

Влияние мутационного груза на здоровье человека: анализ пороговых эффектов в больших популяционных и когортных выборках

Форма 1

Аннотация проекта (не более 0,5 стр., в том числе кратко – актуальность, междисциплинарность, уровень значимости и научная новизна исследования; ожидаемые результаты и их значимость; аннотация будет опубликована на сайте Фонда, если проект получит поддержку)

Геном каждого организма не оптимален, поскольку несет в себе сотни или тысячи слабовредных генетических вариантов (новые мутации и мутации, сегрегирующие в популяции с низкими частотами). Например, геном даже абсолютно здорового человека может содержать несколько мутаций, приводящих к полной потере функции генов, несколько десятков делеций и дупликаций, сотни субоптимальных несинонимичных замещений в белок-кодирующих генах и тысячи редких полиморфизмов в регуляторных районах. Каждый такой слабовредный генетический вариант сам по себе очень мало влияет на здоровье организма, однако все вместе эти варианты (далее «мутационный груз») могут увеличивать вероятность проявления многих полигенных заболеваний, ускорять процесс старения и, в целом, понижать приспособленность. Несмотря на фундаментальную значимость мутационного груза для медицинской, популяционной и эволюционной генетики, мы до сих пор знаем крайне мало как о вариантах, составляющих этот груз, так и о суммарном влиянии этих вариантов на приспособленность. Медицинские генетики хорошо понимают механизмы проявления многих моногенных заболеваний, тогда как анализ рисков сложных полигенных заболеваний, определяемых грузом слабовредных вариантов, является активной «зоной роста» современной геномики.

Основной вопрос, который мы ставим перед собой в данном проекте следующий: **как влияет мутационный груз на здоровье человека?** Для ответа на данный вопрос нам необходимо не только понять из каких классов вредных мутаций состоит мутационный груз, но и как взаимодействуют друг с другом вредные варианты. Для этого мы будем рассматривать следующие более специфические вопросы:

- (i) как взаимодействуют друг с другом вредные варианты (на примере делеций и дупликаций) в одном и том же геноме?
- (ii) возможна ли частичная компенсация делеции или дупликации регуляторными вариантами?
- (iii) может ли сильно вредная мутация частично компенсироваться за счет пониженного числа слабовредных мутаций (модель генетического гандикапа)?
- (iv) как наследуется мутационный груз от родителей к детям (наследование полигенных рисков)?
- (v) какие фенотипические признаки человека лучше всего коррелируют с геномным мутационным грузом?

Для решения поставленных вопросов мы планируем использовать **генетические и фенотипические данные из двух больших популяционных и одной когортной выборки (английский и эстонский популяционные биобанки, состоящие в данный момент из 500 000 и 50 000 индивидуумов; 650 семей с ребенком с трисомией хромосомы 21) и анализировать распределение нескольких классов вредных мутаций в каждом геноме (трисомия, редкие делеции и дупликации, субоптимальные не синонимические и регуляторные варианты, генетические предрасположенности к заболеваниям, подсчитанные по данным полного генома).** С фундаментальной точки зрения, реализация данного проекта поможет нам охарактеризовать лучше мутационный груз человека и оценить силу естественного отбора, действующего в нашей популяции. С прикладной точки зрения, результаты нашего проекта помогут более точно оценивать риски сложных заболеваний и выделять когорты людей с повышенным риском.

Аннотация проекта на английском языке (объемом не более 0,5 стр.; в том числе кратко – актуальность, междисциплинарность, уровень фундаментальности и научная новизна; ожидаемые результаты и их значимость)

Every person's genome is far from being ideal as it usually carries hundreds, maybe even thousands of slightly-deleterious alleles. A typical person - even if perfectly healthy - will have in their genome: several mutations, that can cause a complete loss of function in a gene; a number of deletions and duplications; hundreds of non-synonymous substitutions in protein-coding genes; and thousands of regulatory mutations, that affect gene expression levels. Every one of these alleles does not produce a significant effect on the individual's health by itself, but when combined, these mutations cumulatively (from here on out "mutational burden" or "mutational load") can increase the probability of occurrence of many polygenic diseases, speed up ageing, and lower the adaptability of the individual as a whole.

Despite the fundamental importance of mutational burden for the purposes of medical, population and evolutionary genetics, we still know very little about its constituents and their cumulative effect on fitness. The field of medical genetics is well-versed in the mechanisms of heritability of monogenic diseases, whereas the analysis of inheritance risks for complex polygenic diseases is an actively expanding field of modern genomics with room for growth.

Our main question for this project is the following: how do harmful mutations from same genome interact with each other? In more detail, we consider the following issues:

- (i) how can partial compensation for one highly-deleterious mutation occur at the expense of a reduced number of background slightly-deleterious mutations (genetic handicap model)?
- (ii) how is the mutation load passed from parents to children (inheritance of polygenic risks)?
- (iii) how do deletions interact with each other in the same genome?
- (iv) is it possible to partially compensate for a deletion with certain regulatory options?
- (v) which phenotypic traits are best correlated with the mutation load?

To address the issues raised, we will focus on the analysis of three cohorts: the UK population biobank (500,000 people); Estonian population biobank (50 000) and 650 families with a child with chromosomal trisomy 21. From the fundamental point of view, the implementation of this project will help us to better characterise the mutational load of a person and to gauge the force of natural selection acting in our population. From an applied point of view, the results of our project will help to more accurately assess the risks of complex diseases in the human population and to point out the cohorts of people at increased risk for a particular disease.

Форма 4. Содержание проекта

Описание фундаментальной научной задачи в рамках темы, на решение которой направлено исследование

Основная фундаментальная научная задача нашего проекта - изучение влияния мутационного груза человека на его здоровье. Для этого мы фокусируемся на поиске следов отрицательного отбора в человеческой популяции и изучаем наиболее частые типы взаимодействий между разными классами слабавредных вариантов в геноме человека. Данные знания помогут реконструировать ожидаемый фенотип из геномных данных, что важно как с фундаментальной (понимание механизмов проявления слабавредных вариантов), так и прикладной (предсказание генетических рисков и степени проявления заболеваний) точек зрения.

Анализ современного состояния исследований в данной области (приводится обзор исследований в данной области со ссылками на публикации в научной литературе)

Фенотипы людей с одной и той же вредной мутацией отличаются огромным разнообразием. Для каждой генетической болезни её экспрессивность и пенетрантность могут объясняться как средовыми факторами, так и генетическими. В данном проекте мы фокусируемся на поиске генетических факторов, модифицирующих степень проявления вредных мутаций. Например, все носители гетерозиготной делеции гена *tbx6* диагностируются со сколиозом, однако степень проявления сколиоза меняется от крайне слабой до сильно выраженной формы (Wu et al. 2015). Было показано, что степень проявления сколиоза зависит от точечковой регуляторной мутации, которая увеличивает уровень экспрессии гена *tbx6*, частично ослабляя эффект гетерозиготной делеции этого гена (Wu et al. 2015). Количество хорошо изученных генетических модификаторов вредных мутаций, как в примере со сколиозом, не так велико и, в большинстве случаев, причины фенотипической изменчивости среди носителей одной и той же мутации остаются неизвестными.

Мы полагаем, что мутационный груз (совокупность всех слабовредных вариантов в геноме) может служить важным модификатором вредных мутаций, ослабляя или усиливая их эффект, и, таким образом, влияя на конечный фенотип каждого человека. Недавние масштабные популяционные исследования позволили впервые выявить связь между полногеномным грузом слабовредных мутаций и некоторыми сложными фенотипами человека. Было показано, что количество редких делеций в геноме человека коррелирует с интеллектуальными способностями, а именно с уровнем образования человека: чем выше груз делеций, тем ниже уровень образования (Männik et al. 2015). Также доля генома, находящаяся в гомозиготном состоянии, коррелирует с уровнем образования человека: чем выше гомозиготность генома, тем ниже интеллектуальные способности и уровень образования человека (Joshi et al. 2015; Howrigan et al. 2016). Интересно также, что история расселения нашего вида из Африки также может влиять на предрасположенности разных популяций ко многим сложным заболеваниям: например, европейцы, прошедшие несколько популяционных бутылочных горлышек, могли накопить больше слабовредных вариантов в их популяции по сравнению с африканцами (Henn et al. 2016).

В вышеперечисленных работах авторы анализировали общий мутационный груз человека как общее число делеций или общую протяженность гомозиготных областей генома, не вводя никаких дополнительных предположений о взаимодействии слабовредных вариантов друг с другом. Недавним дополнительным прорывом в изучении мутационного груза явилось открытие частого и сильного отрицательного эпистаза между слабовредными вариантами (Sarkisyan et al. 2016). Математически было доказано (Kondrashov 1995), что отбор, действующий против эпистатически взаимодействующих слабовредных мутаций, может изменять форму распределения числа слабовредных вариантов в геноме, что приводит к уменьшению дисперсии данного распределения (пороговый эффект). Таким образом, отношение дисперсии к среднему будет ниже для вариантов, находящихся под действием отбора, чем для нейтральных мутаций (Kondrashov 1995). Пользуясь данным математическим аппаратом было показано, что в среднем по всему геному слабовредные варианты имеют тенденцию усиливать эффект друг друга (Sohail et al. 2017), а не быть независимыми (аддитивными), как считалось ранее. Данные открытия эпистаза позволяют нам выдвинуть рабочую гипотезу, согласно которой груз слабовредных вариантов приводит к пороговым эффектам: в простейшем случае индивидуумы с грузом меньшим чем пороговое значение здоровы, тогда как индивидуумы с грузом большим чем пороговое значение отсутствуют (встречаются реже, чем ожидается) в здоровой когорте по причине того, что либо болеют либо были элиминированы отбором. Эффект эпистаза приводит к неслучайности распределения вредных мутаций внутри каждого генома человека, что может быть обнаружено и оценено при анализе больших популяционных данных, пользуясь знаниями геномики, эволюции и статистики. Мы планируем использовать этот подход в нашей первой задаче.

Недавно мы расширили этот подход, экстраполируя отрицательный эпистаз как основной тип взаимодействия между сильновредной мутацией и фоновым грузом слабовредных вариантов в геноме. Мы предположили, что эффект сильновредной мутации может быть частично компенсирован пониженным грузом слабовредных мутаций во всем геноме. В этом случае мы ожидаем, что носители сильновредной

мутации могут выжить (иметь ненулевую приспособленность) только в случае пониженного мутационного груза в их геноме. Мы сформулировали гипотезу, согласно которой в случайной популяционной выборке сильно вредная мутация (далее “генетический гандикап”) может сегрегировать преимущественно в геномах с пониженным мутационным грузом и, таким образом, генетический гандикап можем являться маркером пониженного груза слабовредных вариантов в геноме (Popadin et al. 2018). Используя трисомию хромосомы 21 (синдром Дауна) как модель генетического гандикапа в популяции человека, мы протестировали нашу гипотезу. Известно, что около 80% эмбрионов с трисомией хромосомы 21 спонтанно abortируется, и лишь 20% могут родиться. Мы предположили, что пониженный геномный груз слабовредных вариантов может увеличивать вероятность эмбриона с трисомией быть живорожденным. Сравнив выборку европейцев с синдромом Дауна с контрольной выборкой (здоровые европейцы), мы обнаружили, что полногеномный груз слабовредных вариантов действительно понижен у когорты людей с синдромом Дауна (Popadin et al. 2018). Таким образом, наши результаты подтверждают, что сильновредная мутация - на примере трисомии хромосомы 21- взаимодействует эпистатически с фоновым грузом слабовредных вариантов вместе (а не независимо), определяя вероятность рождения ребенка. Мы планируем исследовать детально взаимодействие сильновредной мутации с мутационным грузом в нашей второй и третьей задачах.

В нашем исследовании мы сравнили две независимые и неродственные выборки: здоровые европейцы и европейцы с синдромом Дауна. Однако, данный подход имеет множество недостатков, связанных со сложностью учета популяционных особенностей сравниваемых выборок. Существует другой подход, который позволяет анализировать наследование мутационного груза от родителей к детям - данный подход лишен недостатков популяционных сравнений, однако требует наличия семейных данных. Уже давно известен тест неравновесной передачи аллелей (transmission disequilibrium test), который позволяет анализировать силу отбора против вредного аллеля. Со временем этот тест модифицируется и в данный момент аналог этого теста может применяться на полногеномных данных. Имея в наличии тысячи генотипированных троек (оба родителя и ребенок) в Исландии было показано, что вероятность наследования мутаций, приводящих к потере функции гена ниже чем ожидается (23.5% вместо ожидаемых 25% детей-гомозигот от родителей гетерозигот) (Sulem et al. 2015). Недавно был сделан следующий шаг в эволюции данного теста - было показано, что на основе полногеномных данных можно оценить генетическую предрасположенность родителей к сложным заболеваниям (polygenic risk score) и далее оценить наследуемость этих генетических рисков внутри каждой семьи от родителей к детям (polygenic transmission disequilibrium) (Weiner et al. 2017). При этом, для каждого ребенка можно сравнить ожидаемые наследуемые риски (полученные при анализе геномов родителей) с наблюдаемыми (рассчитанными по геному ребенка), и отличия между этими параметрами будут говорить о силе отбора против многочисленных слабовредных вариантов. Данный подход мы планируем использовать в нашей второй задаче.

Цитируемая литература (список относится ко всему тексту):

- Charlesworth B. 1990. Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet Res* **55**: 199–221.
- Euesden J, Lewis CM, O'Reilly PF. 2015. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**: 1466–1468.
- Gunbin K, Peshkin L, Popadin K, Annis S, Ackermann RR, Khrapko K. 2017a. Data on the time of integration of the human mitochondrial pseudogenes (NUMTs) into the nuclear genome. *Data Brief* **13**: 536–544.
- Gunbin K, Peshkin L, Popadin K, Annis S, Ackermann RR, Khrapko K. 2017b. Integration of mtDNA pseudogenes into the nuclear genome coincides with speciation of the human genus. A hypothesis. *Mitochondrion* **34**: 20–23.
- Henn BM, Botigué LR, Peischl S, Dupanloup I, Lipatov M, Maples BK, Martin AR, Musharoff S, Cann H, Snyder MP, et al. 2016. Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. *Proc Natl Acad Sci U S A* **113**: E440–9.
- Howrigan DP, Simonson MA, Davies G, Harris SE, Tenesa A, Starr JM, Liewald DC, Deary IJ, McRae A, Wright

- MJ, et al. 2016. Genome-wide autozygosity is associated with lower general cognitive ability. *Mol Psychiatry* **21**: 837–843.
- Jacquemont S, Coe BP, Hersch M, Duyzend MH, Krumm N, Bergmann S, Beckmann JS, Rosenfeld JA, Eichler EE. 2014. A Higher Mutational Burden in Females Supports a “Female Protective Model” in Neurodevelopmental Disorders. *Am J Hum Genet* **94**: 415–425.
- Joshi PK, Esko T, Mattsson H, Eklund N, Gandin I, Nütle T, Jackson AU, Schurmann C, Smith AV, Zhang W, et al. 2015. Directional dominance on stature and cognition in diverse human populations. *Nature* **523**: 459–462.
- Kondrashov AS. 1995. Dynamics of unconditionally deleterious mutations: Gaussian approximation and soft selection. *Genet Res* **65**: 113–121.
- Männik K, Mägi R, Macé A, Cole B, Guyatt AL, Shihab HA, Maillard AM, Alavere H, Kolk A, Reigo A, et al. 2015. Copy number variations and cognitive phenotypes in unselected populations. *JAMA* **313**: 2044–2054.
- Narasimhan VM, Hunt KA, Mason D, Baker CL, Karczewski KJ, Barnes MR, Barnett AH, Bates C, Bellary S, Bockett NA, et al. 2016. Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**: 474–477.
- Popadin K, Gunbin KV, Khrapko K. 2014. Mitochondrial DNA mutations and cancer: lessons from the parathyroid. *Am J Pathol* **184**: 2852–2854.
- Popadin K, Peischl S, Garieri M, Sailani MR, Letourneau A, Santoni F, Lukowski SW, Bazykin GA, Nikolaev S, Meyer D, et al. 2018. Slightly deleterious genomic variants and transcriptome perturbations in Down syndrome embryonic selection. *Genome Res* **28**: 1–10.
- Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. 2007. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc Natl Acad Sci US A* **104**: 13390–13395.
- Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, et al. 2016. Local fitness landscape of the green fluorescent protein. *Nature* **533**: 397–401.
- Skuratovskaia DA, Sofronova JK, Zatolokin PA, Popadin KY, Vasilenko MA, Litvinova LS, Mazunin IO. 2018. Additional evidence of the link between mtDNA copy number and the body mass index. *Mitochondrial DNA A DNA Mapp Seq Anal* 1–5.
- Sohail M, Vakhrusheva OA, Sul JH, Pulit SL, Francioli LC, Genome of the Netherlands Consortium, Alzheimer’s Disease Neuroimaging Initiative, van den Berg LH, Veldink JH, de Bakker PIW, et al. 2017. Negative selection in humans and fruit flies involves synergistic epistasis. *Science* **356**: 539–542.
- Sulem P, Helgason H, Oddson A, Stefansson H, Gudjonsson SA, Zink F, Hjartarson E, Sigurdsson GT, Jonasdottir A, Jonasdottir A, et al. 2015. Identification of a large set of rare complete human knockouts. *Nat Genet* **47**: 448–452.
- Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, Samocha KE, Goldstein JL, Okbay A, Bybjerg-Grauholm J, et al. 2017. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet* **49**: 978–985.
- Wu N, Ming X, Xiao J, Wu Z, Chen X, Shinawi M, Shen Y, Yu G, Liu J, Xie H, et al. 2015. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *N Engl J Med* **372**: 341–350.
- Zahavi A. 1975. Mate selection—A selection for a handicap. *J Theor Biol* **53**: 205–214.

Цель и задачи проекта (развернутое описание; форма изложения должна дать возможность эксперту оценить новизну идеи проекта и междисциплинарность исследования)

Основная цель нашего проекта - изучение отбора и анализ пороговых эффектов в распределении слабовредных вариантов в геноме человека. Для этого, пользуясь инструментами биоинформатики и доступными популяционными и семейными геномными данными, мы изучим взаимодействия между разными классами вредных мутаций в геноме человека.

Мы выделяем три основные задачи: (i) анализ распределения делеций в популяциях человека; (ii) анализ наследования полигенных рисков в семьях с ребенком с синдромом Дауна; (iii) анализ модели генетического гандикапа на разных временных масштабах.

1) Анализ распределения делеций в популяциях человека.

Пользуясь генотипированными данными доступными для 500 000 индивидуумов из Английского Биобанка и 50 000 индивидуумов из Эстонского Биобанка мы планируем восстановить делеции для каждого генома и далее проанализировать их распределение. Мы выделяем две основные подзадачи в данной задаче:

1А) Взаимодействие делеций друг с другом - наблюдаемое распределение уже чем ожидаемое

Если делеции распределены случайно в геноме человека, то число делеций в каждом геноме должно быть распределено по закону Пуассона (рисунок 3). Более узкое наблюдаемое распределение чем ожидаемое распределение Пуассона должно говорить о взаимодействии делеций друг с другом (Kondrashov 1995; Sohail et al. 2017). В простейшем случае, если делеции А и В встречается в нашей выборке с частотами 10% и находятся на разных хромосомах (то есть наследуются абсолютно независимо), то мы ожидаем, что частота носителей обоих делеций (А и В) должна быть равна произведению частот независимых делеций (то есть 1%). Однако, если в анализируемой популяции частота носителей обоих делеций значимо меньше чем 1%, то данный результат при корректных статистических проверках может быть интерпретирован как свидетельство отрицательного эпистаза - когда вредный эффект двух вариантов существенно сильнее нежели независимые эффекты этих вариантов. Биологическая интерпретация в простейшем случае может быть такова, что существуют два метаболических пути, которые взаимозаменяемы и каждая отдельная делеция слабо вредная, поскольку инактивирует лишь один из двух путей, однако две делеции (А и В) инактивируют оба пути что приводит к эффекту близкому к летальному. В результате в анализируемой популяции людей мы не обнаружим (или обнаружим значимо меньше чем ожидаем) таких двойных комбинаций. Дополнительно, мы планируем ввести в наш анализ такие параметры как длина делеции, природа удаленных генов (гаплонедостаточность, эволюционная консервативность), частота делеций в популяции.

1Б) Взаимодействие между делециями и регуляторными однонуклеотидными вариантами.

Мы ожидаем, что вредность гетерозиготной делеции может быть модифицирована свойствами оставшейся копии гена: уровнем экспрессии оставшейся копии (рисунок 2) а также “качеством” оставшейся копии - количеством слабовредных несинонимических вариантов в этой копии. Чем выше уровень экспрессии оставшейся копии и чем “качественнее” оставшаяся копия, тем больше шансов, что носитель данной гетерозиготной делеции не будет элиминирован отбором. Таким образом, анализируя распределение делеций в популяции человека, мы ожидаем увидеть неслучайность в распространении делеций - они будут наблюдаться чаще в случаях, когда вторая копия гена высоко экспрессируется и более качественная чем популяционный контроль (копии в диплоидной выборке).

Например, если ген в норме присутствует в геноме в двойном (диплоидном) наборе, но вследствие гетерозиготной делеции был частично или полностью удален с одной хромосомы, то теоретически ожидаемый уровень экспрессии будет 50%. Однако, если оставшийся аллель имеет регуляторный вариант, увеличивающий уровень экспрессии гена (gain of expression cis-eQTL), то он будет частично компенсировать

эффект делеции, увеличивая уровень экспрессии с ожидаемых 50% до, например, 60%. В этом случае среди носителей делеции в популяции мы ожидаем увидеть повышенную частоту данных регуляторных вариантов, частично компенсирующих делецию (рисунок 2). Важно отметить, что регуляторный вариант, увеличивающий уровень экспрессии гена, становится полезным при условии удаления второго аллеля. В контексте предлагаемого проекта интересно будет оценить на масштабе полного генома человека, как часто мы видим такого рода компенсации.

2) Анализ наследования полигенных рисков в семьях с ребенком с синдромом Дауна.

Анализируя выборку людей с трисомией хромосомы 21 мы показали пониженный груз слабовредных вариантов по сравнению с контрольной группой (Popadin et al. 2018). Однако, семейные данные открывают возможность для новых статистически более мощных тестов поскольку позволяют сравнивать генетические свойства детей с родителями. От содружественной лаборатории в США (Stephanie Sherman, Department of Human Genetics, Emory University School of Medicine, USA) мы получили доступ к гентипированным данным семей с ребенком с трисомией хромосомы 21 (650 троек: оба родителя и ребенок с полной трисомией по хромосоме 21). Мы планируем провести три анализа, основанных на семейных данных:

=> полигенное наследование слабовредных однонуклеотидных вариантов

Мы оценим дефицит или избыток слабовредных однонуклеотидных вариантов, унаследованных детьми с синдромом Дауна от родителей пользуясь новым полногеномным тестом по неравновесной передаче слабовредных аллелей (Weiner et al. 2017). В соответствии с нашей гипотезой мы ожидаем увидеть дефицит унаследованных слабовредных вариантов по сравнению со средним числом слабовредных вариантов, обнаруженных у родителей.

=> наследование делеций от родителей во всех аутосомах кроме хромосомы 21

Мы восстановим груз делеций в родителях и детях и проверим, наследует ли ребенок случайное (среднее) число делеций или меньшее чем случайное.

=> наследование регуляторных вариантов (cis-eQTLs) от родителей к ребенку с синдромом Дауна

Генетические варианты, уменьшающие уровень экспрессии генов, находящихся на хромосоме 21, и генов, положительно регулируемых продуктами генов этой хромосомы, могут частично компенсировать эффект трисомии. Мы оценим вероятность их наследования от родителей к детям. В соответствии с нашей гипотезой мы ожидаем увидеть более частое чем ожидается наследование генетических вариантов, уменьшающих уровень экспрессии генов на хромосоме 21.

3) Математический и компьютерный анализ модели генетического гандикапа.

Предложенная нами модель генетического гандикапа показывает что пониженный груз слабовредных вариантов может хотя бы частично компенсировать эффект сильновредной мутации (Popadin et al. 2018). В данном проекте мы планируем провести дополнительные и более глубокие математические популяционные анализы по взаимодействию мутационного груза и сильновредной мутации.

=> Мы введем стохастический дрейф и стохастическую (задающуюся определенным статистическим распределением вероятности) пенетрантность генетического гандикапа, свойственную реальным биологическим популяциям с конечной эффективной численностью, и сможем определить, при какой эффективной численности и какой вероятности проявления признака эффект гандикапа присутствует, а при какой - отсутствует;

=> Некоторые варианты могут изменить свой эффект на приспособленность из-за присутствия гандикапа и стать слабополезными из нейтральных или слабовредных и наоборот, кроме того, мутации составляющие мутационный груз, очевидно, имеют разную вредность (разный коэффициент отбора). В наших портретных симуляциях мы сможем изучить, учитывая эпистатический эффект мутаций, а также его экспрессивность и

пенетрантность, при какой доле безусловно и условно вредных мутаций и какой вероятности проявления эпистатических взаимодействий мы все еще увидим эффект гандикапа, и как влияют различные распределения вредности мутаций генетического груза на эффект гандикапа.

=> Изначально мы ввели эффект гандикапа в рамках одного синхронизированного во времени поколения. Однако, в популяционной модели мы планируем проследить судьбу генов, прошедших через носителя гандикапа в условиях перекрывающихся во времени поколений. Если носители гандикапа могут размножаться, и гандикап – менделирующий признак, то часть потомства потеряет гандикап и будет иметь очень высокую приспособленность из-за низкого числа фоновых слабовредных мутаций. Мы изучим, при каких условиях генетический гандикап сможет сегрегировать в популяции в условиях перекрывающихся во времени поколений.

=> Генетический гандикап, эффективно элиминируя низко-приспособленных особей, может носить очищающий эффект и может увеличивать среднюю популяционную приспособленность, особенно в случае стохастической потери гандикапа через несколько поколений в случае дрейфа генов. Сопоставляя две популяции: контроль и экспериментальную популяцию, в которой стохастически появляются и исчезают генетические гандикапы, мы собираемся сравнить их долгосрочные средние приспособленности и ответить на вопрос: при каких условиях частое появление гандикапа в популяции может быть полезно для данной популяции (в терминах повышенной приспособленности особей, прошедших отбор).

Предлагаемые подходы и методы, их обоснование для реализации цели и задачи исследований (Развернутое описание предлагаемого исследования; форма изложения должна дать возможность эксперту оценить соответствие подходов и методов исследования поставленным целям и задачам, новизну и междисциплинарность идеи)

Предлагаемый нами подход базируется на компенсации (Popadin et al. 2018). Рассмотрим популяцию генетически гетерогенных организмов (рисунок 1). Изначальное распределение слабовредных мутаций отображено серой кривой (правый хвост распределения – организмы с большим количеством слабовредных мутаций и низкой приспособленностью; левый хвост распределения – организмы с низким числом слабовредных мутаций, и, соответственно, высокой приспособленностью). Далее, предположим, что из контрольной популяции мы получаем экспериментальную популяцию, внося в каждый геном генетический гандикап (сильновредную мутацию). Гандикап по своей вредности эквивалентен многим слабовредным мутациям и, поэтому, распределение слабовредных мутаций сдвинется вправо, что отображено красной пунктирной кривой. Далее, отбор элиминирует особей в обеих популяциях, базируясь на мутационном грузе – чем выше груз, тем выше вероятность того, что данный организм не оставит потомство. Для простоты мы предполагаем отсекающий отбор, элиминирующий всех особей с количеством мутаций больше некоторого порога. После отбора существенная часть экспериментальной популяции элиминируется, тогда как большинство организмов из контрольной популяции выживает. В соответствии с нашей гипотезой мы видим, что количество слабовредных мутаций в геномах экспериментальной популяции (выжившие с гандикапом, отмечены красным фоном) меньше, чем в контрольной популяции (выжившие в контрольной популяции, отмечены серым фоном). Важно подчеркнуть, что сам по себе гандикап является сильновредной мутацией, и мы не учитываем его при сравнении груза слабовредных мутаций экспериментальной и контрольной популяций. Сравнивая все организмы, прошедшие отбор в контрольной и экспериментальной популяции, мы видим, что присутствие гандикапа связано с низким числом слабовредных мутаций и таким образом с высокой потенциальной приспособленностью данного генома (приспособленность была бы высокой если бы не было гандикапа). Таким образом, (i) число слабовредных мутаций у носителей гандикапа ниже чем у контрольных организмов. Далее, мы можем заметить, что (ii) чем сильнее гандикап (длина стрелки), тем сильнее ожидаемые отличия между грузом слабовредных мутаций среди выживших в экспериментальной и контрольной популяции.

Важно подчеркнуть, что все наши три задачи (анализ делеций, анализ наследования мутационного груза в семьях с ребенком с синдромом Дауна и математическое описание модели гандикапа) соответствуют нашей модели гандикапа. Отличия лишь в силе гандикапа - это либо трисомия, либо делеции (более слабый гандикап).

Модель генетического гандикапа - новая концепция, разработанная нами (Popadin et al. 2018). Она аналогична концепции эволюционного гандикапа Амоса Захави (Zahavi 1975), где вредный признак рассматривается как маркер высокой приспособленности организма. Классическая популяционная генетика, возможно, не рассматривала проблему сильновредных мутаций так детально, как это планируем мы, поскольку, очевидно, что сильновредные мутации должны быстро и эффективно удаляться очищающим отбором. Однако, наш основной интерес состоит не в том, как долго сильновредная мутация может сегрегировать в популяции, а в том, что присутствие этой мутации у выживших организмов (даже если это происходит в течение всего лишь одного поколения) может служить маркером исходно низкого груза слабовредных мутаций данной особи. Более долгосрочные потенциальные эффекты генетического гандикапа абсолютно не изучены. В нашем проекте мы планируем провести математические эксперименты, чтобы оценить насколько сильно генетический гандикап может влиять на микроэволюционные события.

Основной подход медицинской генетики - объяснить больной фенотип за счет вредных аллелей, ассоциированных с данным фенотипом. Наш подход противоположный - попытаться найти компенсации, облегчающие вредный эффект сильной мутации. Хотя все медицинские генетики понимают, что существует огромная изменчивость в фенотипе пациентов с одной и той же мутацией, и, видимо, фоновый груз слабовредных мутаций может объяснить существенную долю этой изменчивости; эмпирических работ, показывающих такие компенсации, очень мало. Недавно появились первые работы, анализирующие здоровых людей с очень вредными мутациями (Narasimhan et al. 2016), и возрос интерес к геномным вариантам, которые могут компенсировать вредные мутации у человека.

Ожидаемые результаты научного исследования и их научная значимость, возможный инновационный потенциал ожидаемых результатов с детализацией ожидаемых результатов за первый год реализации проекта (форма изложения должна дать возможность провести экспертизу результатов и оценить оригинальность, уникальность (т.е. получение которых возможно только при реализации данного проекта), междисциплинарный характер исследования, соответствие ожидаемых результатов задачам, описанным в аннотации темы),

Данный проект является логичным продолжением нашей недавней статьи, посвященной людям с синдромом Дауна. Мы расширяем применимость нашей концепции генетического гандикапа на другие вредные варианты (делеции, сегрегирующие среди здоровых людей), опробуем и совершенствуем новые подходы (наследование мутационного груза от родителей к детям с синдромом Дауна), а также математически исследуем новые возможные последствия генетического гандикапа на мутационный груз популяции модельных организмов (рассматриваем возможность того, что потомки от родителей с гандикапом могут оказаться очень высоко приспособленными - без самого гандикапа, но с низким фоном слабовредных вариантов).

Наша оригинальная гипотеза генетического гандикапа, уникальные данные которые мы получили от нашей содружественной лаборатории (семьи с ребенком Дауном) и владение современными компьютерными методами обработки данных (предсказание полигенных рисков, наследование мутационного груза) делают данный проект инновационным и актуальным. Все анализы будут проводиться на компьютерном кластере МГУ “Макарьич” (БФУ имени И. Канта заключил договор с МГУ на пользование кластером). Никаких экспериментов в данном проекте не планируется.

В первый год мы планируем решить следующие вопросы из наших трех задач:

- 1) Мы проаннотируем все делеции и дупликации для каждого индивидуума из английского и эстонского биобанка, проведем контроль качества полученных данных и подготовим первые описательные статистики распределения делеций и дупликаций в популяции. Это уже нам позволит ответить на ряд интересных вопросов: сколько делеций и дупликаций в среднем содержит каждый геном? Какова их длина и расположение в геноме? Есть ли отличия в грузе делеций между полами (здесь интересно упомянуть модель согласно которой женский геном может толерировать большее количество делеций оставаясь здоровым - female protective model (Jacquemont et al. 2014)) и возрастами (некоторые делеции могут быть соматическими и тогда мы увидим избыток такого типа делеций у пожилых людей)? Отличается ли средняя длина делеций в геномах с одной, двумя, тремя и более делециями (согласно нулевой гипотезе средняя длина делеций должна быть одинаковой, однако, если происходит эпистатический отбор - средняя длина может падать с числом делеций)? Соответствует ли полученное распределение числа делеций в геноме распределению Пуассона или нет?
- 2) Для оценки полигенных рисков болезней для каждого из 1950 генома (650*3) мы скачаем полный каталог GRASP (каталог, где систематически хранятся результаты всех больших genome-wide association studies). В каталоге GRASP хранится информация о полиморфизмах, ассоциированных с сотней частых болезней (диабет второго типа, болезнь Альцгеймера, шизофрения и так далее) а также сложных фенотипов (рост, предрасположенность к высоким интеллектуальным способностям, продолжительность жизни и так далее) - смотрите полный список доступных фенотипов в GRASP каталоге (<https://grasp.nhlbi.nih.gov/Overview.aspx>). Используя веса полиморфизмов мы оценим генетические предрасположенности (polygenic risk score) к сотне фенотипов для каждого генома - для каждого члена семьи из доступных 650 семей. Для этого мы будем использовать программы, такие как PLINK, а также программы по оценке полигенного риска (Euesden et al. 2015). Далее для каждого ребенка сравнивая ожидаемые генетические риски и наблюдаемые (подход описан в статье (Weiner et al. 2017)), мы получим контраст, который будет говорить о том, что ребенок унаследовал больше или меньше аллелей, ассоциированных с тем или иным фенотипом. После пермутаций и дополнительных анализов (смотрите логику также в (Weiner et al. 2017)) мы выделим несколько фенотипов, генетические варианты которых значимо пере- или

недо- унаследованы детьми с синдромом Дауна. Список таких фенотипов будет представлять собой уникальную информацию. Например, согласно нашим первоначальным данным (на данный момент мы проверили лишь десять разных фенотипов), дети с синдромом Дауна от своих родителей наследуют повышенное число аллелей (повышенный риск), ассоциированных с высокой скоростью прохождения 10 метрового участка (gate speed - измеряется время за которое человек проходит десять метров начиная движение с места и заканчивая полной остановкой; несмотря на свою простоту эта характеристика, этот фенотип хорошо коррелирует с общим статусом здоровья пожилого населения). О чем это может говорить? Возможно, что тысячи аллелей, ассоциированные с высокой скоростью ходьбы, взаимодействуют с трисомией и увеличивают вероятность того, что данный эмбрион с трисомией будет живорожденным. Другими словами, мы сможем изучать генетическую компоненту эмбрионального отбора плодов с трисомией. С фундаментальной точки зрения это нам позволит выявить сложные взаимодействия между трисомией и грузом аллелей, ассоциированных с тем или иным фенотипом (например gate speed). С социальной точки зрения это нам позволит реконструировать фенотип ребенка с синдромом Дауна и ответить на вопрос - каким бы он был, если бы у него не было трисомии (в расчет полигенных рисков мы не берем хромосому 21)?

- 3) Мы планируем промоделировать отбор против генетического гандикапа вводя в нашу симуляционную модель (i) разные коэффициенты отбора против слабовредных вариантов, (ii) разные коэффициенты эпистаза между слабовредными вариантами а также между слабовредными мутациями и генетическим гандикапом, (iii) разную долю мутаций условно и безусловно вредных. Результаты этой модели покажут нам насколько универсальна может быть наша гипотеза гандикапа (насколько применима к разным организмам в разных условиях среды).

В целом, наши результаты прольют свет на взаимодействие между разными классами вредных вариантов, а именно - между сильновредной мутацией и грузом слабовредных мутаций. Наши подходы могут быть использованы для увеличения эффективности персонализированной и превентивной медицины, поскольку помогут выявлять по генетическим данным подвыборки людей, предрасположенных к разным заболеваниям.

Общий план исследований на весь срок реализации проекта (форма представления информации должна дать возможность эксперту оценить реализуемость заявленного плана исследований и риски его невыполнения; общий план реализации проекта даётся с разбивкой по годам с учетом вклада каждого из участников коллектива, предполагаемых поездок)

Все работы будут проведены на вычислительном кластере МГУ с использованием наших собственных программ, которые реализованы на языках PERL, Python и R.

1) Анализ распределения делеций в популяциях человека.

Основные исполнители: Константин Гунбин, Константин Попадьин и студенты.

=> Первый год: аннотация делеций и дупликаций в каждом индивидууме из английского и эстонского биобанков. Контроль качества полученных данных. Описательная статистика.

=> Второй год: анализ ассоциаций регуляторных вариантов с делециями и дупликациями.

Планируется посещение конференций: ASHG (American Society of Human genetics) и BGRS\SB (Bioinformatics of Genome Regulation and Structure\Systems Biology).

=> Третий год: анализ ассоциации груза несинонимических мутаций с делециями и дупликациями.

Планируется посещение конференций: ASHG (American Society of Human genetics) и MCCMB (Moscow Conference on Computational Molecular Biology).

2) Анализ наследования полигенных рисков в семьях с ребенком с синдромом Дауна.

Основные исполнители: Константин Попадьин, Илья Мазунин и студенты.

=> Первый год: полигенное наследование слабовредных одонуклеотидных вариантов (ожидаемые риски сложных полигенных фенотипов против наблюдаемых).

=> Второй год: наследование делеций от родителей к детям во всех аутосомах кроме хромосомы 21.

Планируется посещение конференции - ESHG (European Society of Human genetics) и BGRS\SB (Bioinformatics of Genome Regulation and Structure\Systems Biology);

=> Третий год: наследование регуляторных вариантов (cis-eQTLs) от родителей к ребенку с синдромом Дауна. Планируется посещение конференций: ESHG (European Society of Human genetics) и MCCMB (Moscow Conference on Computational Molecular Biology).

3) Математический и компьютерный анализ модели генетического гандикапа.

Основные исполнители - Константин Гунбин, Константин Попадьин и студенты.

Классическая популяционная генетика будет являться первоосновой для аналитических расчетов и симуляции, однако вопрос, который мы ставим перед собой - новый. Для детерминистических (непрерывных) моделей с и без применения аппарата дифференциальных уравнений будет использована среда моделирования Mathematica или MATLAB. Для стохастических моделей популяционной динамики будет адаптирована новейшая среда SLiM (<https://messerlab.org/slim/>) позволяющая вводить в модель все необходимые детализации микроэволюционного процесса такие как генетический дрейф, эпистаз между мутациями, изменчивость в давлении отбора, структурированность популяции, задавать вероятность экспрессивности и пенетрантности признаков (например, с использованием директив individuals.tag и их статистических распределений).

=> Первый год: стохастические компьютерные симуляции (конечные популяции).

Планируется посещение конференции: SBE (Society for Molecular Biology and Evolution).

=> Второй год: математическая детерминистическая модель. При каких параметрах эффект гандикапа наиболее сильный (эпистаз, распределение коэффициентов отбора вредных мутаций)?

Планируется посещение конференции: SBE (Society for Molecular Biology and Evolution).

=> Третий год: эффект гандикапа в микроэволюции (компьютерные симуляции).

Планируется посещение конференции: SBE (Society for Molecular Biology and Evolution).

Имеющийся у коллектива научный задел по проекту (указываются полученные результаты, разработанные программы и методы, экспериментальное оборудование, материалы и информационные ресурсы, имеющиеся в распоряжении коллектива для реализации проекта)

Наш коллектив сформировался меньше одного года назад, когда Константин Попадьян и Константин Гунбин получили позиции в БФУ и начали читать лекции и работать со студентами. Уже сейчас у нас есть несколько интересных результатов.

Основная концепция проекта базируется на нашей недавней статье (Popadin et al. 2018).

Наша научная команда имеет успешный опыт совместной работы в прошлом. Руководитель проекта опубликовал совместные статьи в высокорейтинговых журналах с каждым из основных исполнителей: с Константином Гунбиным: (Popadin et al. 2007, 2014; Gunbin et al. 2017b, 2017a) и Илей Мазуниным (Skuratovskaia et al. 2018).

По каждой из трех задач у нас имеются некоторые наработки.

1) Анализ распределения делеций в популяциях человека.

1А) Пилотный анализ (3953 индивидуумов, Эстонский биобанк, рисунок 3) взаимодействия делеций друг с другом по данным эстонского биобанка показал интересные результаты и был представлен на международном генетическом конгрессе в Киото, Япония в виде устного доклада (Popadin. 2016. Touching the limits of being alive: the distribution of genome-wide CNV loads in healthy human cohort is right truncated due to ongoing purifying selection. 2016. International Congress of Human Genetics. Kyoto, Japan).

Проанализировав все редкие делеции, которые перекрываются с белок-кодирующими генами, мы увидели, что большинство индивидуумов не имеют делеций вообще ($N = 3596$), некоторые имеют только одну делецию ($N = 346$) и только одиннадцать человек содержат 2 делеции. Что мы ожидаем, если делеции распределены случайно и независимо друг от друга? Проведя случайную пертурбацию данных (случайно распределив число делеций на то же самое число геномов) 10000 раз, мы получили ожидаемые величины (рисунок 3). Интересно, что количество ожидаемых индивидуумов с двумя или более делециями больше, чем количество наблюдаемых. Анализ показал, что эффект статистически достоверен ($p < 0.01$). Мы полагаем, что более тонкий “правый хвост” наблюдаемого распределения говорит об отборе, который элиминировал индивидуумы со слишком высоким грузом делеций. На данный момент, имея в 20 раз больше данных из Эстонского биобанка и в 100 раз больше данных из Английского биобанка, мы ожидаем более сильных результатов, которые мы сможем углубить учитывая функцию делетированного района (наличие/отсутствие энхансеров и сайленсеров) и свойства делетированных генов (гаплонедостаточность, толерантность к делециям, количество белковых связей, вовлеченность в разные метаболические пути, консервативность) и трехмерную конформацию района (топологические домены).

1В) Пилотный анализ (120 000 индивидуумов, Английский биобанк, схема анализа на рисунке 2) взаимодействия делеций с регуляторными вариантами показал компенсацию 4% делеций и был представлен на американском обществе человеческой генетики в Орландо, США в виде устного доклада (Popadin K. 2017. Evidence of compensatory variations on the remaining allele of rare deletions. American Society of Human Genetics, Orlando, USA.).

2) Анализ наследования полигенных рисков в семьях с ребенком с синдромом Дауна.

Мы уже получили от наших коллег генотипированные данные для 650 семей. Константин Гунбин со студентами начал анализ полигенных рисков (polygenic risk score) и мы проанализировали полигенные риски десяти сложных фенотипов.

3) Математический и компьютерный анализ модели генетического гандикапа (смотрите рисунок 1).

Ниже мы приводим наши предварительные математические результаты, показывающие, что гипотеза гандикапа работает даже в случае, если контрольная и экспериментальная популяции отбираются независимо друг от друга. Мы доказываем, что в случае отрицательного эпистатического взаимодействия между слабовредными мутациями концепция гандикапа работает. Графически отрицательный эпистаз между слабовредными мутациями отображается в виде “отсекающего отбора”, когда все организмы с мутационным грузом выше некоторого порога элиминируются отбором

Мы начали с классической модели отбора против слабовредных мутаций (Charlesworth 1990). Полагая, что все мутации имеют одну и ту же вредность (коэффициент отбора α), различие в среднем числе слабовредных мутаций до и после отбора описывается следующей формулой:

$$\Delta n = -(\alpha + \beta \cdot n) \cdot V / (1 + \beta \cdot V) \quad (1),$$

где α – коэффициент отбора против слабовредной мутации, β – коэффициент эпистаза между слабовредными мутациями, n и V среднее число и изменчивость в числе слабовредных мутаций перед отбором соответственно.

Полагая, что генетический гандикап эквивалентен k слабовредным мутациям (вредность гандикапа = k), среднее число слабовредных мутаций в экспериментальной популяции увеличивается на k , тогда как изменчивость остаётся прежней. Таким образом, мы можем переписать формулу 1 для организмов из экспериментальной популяции:

$$\Delta n = -(\alpha + \beta \cdot (n+k)) \cdot V / (1 + \beta \cdot V) \quad (2)$$

Далее, для оценки эффекта гандикапа (H) как разницы между эффективностями отбора в экспериментальной и контрольной популяциях мы вычитаем Δn из Δn :

$$H = \Delta n - \Delta n = V \cdot \beta \cdot k / (1 + \beta \cdot V) \quad (3)$$

Мы видим, что, если нет изменчивости в числе слабовредных мутаций в популяции перед отбором ($V=0$), или нет эпистаза ($\beta = 0$), генетический гандикап не работает ($H = 0$). Однако, если обе переменные – дисперсия и эпистаз – выше нуля, то эффект гандикапа (H) увеличивается линейно с его вредностью (k).

Чтобы отобразить данные результаты в виде графика и сделать их близкими к человеческой популяции, мы предположили, что средний человек в своем геноме содержит 1000 слабовредных мутаций, популяционное распределение близко к распределению Пуассона, коэффициент эпистаза меняется от 0 (нет эпистаза) до 0.1 (Charlesworth 1990), и вредность гандикапа увеличивается от 0 до 200 эквивалентов слабовредных мутаций. В итоге мы видим, что эффект гандикапа увеличивается как с вредностью гандикапа, так и с силой эпистаза. Мы заключили, что эффект гандикапа при биологически осмысленных мутационных параметрах работает и может достигать 20% общего мутационного груза (200 из 1000).

На первом году проекта мы планируем провести компьютерные симуляции, которые нам могут подсказать новые интересные параметры или комбинации параметров, которые мы сможем включить в детерминистическую модель. На втором году мы планируем доработать данную модель и опубликовать её. В настоящее время уже создан прототип модели в среде SLiM, учитывающей отрицательный эпистаз и имеющей возможность задания стохастическим образом вероятности проявления признака у разных особей популяции.

Список основных публикаций членов коллектива, наиболее близко относящихся к предлагаемому проекту за последние 5 лет (не более 15, каждая с новой строки, с указанием импакт-фактора журнала за 2017 год и базы данных ИФ

Popadin K., Peischl S., Garieri M., Sailani MR, Letourneau. A., Santoni F., Lukowski SW, Bazykin G., Nikolaev S., Meyer D., Excoffier L., Reymond A., Antonarakis SE. Slightly deleterious genomic variants and transcriptome perturbations in Down syndrome embryonic selection. *Genome Research*, 2018. Jan;28(1):1-10. doi: 10.1101/gr.228411.117

Skuratovskaia DA, Sofronova JK, Zatolokin PA, Popadin KY, Vasilenko MA, Litvinova LS, Mazunin IO. Additional evidence of the link between mtDNA copy number and the body mass index. **Mitochondrial DNA A DNA Mapp Seq Anal.** 2018 Feb 10:1-5. doi: 10.1080/24701394.2018.1436170.

Maria Nicla Loviglio, Thomas Arbogast, Aia Elise Jøneh, Stephan C. Collins, Konstantin Popadin, Camille S. Bonnet, Giuliana Giannuzzi, Anne M. Maillard, Sébastien Jacquemont, 16p11.2 Consortium, Binnaz Yalcin, Nicholas Katsanis, Christelle Golzio, Alexandre Reymond. The immune signaling adaptor LAT contributes to the neuroanatomical phenotype of 16p11.2 BP2-BP3 CNVs. **AJHG**, 2017.

Gaspar, Howald, Popadin, Maier, Mauvoisin, Moriggi, Gutierrez-Arcelus, Falconnet, Borel, Kunz, Kramer, Gachon, Dermitzakis, Antonarakis, Brown. The genomic landscape of human cellular circadian variation points to a novel role for the signalosome. **Elife**. 2017 Sep 4;6. pii: e24994. doi: 10.7554/eLife.24994.

Gunbin K, Peshkin L, Popadin K, Annis S, Ackermann RR, Khrapko K. Data on the time of integration of the human mitochondrial pseudogenes (NUMTs) into the nuclear genome. **Data Brief**. 2017. doi: 10.1016/j.dib.2017.05.024.

Gunbin K, Ponomarenko MP, Suslov VV, Gusev F, Fedonin GG, Rogaev EI. Evolution of Brain Active Gene Promoters in Human Lineage Towards the Increased Plasticity of Gene Regulation. **Mol Neurobiol**. 2018 Mar;55(3):1871-1904. doi: 10.1007/s12035-017-0427-4. Epub 2017 Feb 24.

Gunbin K, Peshkin K, Popadin K, Annis S, Ackermann RR, Khrapko K. Integration of mtDNA pseudogenes into the nuclear genome coincides with speciation of the human genus. A hypothesis. **Mitochondrion**. 2017. doi: 10.1016/j.mito.2016.12.001

Polishchuk LV*, Popadin K*, Baranova M & Kondrashov A. *A genetic component of extinction risk in mammals*, **OIKOS**, 2015. Doi 10.1111/oik.01734

Gunbin KV, Afonnikov DA, Kolchanov NA, Derevianko AP, Rogaev EI. The evolution of Homo sapiens denisova and Homo sapiens neanderthalensis miRNA targeting genes in the prenatal and postnatal brain. **BMC Genomics**. 2015;16 Suppl 13:S4. doi: 10.1186/1471-2164-16-S13-S4. Epub 2015 Dec 16.

Sailani MR, Santoni FA, Letourneau A, Borel C, Makrythanasis P, Hibaoui Y, Popadin KY, Bonilla X, Guipponi M, Gehrig C, Vannier A, Carre-Pigeon F, Feki A, Nizetic D, Antonarakis SE. DNA-Methylation Patterns in Trisomy 21 Using Cells from Monozygotic Twins. **Plos One**. 2015. 10(8):e0135555.

Popadin K, Maria Gutierrez-Arcelus, Tuuli Lappalainen, Alfonso Buil, Sergey I. Nikolaev, Samuel W. Lukowski, Georgii A. Bazykin, Vladimir B. Seplyarskiy, Panagiotis Ioannidis, Evgeny Zdobnov, Emmanouil T. Dermitzakis, Stylianos E. Antonarakis. Gene age predicts the strength of purifying selection acting on gene expression variation in humans. **AJHG**. 2014. 95(6):660-674. **Highlighted IN BRIEF in NATURE REVIEWS GENETICS: "Gene Age Begets Constraints", 2015, vol 16**

Audrey Letourneau, Federico A. Santoni, Ximena Bonilla, M.Reza Sailani, David Gonzalez, Jop Kind, Claire Chevalier, Robert Thurman, Richard S. Sandstrom, Youssef Hibaoui, Popadin K, Emilie Falconnet, Maryline Gagnebin, Corinne Gehrig, Anne Vannier, Michel Guipponi, Laurent Farinelli, Daniel Robyr, Marco Garieri, Eugenia Migliavacca, Christelle Borel, Samuel Deutsch, Anis Feki, John A. Stamatoyannopoulos, Yann Herault, Bas van Steensel, Roderic Guigo, Stylianos E. Antonarakis. *Domains of genome wide gene expression dysregulation in Down syndrome*. **Nature**. 2014. 508(7496):345-50.

Popadin K*, Gutierrez-Arcelus M*, Dermitzakis ET, Antonarakis SE. *Genetic and epigenetic regulation of human LincRNA gene expression*. **AJHG**. 2013. 93(6):1015-26.

Reza Sailani, Periklis Makrythanasis, Armand Valsesia, Federico A. Santoni, Samuel Deutsch, Popadin K, Christelle Borel, Eugenia Migliavacca, Andrew J Sharp, Genevieve Duriaux Sail, Emilie Falconnet, Kelly Rabionet, Clara Serra-Juh, Stefano Vicari, Daniela Laux, Yann Grattau, Guy Dembour, Andre Megarbane, Renaud Touraine, Samantha Stora, Sofia Kitsiou, Helena Fryssira, Chariklia Chatzisevastou-loukidou, Emmanouel Kanavakis, Giuseppe Merla, Damien Bonnet, Luis A. Perez-Jurado, Xavier Estivill, Jean M. Delabar and Stylianos E. Antonarakis. *The complex SNP and CNV genetic architecture of the increased risk of congenital heart defects in Down syndrome*. **Genome Research**. 2013. 23(9):1410-21.

Popadin K*, Nikolaev SI*, Junier T, Baranova M & Antonarakis SE. *Purifying Selection in Mammalian Mitochondrial Protein-Coding Genes Is Highly Effective and Congruent with Evolution of Nuclear Genes*. **Molecular Biology and Evolution**. 2013. 30(2):347-55.

