Классификация мутаций

Точковые мутации

- Замена
  - Синонимичная
  - Миссенс
  - Нонсенс
- Делеция
- Инсерция
- Амлификация тринуклеотидных повторов

Хромосомные мутации

- Транслокация
- Делеция
- Инверсия
- Дупликация

Геномные мутации

- Моносомия
- Трисомия

Johan T den Dunnen

Professor Medical Genomics in the
Leiden University Medical Center

Номенклатура названий мутаций в соответствии с правилами HGVS

Human Genome Variation Society

HGVS
HUMAN GENOME VARIATION SOCIETY

Definitions

- **prevent confusion**
  - *do not use "mutation"*
    - *use variant, disease-associated variant*
  - *do not use "polymorphism"*
    - *use variant, not disease-associated variant*
  - *do not use "pathogenic"*
    - *use disease-associated, a disease-associated variant*

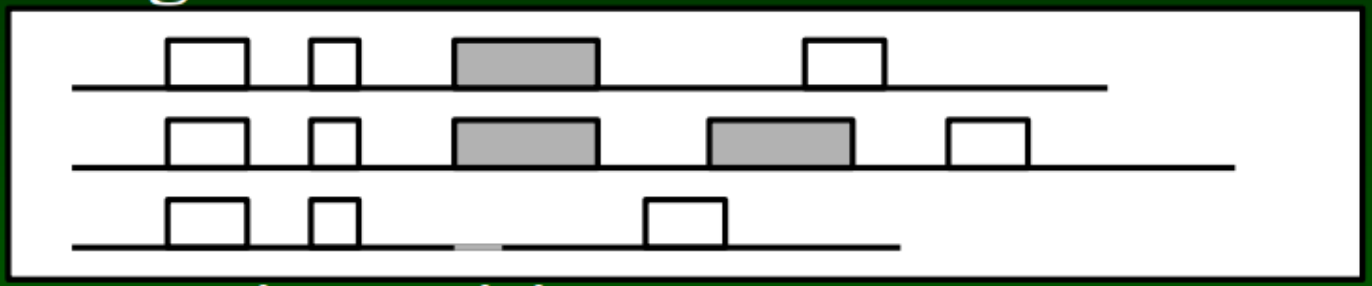
- **better use neutral terms**
  - *sequence variant*
  - *alteration*
  - *CNV (Copy Number Variant)*
  - *SNV (Single Nucleotide Variant, not SNP)*

Variant types

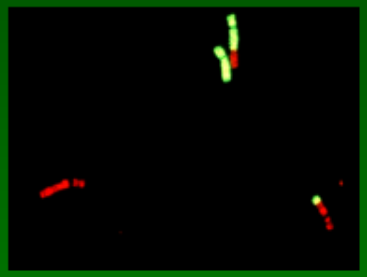
- change in sequence

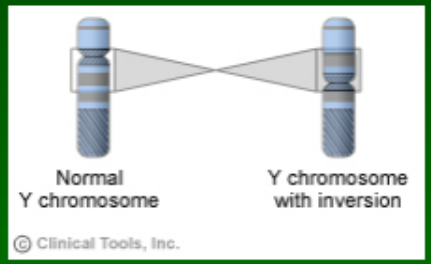
```
ACATCAGGAGAAGATGTTC GAGACTTTGCCA
ACATCAGGAGAAGATGTTT GAGACTTTGCCA
ACATCAGGAGAAGATGTT  GAGACTTTGCCA
ACATCAGGAGAAGATGTTCCGAGACTTTGCCA
```

- change in amount *(Copy Number Variation)*



- change in position





Structural Variation (SV)

Reference sequence

- use official HGNC gene symbols

- provide reference sequence
  - *covering complete sequence*
    - *largest transcript*
  - *preferably a LRG*
    - *e.g. LRG_123*
  - *give accession.version number*
    - *e.g. NM_012654.3*

- indicate type of Reference Sequence
  - *DNA*
    - *coding DNA* *c.*
    - *genomic* *g.*
    - *mitochondrial* *m.*
    - *non-coding RNA* *n.*
  - *RNA* *r.*
  - *protein* *p.*

The LRG

Dalgleish *et al. Genome Medicine* 2010, 2:24
http://genomemedicine.com/content/2/4/24

Genome Medicine

CORRESPONDENCE Open Access

Locus Reference Genomic sequences: an improved basis for describing human DNA variants

Raymond Dalgleish[1]*, Paul Flicek[2], Fiona Cunningham[2], Alex Astashyn[3], Raymond E Tully[3], Glenn Proctor[2], Yuan Chen[2], William M McLaren[2], Pontus Larsson[2], Brendan W Vaughan[2], Christophe Béroud[4], Glen Dobson[5], Heikki Lehväslaiho[6], Peter EM Taschner[7], Johan T den Dunnen[7], Andrew Devereau[5], Ewan Birney[2], Anthony J Brookes[1] and Donna R Maglott[3]

Abstract

As our knowledge of the complexity of gene architecture grows, and we increase our understanding of the subtleties of gene expression, the process of accurately describing disease-causing gene variants has become increasingly problematic. In part, this is due to current reference DNA sequence formats that do not fully meet present needs. Here we present the

Introduction

In 1993 Ernest Beutler
editor of the *American*
lighting the deficienci
describe DNA variants
Human Mutation invit
Tsui to produce a nom
proteins [2]. From the
years have borne witn

EDITORIAL

nature genetics

Conventional wisdom

Recent agreement on stable reference sequences for reporting human genetic variants now allows us to mandate the use of the allele naming conventions developed by the Human Genome Variation Society.

By agreement between stakeholders and two principal databases, it has been proposed (R. Dalgleish *et al.*, *Genome Med.* 2, 24, 2010, doi:10.1186/gm145) that human genetic variants be reported relative to a new set of stable reference sequences, "Locus Reference, Genomic" (LRG, pronounced "large" http://www.lrg-sequence.org/page.php). These sequences have been developed from the initial NCBI RefSeqGene concept and are provided by NCBI and EBI according to agreed rules

age, resequencing and marker association studies and so keep allele descriptions commensurate with the method by which their data were generated.

The LRG reference sequences should be used in conjunction with standard HGNC gene abbreviations (http://www.genenames.org/) that we already require as a condition of publication. All human genetic variants must now be described—in abstracts and at first use—in accor-

EBI, NCBI, Gen2Phen

GEN2PHEN

Сборки генома человека

GRCh37:

First release:
Feb 27, 2009

Latest patch:
Jun 28, 2013 (p13)

GRCh38:

First release:
Dec 24, 2013

Latest patch:
Oct 14, 2014 (p1)

1. Качество
2. «Дырки»
3. Корректное картирование отдельных локусов

Сборки генома человека

May 27, 2021

bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results

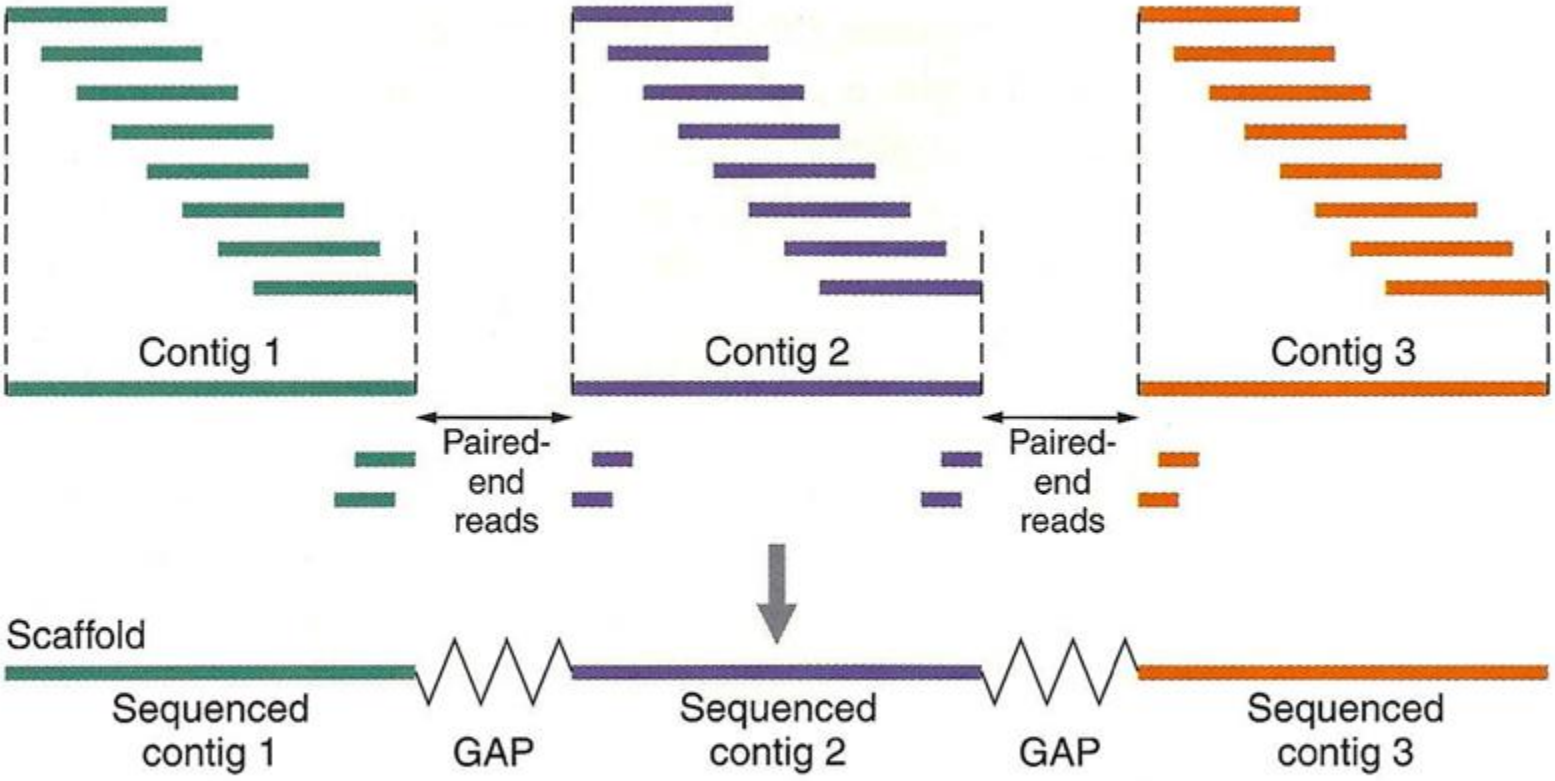
The complete sequence of a human genome

Sergey Nurk, Sergey Koren, Arang Rhie, Mikko Rautiainen, Andrey V. Bzikadze, Alla Mikheenko, Mitchell R. Vollger, Nicolas Altemose, Lev Uralsky, Ariel Gershman, Sergey Aganezov, Savannah J. Hoyt, Mark Diekhans, Glennis A. Logsdon, Michael Alonge, Stylianos E. Antonarakis, Matthew Borchers, Gerard G. Bouffard, Shelise Y. Brooks, Gina V. Caldas, Haoyu Cheng, Chen-Shan Chin, William Chow, Leonardo G. de Lima,

- Практически полностью досеквенированы 8% генома которые не могли никак досеквенировать
- Новая сборка не имеет гэпов для всех 22 аутосом и хромосомы X, с исправлениями многочисленных ошибок и имеет почти 200 миллионов п.н. новой последовательности, содержащей 2226 копий паралогов, 115 из которых, как предполагается, кодируют белок.
- Новые отсеквенированные области включают все центромерные сателлитные массивы и короткие плечи всех пяти акроцентрических хромосом, что впервые открывает доступ к этим сложным областям генома для вариационных и функциональных исследований.

https://www.biorxiv.org/content/10.1101/2021.05.26.445798v1.full.pdf

Контиги и гэпы



Референсные последовательности

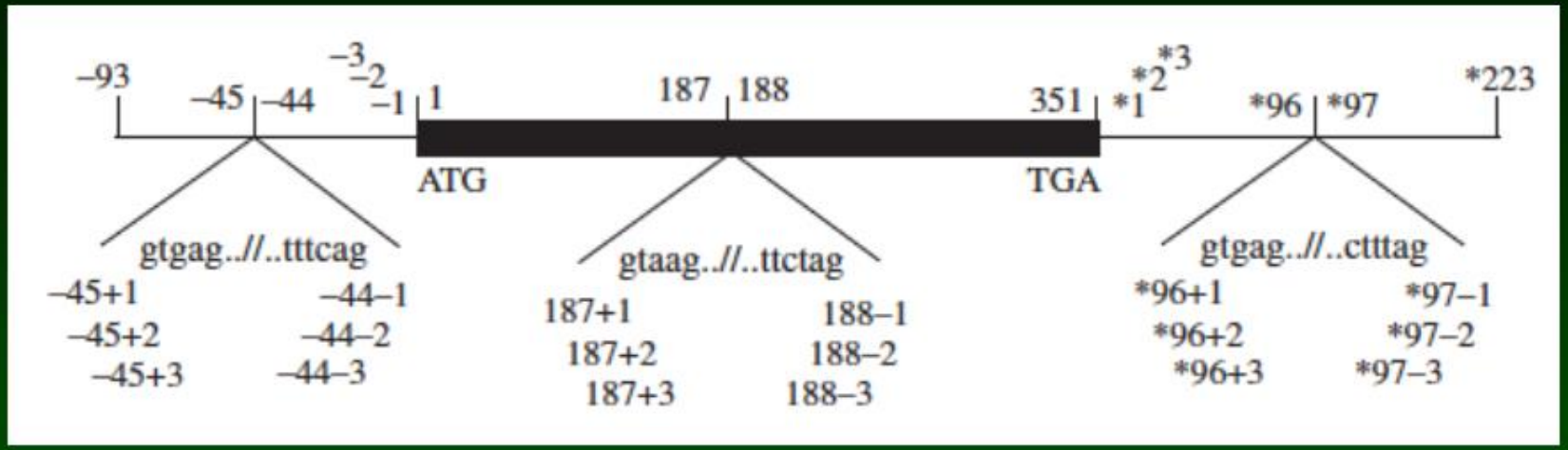
Recommended Reference Sequences types are:

- RefSeq sequences with the prefixes NC_, NT_, NW_,NG_, NM_, NR_ or NP_
  - chromosome - NC_000023.11
  - genomic contigs or scaffolds - NT_010718.17, NW_003315950.2
  - gene/genomic region - NG_012232.1
  - coding transcript - NM_004006.2
  - non-coding transcript - NR_004430.2
  - protein - NP_003997.1
- LRG sequences with the prefixes LRG_#, LRG_#t#, LRG_#p# (see examples below)
  - gene/genomic region - LRG_199
  - coding transcript (or non-coding transcript) - LRG_199t1
  - protein - LRG_199p1
- Ensembl transcript (ENST) and protein (ENSP) which are not identified by Ensembl as being incomplete,
  - gene/genomic region - ENSG00000198947.15
  - coding transcript - ENST00000357033.8
  - non-coding transcript - ENST00000383925.1
  - protein - ENSP00000354923.3

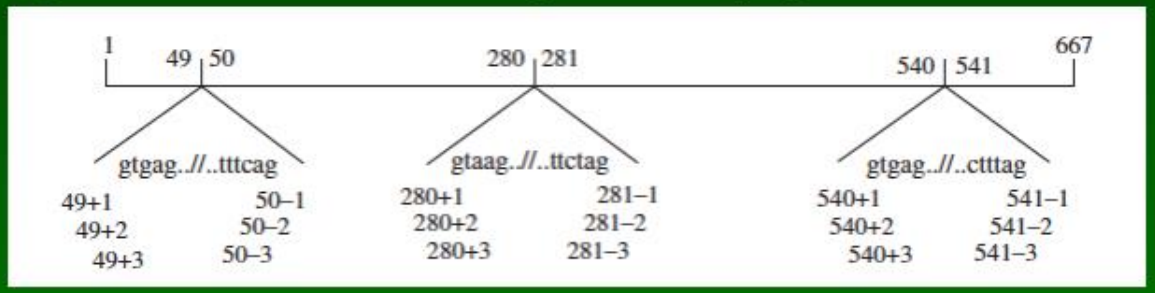
Где взять эти референсные последовательности?

Reference Sequence

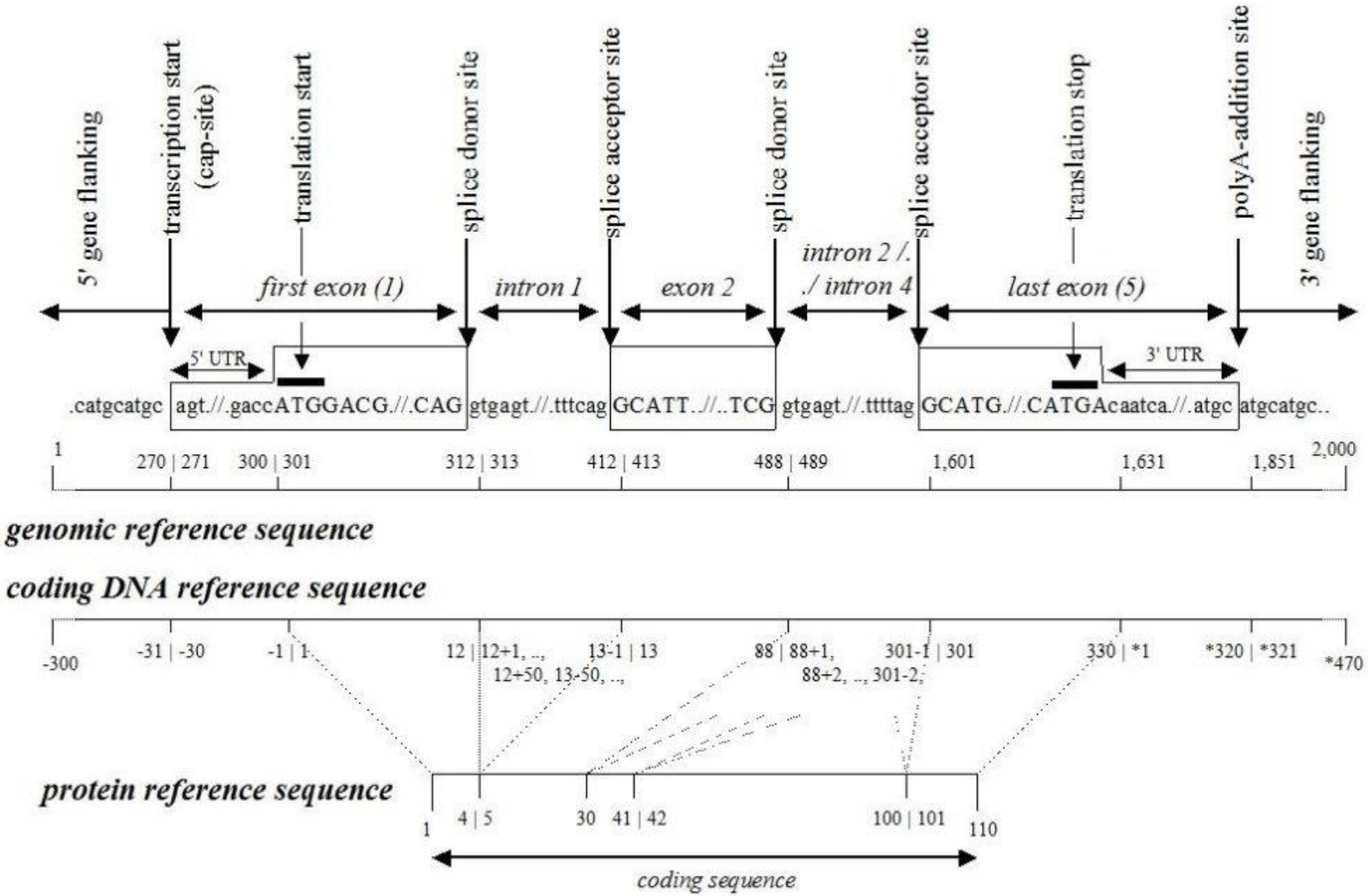
coding DNA reference sequence (c.)



non-coding DNA reference sequence (n.)



Структура гена



http://www.hgvs.org/mutnomen/RefSeq.jpg

Условные обозначения

Знак	**Значение**
g.	Геномная ДНК
c.	Кодирующая ДНК
m.	Митохондриальная ДНК
p.	Последовательность аминокислот в белке
*	Терминирующий кодон (стоп-кодон)
>	Замена в последовательности ***ДНК или РНК, но не протеина***
–	Ряд нуклеотидов

Замены в последовательности ДНК и нумерация

Локализация изменения	По последовательности геномной ДНК*	По последовательности кодирующей ДНК
5’-UTR	g.13793T>G	c.-5T>G
Кодирующий регион	g.59883A>G	c.1445A>G
Интрон	g.13874G>T	c.76+1G>T
	g.59871A>C	c.1434-2A>C
3' UTR и 3' – фланкирующий регион	g.108350G>C	c.*49G>C

Замены в кДНК и в аминокислотной последовательности

Изменение	По последовательности кодирующей ДНК	По последовательности аминокислот
Повреждение промотера или изменение в инициирующем ATG-кодоне	e.g. c.2T>G	p.0
		p.0?
Синонимичная замена	c.1311T>C	p.Ile437= (p.I437=)
Миссенс-замена	c.1445A>G	p.Asp482Gly (p.D482G)
Нонсенс-замена	c.1405C>T	p.Gln469* (p.Q469*, p.Gln469Ter, p.Gln469Term)
Замена стоп-кодона на смысловой кодон	c.3964T>C	p.*1322Argext*17
		p.*1322Argext*?
Замена двух (и более) нуклеотидов подряд	[c.1311T>G; c.1312C>G]	[p.Ile437Met; p.Leu438Val]

Делеции в последовательностях ДНК

g/c. <first deletion nucleotide №>_<last deletion nucleotide № >del(NNN)

- g.301del (g.301delA), g.301_304del (g.301_304delAGTG);
- c.330del (c.330delG), c.330_331del (c.330_331delGC);
- c.120_123+48del;
- c.124-12_129del;
- c.-11_-4del;
- c.*8_*21del.

ATGTTG**TG**CC -> ATGTTG_CC

c.7_8del (c.7_8delTG), а не c.5_6delTG

Дупликации в последовательностях ДНК

g/c. <first duplication nucleotide №>_<last duplication nucleotide №>dup(NNN)

- g.301dup (g.301dupA), g.301_304dup (g.301_304dupAGTG);
- c.330dup (c.330dupG), c.330_331dup (c.330_331dupGC);
- c.120_123+48dup;
- c.124-12_129dup;

ATGTTGTGCC -> ATGTTGTG**TG**CC

c.7_8dup (c.7_8dupTG), а не c.5_6dupTG

Инсерции в последовательностях ДНК

У вставившегося нуклеотида нет своего номера в референсной последовательности!

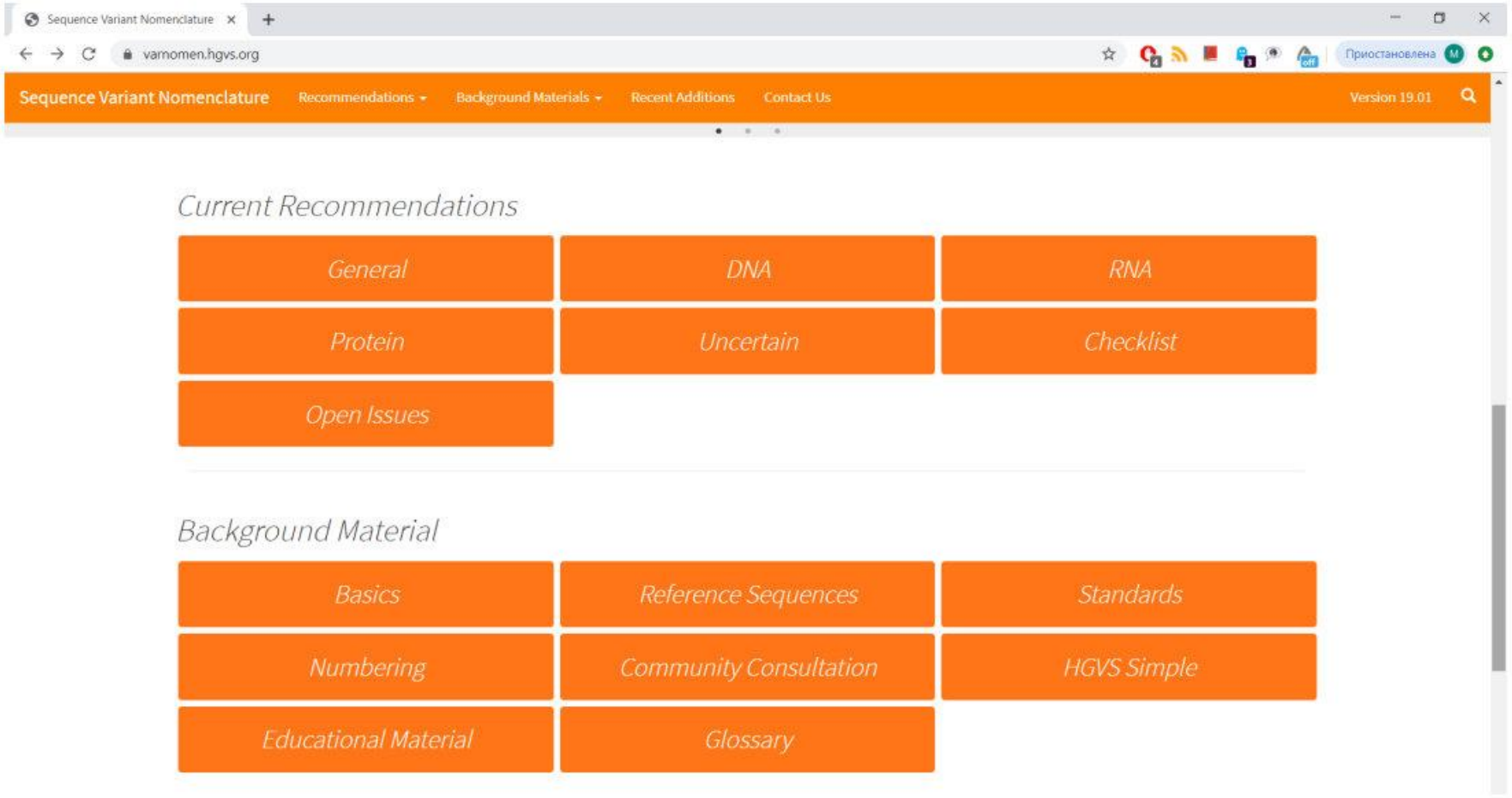
g/c. <previous nucleotide №>_<next nucleotide №>insNNN

Вставки, повторяющие предыдущие нуклеотиды референсной последовательности, всегда описываются как дупликации, а не как инсерции

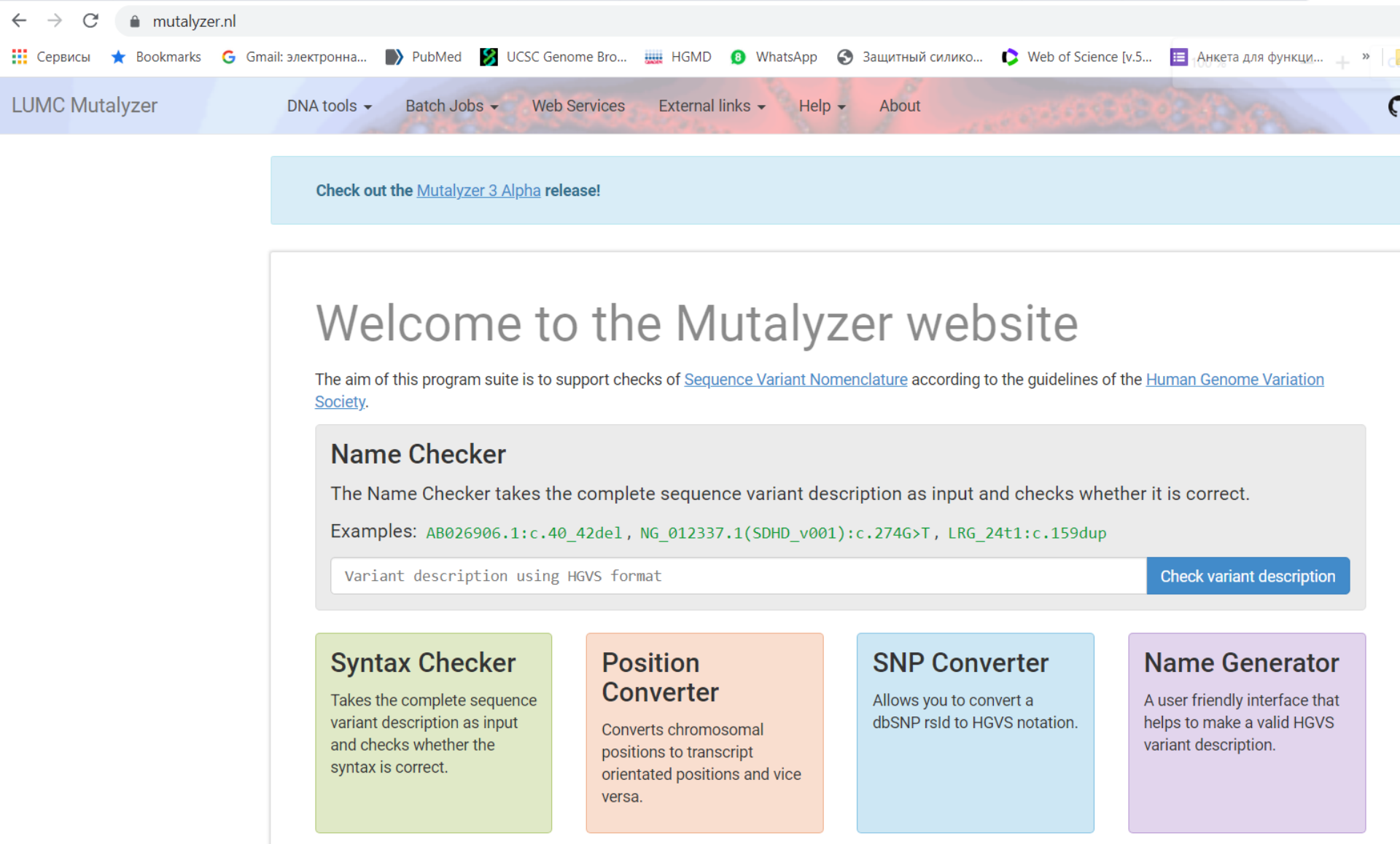
- g.451_452insT, c.51_52insT;
- g.451_452insGAGA, c.51_52insGAGA

Sequence Variant Nomenclature

Human Genome Variation Society



- https://varnomen.hgvs.org/


mutalyzer.nl
Сервисы
Bookmarks
Gmail: электронна...
PubMed
UCSC Genome Bro...
HGMD
WhatsApp
Защитный силико...
Web of Science [v.5...
Анкета для функци...
LUMC Mutalyzer
DNA tools
Batch Jobs
Web Services
External links
Help
About
Check out the Mutalyzer 3 Alpha release!
Welcome to the Mutalyzer website
The aim of this program suite is to support checks of Sequence Variant Nomenclature according to the guidelines of the Human Genome Variation Society.
Name Checker
The Name Checker takes the complete sequence variant description as input and checks whether it is correct.
Examples: AB026906.1:c.40_42del, NG_012337.1(SDHD_v001):c.274G>T, LRG_24t1:c.159dup
Variant description using HGVS format
Check variant description
Syntax Checker
Takes the complete sequence variant description as input and checks whether the syntax is correct.
Position Converter
Converts chromosomal positions to transcript orientated positions and vice versa.
SNP Converter
Allows you to convert a dbSNP rsId to HGVS notation.
Name Generator
A user friendly interface that helps to make a valid HGVS variant description.

VariantValidator

Accurate validation, mapping and formatting of sequence variants using HGVS nomenclature.

What We Do

We validate HGVS sequence variation descriptions, accurately mapping between transcript and genomic variants. We also automate conversion of genomic (VCF) sequence variation descriptions into the HGVS format and vice-versa.

VariantValidator auto-corrects your mistakes if it can and helps you correct your own if it can't. We provide a range of tools to meet your needs including batch processing, a VCF file converter and API access.

Powered By

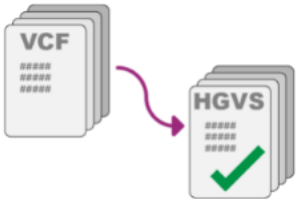
VariantValidator
version 1.0.4.dev120+g683823d
vv_hgvs
version 1.2.5.vv1
UTA
release uta_20180821
SeqRepo
release 2018-08-21

Validator

Validate your variant descriptions using HGVS nomenclature.

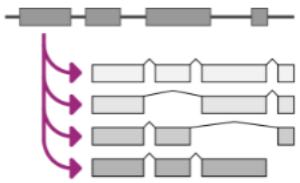
Batch Validator

Validate multiple variant descriptions at once.



VCF to HGVS

Convert VCF files to validated HGVS variants.



Gene to Transcript

Identify all transcripts from a gene symbol.