



НОВ БЪЛГАРСКИ УНИВЕРСИТЕТ

Департамент Информатика

Бакалавърка програма Информатика

**Автоматизиран биоинформатичен анализ
на генетични варианти, потенциално
свързани със стареенето**

Дипломна работа на
Михаил М. Здравков

Научни ръководители:
доц. д-р Милена Георгиева
Момчил Топалов

Ръководител катедра:
гл. ас. д-р Методи Трайков

София 2022

Съдържание

1	Увод	6
2	Литературен обзор	8
2.1	Значение на стареенето	8
2.1.1	Дефиниция	8
2.1.2	Физиологични ефекти	8
2.1.3	Демографски и икономически ефекти	9
2.2	Молекулярно-биологични теории	
	за стареенето	10
2.2.1	Натрупване на геномни изменения	10
2.2.2	Скъсяване на теломерите	10
2.2.3	Клетъчно стареене	11
2.2.4	Епигенетични изменения	11
3	Генетични фактори,	
	влияещи на процеса	
	на стареене	12
3.1	Съществуване	12
3.2	Начин на действие	12
3.3	Проблеми на изследването им	12
4	Обзор на съществуващи	
	биоинформатични решения	13
4.1	Анотация	14
4.1.1	snpEff	14
4.1.2	VEP	14
4.1.3	Annotar	14
4.2	Филтриране и анализ на генетични	
	варианти	14
4.2.1	snpSift	14
4.2.2	???	14

4.3	Геномни браузъри	14
4.3.1	UCSC Genome Browser	14
4.3.2	IGV Genome Browser	14
4.4	Нагъване на протеини	14
4.4.1	Подходи	14
4.4.2	AlphaFold	14
4.5	Интегрирани софтуерни решения	14
4.5.1	Galaxy Project	14
5	Цели	15
6	Задачи	16
7	Представяне на софтуерното решение	17
7.1	Софтуерна архитектура	17
7.2	Структура на базата данни	17
7.3	Управление на генни множества	17
7.4	Обработка на VCF при импортиране	17
7.5	Операции върху импортиран VCF	17
8	Дискусия	18
9	Изводи	19
	Използвани външни библиотеки и софтуер	20
	Библиография	22

1. Увод

Стареенето е естествен процес, който има огромно значение както за отделния индивид, така и за обществото като цяло. С напредването на възрастта, рискът от разнообразни заболявания като рак, болест на Алцхаймер, диабет, сърдечно-съдови заболявания и др. нараства значително. Смята се, че около две-трети от смъртните случаи при хора се дължат на заболявания, свързани с възрастта. Същевременно, с глобалното нарастване на средната продължителност на живота, проблемите на стареенето засягат все повече хора и имат все по-голямо обществено значение. От социална гледна точка, стареенето оказва значителен икономически и демографски ефект.

Установено е, че процесът на стареене се влияе както от генетични, така и от епигенетични фактори. Въпреки това, този процес все още не е достатъчно добре разбран от науката, поради което е трудно да се създадат ефективни методи за терапия и справяне с негативните му ефекти.

Настоящата дипломна работа се фокусира върху генетичната основа на стареенето. Основен подход при нейното изследване е анализът на генетични варианти. При такива изследвания е необходима обработката на големи обеми от данни, което налага нуждата от използване на специализиран биоинформатичен софтуер. Налични са множество различни инструменти, покриващи различни аспекти от обработката на файлове с генетични варианти - анотация, филтриране, анализ и тн. Повечето от тях, обаче, изискват значителни технически познания, което ги прави трудни за използване от специалисти в други области, като биология и генетика.

Целта на настоящата дипломна работа е създаването на интегрирана софтуерна система за биоинформатични изследвания на генетични варианти и предсказване на тяхната потенциална асоциация с процеса на стареене. Надяваме се, чрез създаване на по-достъпен инструмент, да допринесем за

бъдещи изследвания на процеса на стареене и за търсенето на ефективни терапии против негативните му ефекти.

2. Литературен обзор

2.1 Значение на стареенето

2.1.1 Дефиниция

Въпреки, че концепцията за стареене е универсално разбираема, формалната ѝ дефиниция не е тривиална и множество автори дават твърде различни определения за този термин. Аркинг (2006, стр. 11) прави преглед на наличната литература и, в резултат, предлага следната дефиниция[1]:

„Стареенето е независима от времето поредица от кумулативни, прогресивни, свойствени и вредящи структурни и функционални промени, които обикновено започват да се изразяват при репродуктивната зрялост и приключват със смъртта.“

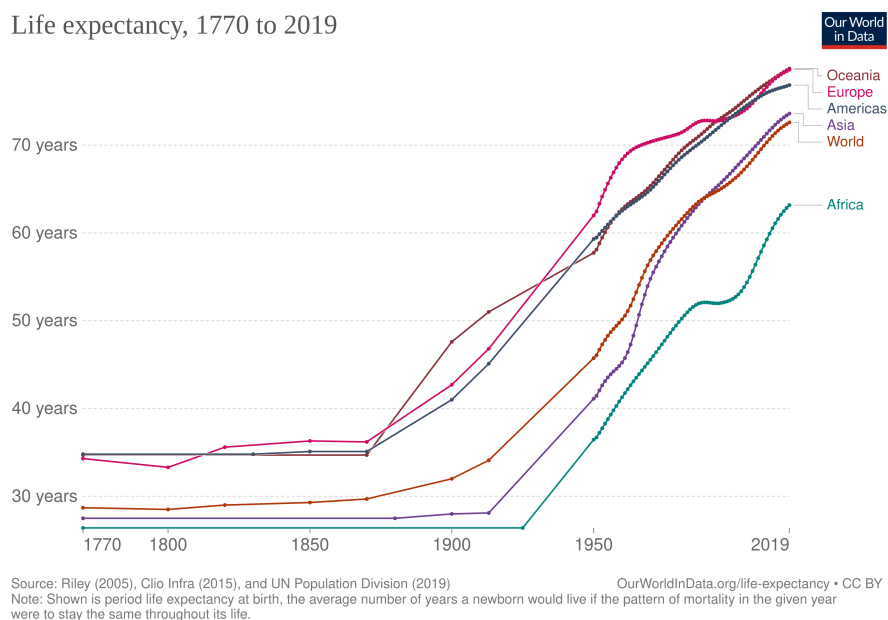
Макар времето да няма каузална връзка с ефектите на стареенето, то корелацията помежду им е причина обикновено да се говори за ефектите на стареенето като за нещо, настъпващо с напредването на възрастта.

2.1.2 Физиологични ефекти

Стареенето оказва изключително голям ефект върху човешкото тяло. То обикновено включва широк спектър от различни физиологични промени, които влошават жизнеността и качеството на живот на индивида. Примери за това са понижена фертилност при жените[4]; загуба на телесна маса[16]; влошен слух[3]; повишен риск от хронични заболявания[7][15]; хронична болка[12]; загуба на сила и еластичност в мускулно-скелетната система; понижената способност за устояване на инфекции, екстремни температури и др. видове стрес; влошаване на зрението; загуба на неврологични функции[18] и др.

2.1.3 Демографски и икономически ефекти

През последния един век очакваната продължителност на живота в целия свят драстично се е повишила[20] (виж фиг. 2.1). Освен безспорните ползи, това води и до редица проблеми. Удължаването на продължителността на живота, в комбинация с наблюдавания спад на раждаемостта, се очаква да доведе до застаряване на населението[8]. Световната Здравна Организация (СЗО) предупреждава, че се очаква между 2015 и 2050 броят на хората над 60-годишна възраст да се повиши от 12% от населението до 22%[13]. Същевременно, по данни на СЗО, увеличаването на продължителността на живота (с 6 години за периода между 2000 и 2019) изпреварва увеличаването в продължителността на здравословния живот (с 5.4 години за същия период)[14].



Фигура 2.1: Очаквана продължителност на живота за различни региони

Застаряването на населението би оказало неблагоприятен ефект и върху икономиката на държавите. Първо, заради увеличаването на дяла на хора, които не участват в работната сила. Второ, поради това, че здравните системи ще бъдат допълнително натоварени с по-голям брой хора в напреднала възраст, за които рисковете от хронични заболявания са значително по-големи.

2.2 Молекулярно-биологични теории за стареенето

Стареенето е въпрос, който вълнува учените от дълго време. През 1990-та, Медведев твърди, че вече съществуват над 300 теории за стареенето[10]. Въпреки постигнатият значителен прогрес през последните години в областта на геронтологията, причините за стареенето все още остават ненапълно обяснени. Това се дължи на факта, че стареенето е сложен процес, в който са намесени множество фактори. Все още липсва голяма обединяваща теория на стареенето, която да обясни изцяло процеса, но съществуват множество теории, които дават добра представа за различни негови аспекти[18]. Следва кратък преглед на основните теории:

2.2.1 Натрупване на геномни изменения

Изменения в ДНК молекулите могат да настъпят както в следствие на вътрешноклетъчни фактори, така и поради въздействието на външни мутагени. Примери за вътрешноклетъчни фактори са случайни грешки при репликация и оксидативния стрес, предизвикан от натрупването на свободни радикали[19]. Външните мутагени могат да бъдат разделени на три вида - физични, химични и биологични. Пример за физичен мутаген е радиацията[2], а за биологичен вирусните инфекции, които също могат да предизвикат генетични мутации. Измененията в ДНК молекулите включват различни видове мутации като точкови мутации, делеции и инсерции, транслокации, инверсии и др.

Съществуват механизми, чрез които клетките засичат мутациите и ги поправят. Основни такива механизми са гените ATM и TP53. Все пак, тези механизми не са ефективни на 100% и ефективността им допълнително спада с възрастта[9]. В резултат, в течение на времето, ДНК молекулите акумулират все повече мутации. Смята се, че тази геномна нестабилност е един от основните фактори, допринасящи за процеса на стареенето[17].

2.2.2 Скъсяване на теломерите

Теломерите са регион, намиращ се в края на хромозомите, в който се съдържат повтарящи се поредици от нуклеотидни бази. Те служат за запазване на хромозомата от рекомбинация и постепенна деградация и дават възможност на клетката да различава края на хромозомата от слу-

чайни прекъсвания, при които биха били активирани механизмите за поправка на ДНК[5]. При всеки цикъл на делене на клетката, теломерите се скъсяват поради непълното синтезиране на изоставащата нишка от ДНК полимеразата[6]. Този проблем се компенсира донякъде от ензима теломераза, който пренася своя собствена РНК молекула и я използва като шаблон, спрямо който да удължи скъсения теломер. Въпреки това, недостатъчната експресия на теломеразата води до постепенното скъсяване на теломерите. Това може да доведе до загуба на репликативна способност на клетката и блокирането на клетъчния ѝ цикъл, процес известен като клетъчно стареене[11].

2.2.3 Клетъчно стареене

TODO

2.2.4 Епигенетични изменения

TODO

3. Генетични фактори, влияещи на процеса на стареене

В секция 2.2 беше представен кратък обзор на различните биологични процеси, които способстват процеса на стареене. Уместен е въпросът дали има определени генетични фактори, които оказват въздействие на тези процеси. Ако това е така, бихме могли да очакваме, че съществуват генни алели, които забързват или забавят стареенето. В текущата глава ще разгледаме въпроса за съществуването на такива генни алели, както и за начините им на действие и методите за изследването им.

3.1 Съществуване

3.2 Начин на действие

3.3 Проблеми на изследването им

4. Обзор на съществуващи биоинформатични решения

4.1 Анотация

4.1.1 snpEff

4.1.2 VEP

4.1.3 Annovar

4.2 Филтриране и анализ на генетични варианти

4.2.1 snpSift

4.2.2 ???

4.3 Геномни браузъри

4.3.1 UCSC Genome Browser

4.3.2 IGV Genome Browser

4.4 Нагъване на протеини

4.4.1 Подходи

4.4.2 AlphaFold

4.5 Интегрирани софтуерни решения

4.5.1 Galaxy Project

5. Цели

6. Задачи

7. Представяне на софтуерното решение

- 7.1 Софтуерна архитектура
- 7.2 Структура на базата данни
- 7.3 Управление на генни множества
- 7.4 Обработка на VCF при импортиране
- 7.5 Операции върху импортиран VCF

8. Дискусия

9. Изводи

Използвани външни библиотеки и софтуер

Библиография

- [1] R. Arking. *Biology of Aging: Observations and Principles*. Oxford University Press, 2006.
- [2] L. H. Breimer. Ionizing radiation-induced mutagenesis. *Br J Cancer*, 57(1):6–18, Jan 1988.
- [3] K. Feder, D. Michaud, P. Ramage-Morin, J. McNamee, and Y. Beaugerard. Prevalence of hearing loss among Canadians aged 20 to 79: Audiometric results from the 2012/2013 Canadian Health Measures Survey. *Health Rep*, 26(7):18–25, Jul 2015.
- [4] K. George and M. S. Kamath. Fertility and age. *J Hum Reprod Sci*, 3(3):121–123, Sep 2010.
- [5] Jack D Griffith, Laurey Comeau, Soraya Rosenfield, Rachel M Stansel, Alessandro Bianchi, Heidi Moss, and Titia de Lange. Mammalian telomeres end in a large duplex loop. *Cell*, 97(4):503–514, 1999.
- [6] Alexander K. Koliada, Dmitry S. Krasnenkov, and Alexander M. Vaiserman. Telomeric aging: mitotic clock or stress indicator? *Frontiers in Genetics*, 6, 2015.
- [7] E. B. Larson, K. Yaffe, and K. M. Langa. New insights into the dementia epidemic. *N Engl J Med*, 369(24):2275–2277, Dec 2013.
- [8] W. Lutz, W. Sanderson, and S. Scherbov. The coming acceleration of global population ageing. *Nature*, 451(7179):716–719, Feb 2008.
- [9] Mark T. Mc Auley, Alvaro Martinez Guimera, David Hodgson, Neil McDonald, Kathleen M. Mooney, Amy E. Morgan, and Carole J. Proctor. Modelling the molecular mechanisms of aging. *Bioscience Reports*, 37(1), 02 2017. BSR20160177.

- [10] Zhores A Medvedev. An attempt at a rational classification of theories of ageing. *Biological Reviews*, 65(3):375–398, 1990.
- [11] Keiko Muraki, Kristine Nyhan, Limei Han, and John P Murnane. Mechanisms of telomere loss and their consequences for chromosome instability. *Frontiers in oncology*, 2:135, 2012.
- [12] AGS Panel on Persistent Pain in Older Persons. The management of persistent pain in older persons. *Journal of the American Geriatrics Society*, 50(6 Suppl):S205–224, Jun 2002.
- [13] World Health Organization. *World report on ageing and health*. World Health Organization, 2015.
- [14] World Health Organization. Life expectancy and healthy life expectancy - data by country. *Published online*, 2020.
- [15] Sahdeo Prasad, Bokyung Sung, and Bharat B. Aggarwal. Age-associated chronic diseases require age-old medicine: Role of chronic inflammation. *Preventive Medicine*, 54:S29–S37, 2012. Dietary Nutraceuticals and Age Management Medicine.
- [16] R. P. Spencer. Organ/body weight loss with aging: evidence for co-ordinated involution. *Med Hypotheses*, 46(2):59–62, Feb 1996.
- [17] J. Vijg and Y. Suh. Genome instability and aging. *Annu Rev Physiol*, 75:645–668, 2013.
- [18] Jose Viña, Consuelo Borrás, and Jaime Miquel. Theories of ageing. *IUBMB life*, 59(4-5):249–254, 2007.
- [19] David Wang, Deborah A. Kreutzer, and John M. Essigmann. Mutagenicity and repair of oxidative dna damage: insights from studies using defined lesions. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 400(1):99–115, 1998.
- [20] Richard Zijdeman and Filipa Ribeira da Silva. Life Expectancy at Birth (Total). 2015.