S.A., Ciriello G., … TCGA Research Network. The somatic genomic landscape of glioblastoma. Cell, 2013; 155(2), 462–477. URL: https://doi.org/10.1016/j.cell.2013.09.034
18. Xu H., Zong H., Ma C., Ming X., Shang M., Li K., He X., Du H. & Cao L. Epidermal growth factor receptor in glioblastoma. Oncologyletters, 2017; 14(1), 512–516. URL: https://doi.org/10.3892/ol.2017.6221
19. Hatanpaa K.J., Burma S., Zhao D. & Habib A.A. Epidermal growth factor receptor in glioma: signal transduction, neuropathology, imaging, and radioresistance. Neoplasia (New York, N.Y.), 2010; 12(9), 675–684. URL: https://doi.org/10.1593/neo.10688
20. An Z., Aksoy O., Zheng T., Fan Q.W. & Weiss W.A. Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. Oncogene, 2018; 37(12), 1561–1575. URL: https://doi.org/10.1038/s41388-017-0045-7
21. Binder Z.A., Thorne A.H., Bakas S., et al. Epidermal Growth Factor Receptor Extracellular Domain Mutations in Glioblastoma Present Opportunities for Clinical Imaging and Therapeutic Development. CancerCell. 2018; 34(1): 163–177.e7. Doi:10.1016/j.ccell.2018.06.006

Section 6. Physiology

https://doi.org/10.29013/ELBLS-20-4-79-84

Pozilov Mamurjon Komiljonovich,
Senior research, Institute of Biophysics and Biochemistry at the National University of Uzbekistan named after Mirzo Ulugbek
E-mail: mamurjon2281@mail.ru

Ernazarov Zafar Mamurovich,
Junior research, Institute of Biophysics and Biochemistry at the National University of Uzbekistan named after Mirzo Ulugbek
E-mail: Zafarbek1985@gmail.com

Raximov Akmal Dilshod o'gli,
Institute of Biophysics and Biochemistry at the National University of Uzbekistan named after Mirzo Ulugbek

Kukanova Nargiza Fahritdinovna,
Institute of Biophysics and Biochemistry at the National University of Uzbekistan named after Mirzo Ulugbek

Asrarov Muzaffar Islamovich,
Professor, Institute of Biophysics and Biochemistry at the National University of Uzbekistan named after Mirzo Ulugbek

Makhmudov Rustamjon Rasuljonovich,
A. S. Sadykov Institute of Bioorganic Chemistry, Academy of Sciences of Uzbekistan

INFLUENCE OF NATURAL POLYPHENOLIC COMPOUND GOSSITAN ON ION CHANNELS OF MITOCHONDRIA OF THE HEART AND PANCREAS IN STREPTOZOTOCIN-INDUCED DIABETES

Abstract. The influence of natural polyphenolic compound of gossitan on mitochondrial function was investigated. It was shown that in streptozotocin (STZ)-induced diabetes damaged functional systems of rat hearth and pancreas mitochondria: mitochondrial permeability transition pore (mPTP) and ATP-dependent potassium channel (mitoK_{ATP}-channel). Pharmacotherapy with gossitan (intraperitoneally in dose of 10 mg/kg body weight) for 8 days has a protective effect on mitochondria in experimental diabetes, correction membrane disorders.

Keywords: Hearth, pancreas, mitochondria, mPTP, Streptozotocin-induced diabetes, gossitan.

The biological activity of polyphenol compounds isolated from plants is very high, and their mechanisms of action vary with each other [1, 1–3]. Quercetin, campherol, and epicatechins inhibit the formation of H_2O_2 in the mitochondria and the activity of the mitochondrial respiratory-chain complex 1 [2, 1562–1572]. Polyphenol compounds restore morphological changes in the mitochondrial matrix in experimental diabetic conditions. It also reduces the formation of free radicals, increases ATF synthesis and insulin resistance [3, 3135–3136].

Polyphenols are also effective in mitochondrial bioenergetics dysfunction, inhibition of uncoupling protein (UCP) activity, mPTP, and damage to proteins, and lipids that causes changes in various diseases [4, 67–78]. Although the antidiabetic activities of these compounds have been extensively studied, their effects on mitochondrial functional changes remain unexplored.

We know that mitochondria are significantly damaged in heart muscle injury in diabetes and ischemia. In diabetes, disruption of the homeostasis of calcium ions in the mitochondria occurs with the activation (or opening) of nonspecific pores in the inner membrane of the mitochondria, with the loss of various substances and ions from the mitochondrial matrix, as well as changes in the outer membrane [5, 1009–1010].

These processes can result in decreased cellular ATF synthesis, significant changes in mitochondrial membranes, and consequent cell death. The formation of nonspecific pores associated with calcium ions in the mitochondria and the formation of reactive oxygen species (ROS) directly accelerates the process of cell death [6, 874–875] At present, although the role of cardiac mitochondria in the role of mitochondria in the pathogenesis of diabetes has been Despite extensive research, their correction with polyphenol compounds has not been adequately studied.

The aim of the study was to study the corrective effect of polyphenol gossitan [7, 109–110] isolated from the *Gossypium hirsutum L.* plant on the dysfunction of mPTP and mitoK_{ATP}-channel of rat heart and pancreas mitochondria under conditions of STZ-diabetes.

Material and Methods. For screening and detailed study of the mechanism of action of pharmacological agents are widely used various experimental models of diabetes caused by administration of alloxan and STZ al., cytotoxic activity on β-cells of the pancreas. We have in this study used an experimental model of diabetes induced by STZ.

Experiments were performed white mongrel male rats weighing 180–200 g. The animals were divided into three groups: I group – control, II group – the animals with experimental diabetes, which once were injected intraperitoneally with an STZ (50 mg/kg body weight intraperitoneally in a 0,1 mol/L citrate buffer, pH 4,5) and III group – STZ-induced diabetes+gossitan (intraperitoneally dose of 10 mg/kg body weight) for 8 days starting from 12 days after administration of STZ and reaching a predetermined level of hyperglycemia. Blood glucose was determined using glucose oxidase method set «Glucose – enzymatic-colorimetric test» (Cypress diagnostic, Belgium).

Mitochondria isolated from rat hearth and pancreas by differential centrifugation according to [8, 30–32; 9, 71–72]. Nuclei and cellular fragments were removed by centrifugation at 600 g for 7 minutes in a centrifuge. The mitochondria are pelleted at 7000 g for 15 minutes at the same temperature. The mitochondrial pellet was washed twice in the isolation EDTA-free medium.

MPTP condition assessed by the speed of Ca^{2+}-dependent swelling of mitochondria, the mitochondrial suspension recording light scattering at 540 nm. Experiments at 26 °C in a swelling medium of 200 mM sucrose, 20 μM EGTA, 5 mM succinate, 2 μM rotenone, 1 μg/ml oligomycin, 20 mM Tris, 20 mM HEPES, and 1 mM KH_2PO_4, pH 7,2 [10, 16755–16760]. The concentration of mitochondria in the swelling experiments was 0,5 mg protein/ml.

The content of mitochondrial protein was determined by the Lowry method in the modification of the Peterson [11, 346–347].

Mitochondrial swelling induced activation mito-K_{ATP}–channel was recorded using a change in light scattering at a wavelength of 540 nm. Mitochondria were added to the standard incubation medium of the following composition: 125 mM KCl, 10 mM Hepes, 5 mM succinate, 1 mM $MgCl_2$, 2,5 m M K_2HPO_4, 2,5 mM KH_2PO_4, rotenone 1 μM/ ml, oligomycin 1 μ/ml, pH 7,4.

The results were statistically processed using the Origin 6.1 program. The P value < 0.05 was considered as an indicator of significant differences.

Results and discussion. The investigation on the effect of hypoglycemic polyphenol gossitan on mitochondrial distend isolated from animal hearts, called the STZ-diabetes model, was studied. A concentration of 20 μM Ca^{2+} ions was applied as an inducer for the induction of cardiac mitochondrial swelling. The results obtained, STZ-diabetes (group II), the distant of mitochondria increased by 86.1 ± 7.1% compared with control of (group I) was determined (fig. 1).

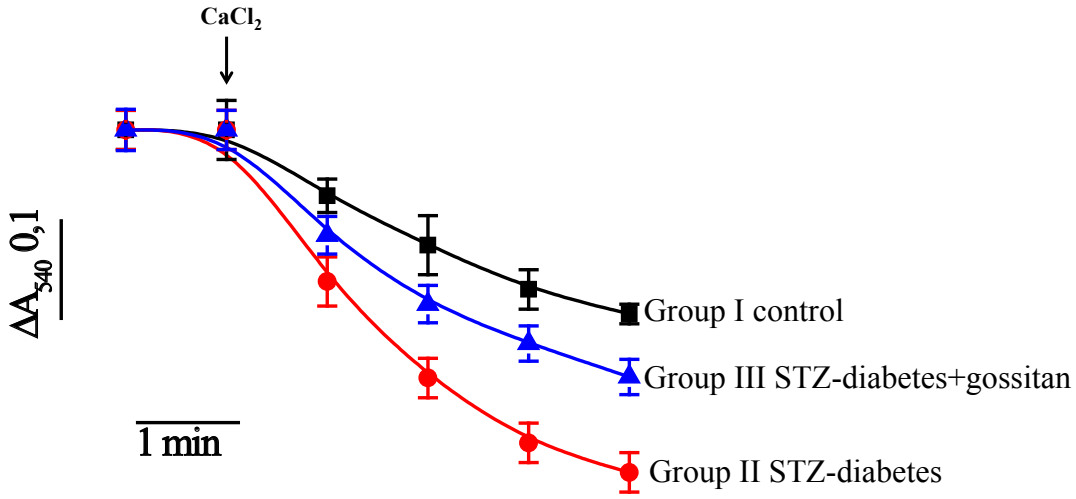


Figure 1. The effect of gositan on mitochondrial swelling in the heart of a rat with STZ-diabetes

As the result of diabetes the is increased Ca^{2+} retention capacity on cardiac mitochondria and a loss of membrane stability was observed, which provides a state of high permeability of mPTP. As a result, oral supplement of gossitan polyphenols (10.0 mg/kg) once daily for 8 days for STZ-diabetic rats was revealed and remarkable decrease can be seen on the swelling of mitochondria isolated from the heart. As a consequence, distend of mitochondria under the influence of gossitan was inhibited by 51.9±4.8% compared with STZ-diabetes (group II) was found (Fig. 1). The main causes for the opening of mPTP in STZ-diabetes include the development of oxidative stress, prooxidants, LPO induction, and oxidation of thiol groups in the mPTP complex. Due to the strong antiradical property of polyphenolic compounds, it can reduce the amount of free radicals in mitochondria and possess capability of controlling the inhibitory properties of cyclosporin A by binding to the CyP-D matrix domain.

Through our next experiment the effect of the polyphenol gossitan on the pancreas mitochondria in the conditions of STZ-diabetic rats was also investigated. STZ is estimated as a specific target of pancreas, its plasma membrane can increased LPO both disrupt the activity of ion channels. On the other hand, it intensifies the generation of free radicals and causes impaired insulin secretion as a result of reduced ATF synthesis in the mitochondria.

The pharmacotherapy on animals with STZ-diabetes were per os with polyphenolic compounds. Once glucose levels approached towards normal position,

they were decapitated and mitochondria were isolated from pancreas. The results obtained, the distend of the pancreas mitochondria increased by 95.5±5.4% compared with the control in STZ-diabetes (Fig. 2).

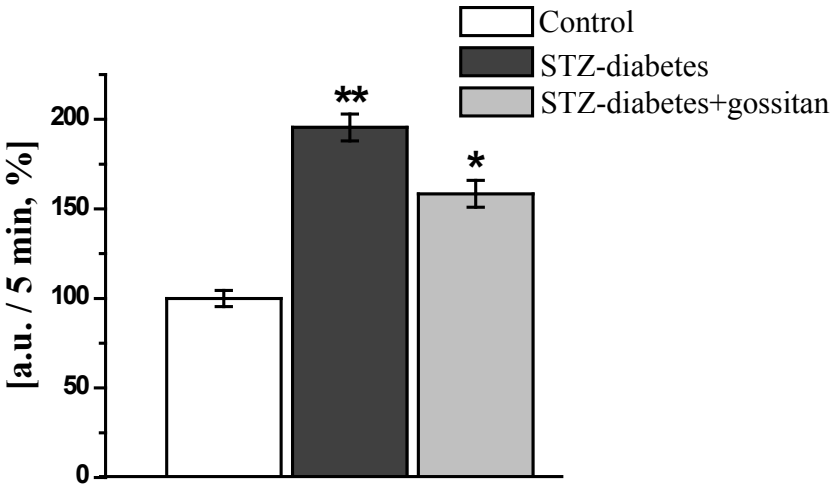


Figure 2. The effect of gossitan on pancreas mitochondria swelling with STZ-induced diabetic rats (*p<0.05; **p<0.01; n=5)

After conducting pharmacotherapy on animals of group III with STZ-diabetes with gossitan polyphenols was apparent that their mitochondrial swelling pancreas was inhibited by 37.1±2.7% contrasted with group II. Polyphenolic compounds within possession of hypoglycemic properties which in experiments, can recreated mitochondrial dysfunction of pancreas in diabetes conditions.

Effect of gossitan on activity mitoK$_{ATP}$-channels in mitochondria of rat heart in experimental diabetes. According to modern researches, owing to priority of the functional significance of the mitoK$_{ATP}$-channel in ischemia and hypoxia it has been frequent studied in experiments. MitoK$_{ATP}$-channels is considered as a selective channel in mitochondria and plasma membrane, usually both channels exist common activators and inhibitors. Although such a pharmacological agents currently have been identified in a large scope, the selectivity of the modulator effectiveness actually depends on the type of cells which is investigated and the experimental conditions.

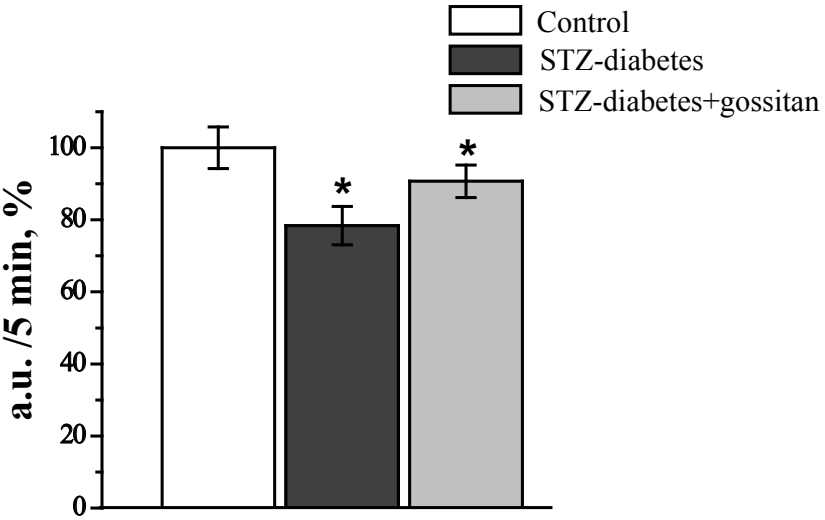


Figure 3. Effect of gossitan on rat heart mitoK$_{ATP}$-channel with STZ-induced diabetes (*p<0.05; n=6)

It can be hardly found data on modulators that affect the activity of the mitoK_{ATP}-channel of the liver and heart in experimental diabetes. It is stated only by a number of scientists that glibenclamide owe a productive effect on the activity of the mitoK_{ATP}-channel in diabetes [12, 485–488; 13, 101–105]. At present, according to the latest literature data, the mechanisms of inhibition of the K_{ATP}-channel of the plasma membrane of pancreatic cells in diabetes mellitus are widely studied [14, 161–162]. However, changes in the flow of K^+ ions in the mitochondrial channel of the heart in diabetes and the polyphenolic compounds affecting them have not been sufficiently studied. In this regard, our experiments possess significant value in studying the effect of gossitan polyphenol in *in vivo* experiments in condition diabetes as well as pathological changes in the mitoK_{ATP}-channel.

It was obvious through the results defined that the mitoK_{ATP}-channel isolated from the hearts of animals with STZ-induced group II diabetes was inhibited by 21.6±1.8% in comparison group I (Fig 3.). Inhibition of the mitoK_{ATP}-channel channel of the heart, in contrast to the mitoK_{ATP}-channel channel of the liver, causes a deficiency energy supply in both normal and pathological conditions.

A decrease in the size of the cardiac mitochondrial matrix is estimated as the result of inhibition of K_{ATP}-channels leads slowing down of respiration, and a sharp decrease in ATF synthesis could be noticed. When STZ-diabetic group III animals were treated with gossitan, the activation of the mitoK_{ATP}-channel separated from their heart was detected during the investigation. Consequently, the polyphenols of gossitan were productively effected in the mitochondrial channel of the heart in STZ-diabetes. Gossitan activates the reduction in the permeability of the channels as a result of diabetes. As a result, it can inhibit oxidative stress can emerge apoptosis in cardiomyocytes. In experiments modulators are exist within capability of various effects on plasma and mitoK_{ATP}-channels. They act as an activator of the K_{ATP}-channel of the plasma membrane of smooth muscles, as for mitoK_{ATP}-channel it functions as an inhibitor as well [15, 1183–1185]. It requires deep further investigation on the effect of polyphenolic compounds on the mitochondrial channel of the heart as an inhibitor and activator of condition of diabetes.

Conclusions: Gossitan polyphenol could afford to inhibite mPTP in rat liver and pancreas mitochondria in STZ-induced diabetic conditions.

By applying the polyphenols of gossitan to STZ-induced diabetic rats, their mitochondria which is seperated from their heart, the activeness of the mitoK_{ATP}-channel was determined.

References:

1. Serrano J. C. E., Cassanye A., Martín-Gari M., Granado-Serrano A. B., Portero-Otín M. Effect of dietary bioactive compounds on mitochondrial and metabolic flexibility // Diseases. 2016. – V. 4(2). – P. 1–14.
2. Lagoa R., Graziani I., Lopez-Sanchez C., Garcia-Martinez V., Gutierrez-Merino C. Complex I and cytochrome c are molecular targets of flavonoids that inhibit hydrogen peroxide production by mitochondria // Biochim. Biophys. Acta. 2011. – V. 1807(12). – P. 1562–1572.
3. Heinonen S., Buskova J., Muniandy M., Kaksonen R., Ollikainen M., Ismail K., Hakkarainen A., Lundbom J., Lundbom N., Vuolteenaho K., Moilanen E., Kaprio J., Rissanen A., Suomalainen A., Pietiläinen K. H. Impaired mitochondrial biogenesis in adipose tissue in acquired obesity // Diabetes. 2015. – V. 64(9). – P. 3135–3145.
4. Dorta D. J., Pigoso A. A., Mingatto F. E., Rodrigues T., Prado I. M., Helena A. F., Uyemura S. A., Santos A. C., Curti C. The interaction of flavonoids with mitochondria: Effects on energetic processes // Chem. Biol. Interact. 2005. – V. 152(2–3). – P. 67–78.

5. Sloan R. C., Moukdar F., Frasier C. R., Patel H. D., Bostian P. A., Lust R. M., Brown D. A. Mitochondrial permeability transition in the diabetic heart: Contributions of thiol redox state and mitochondrial calcium to augmented reperfusion injury // J Mol Cell Cardiol.2012.– V. 52(5).– P. 1009–1018.
6. Costa A. D., Garlid K. D. Intramitochondrial signaling: interactions among mito K_{ATP}, PKCε, ROS and MPT // Am J Physiol Heart Circ Physiol.2008.– V. 295(2).– P. 874–882.
7. Salikhov S. I., Mavlyanov S., Abdulladjanova N. G., Pirniyazov A. J., Dalimov D. N., Salakhutdinov B. A., Kurmukov A. G. Polyphenols of some tannin containing plants and creation on their base drug remedies // New research on Biotech. Med.2006.– P. 109–117.
8. Алматов К. Т., Ахмеров Р. Н., Аулов Д. М., Рахимов М. М. Выделение митохондрий из поджелудочной железы // Узб. биол. журн.1977.– № 3.– С. 30–32.
9. Ахмеров Р. Н. Размельчитель ткани (комбинированный гомогенизатор) с резьбовым ножевым блоком и тканеподающим устройством // Узб. биол. журн.1979.– № 5.– С. 71–72.
10. He L., Lemasters J. J. Heat shock suppresses the permeability transition in rat liver mitochondria // J. Biol. Chem.2003.– V. 278(19).– P. 16755–16760.
11. Peterson G. L. A simplification of the protein assay method of Lowry et al. which is more generally applicable // Analytical biochemistry.1977.– 83(2).– P. 346–356.
12. Proks P., Lippiat J. D. Membrane ion channels and diabetes // Current Pharmaceutical Design.2006.– V. 12(4).– P. 485–501.
13. Szewczyk A., Wojtczak L. Mitochondria as a pharmacological target // Pharmacol Rev.2002.– 54(1).– P. 101–127.
14. Wang Q., Cai Y., Casteele M. V., Pipeleers D., Ling Z. Interaction of glibenclamide and metformin at the level of translation in pancreatic b cells // Journal of Endocrinology.2011.– V. 208(2).– P. 161–169.
15. Sasaki N., Murata M., Guo Y., Jo S. H., Ohler A., Akao M., O'Rourke B., Xiao R. P., Bolli R., Marbán E. MCC-134, a single pharmacophore, opens surface atp-sensitive potassium channels, blocks mitochondrial ATP–sensitive potassium channels, and suppresses preconditioning // Circulation.2003.– V. 107(8).– P. 1183–1188.

Contents