

NGS в медицинской иммунологии

Смирнов Антон Сергеевич

2025-10-19

О себе



Смирнов Антон Сергеевич

- МБФ, медицинская кибернетика'23
- ассистент кафедры биоинформатики МБФ Пироговского университета
- руководитель студенческого научного кружка каф.биоинформатики
- лаборант-исследователь лаборатории эпигенетики ожирения и диабета + специалист по информационным технологиям МГНЦ РАН



План лекции

1. Место NGS в медицине
2. Болезни иммунитета и методы их диагностики
3. Пайплайн обработки
 1. Форматы данных (sam, vcf)
 2. Аннотация vcf
4. Интерпретация результатов
 1. Критерии патогенности вариантов
 2. Алгоритмы по интерпретации результатов NGS и инструменты для анализа
5. HLA-типирование

Место NGS в медицине

Диагностика

1. Физикальная (опрос, осмотр, пальпация, перкуссия, аускультация)
2. Функциональная (ЭКГ, спирометрия, ЭхоКГ и т.д.)
3. Лабораторная (анализы крови, мочи, кала и т.д.)
4. Инструментальная (УЗИ, рентгендиагностика, МРТ и т.д.)

Диагноз

- Клинический - конкретное заболевание или синдром, т.е. фенотип
- Молекулярно-генетический - установлено изменение генотипа, приводящее к такому фенотипу

Диагностика с помощью NGS

Возможности

- Однонуклеотидные мутации (варианты генетической последовательности)
- Небольшие инделы (до 500 п.о.)
- Варианты числа копий вплоть до анеуплоидий

Ограничения

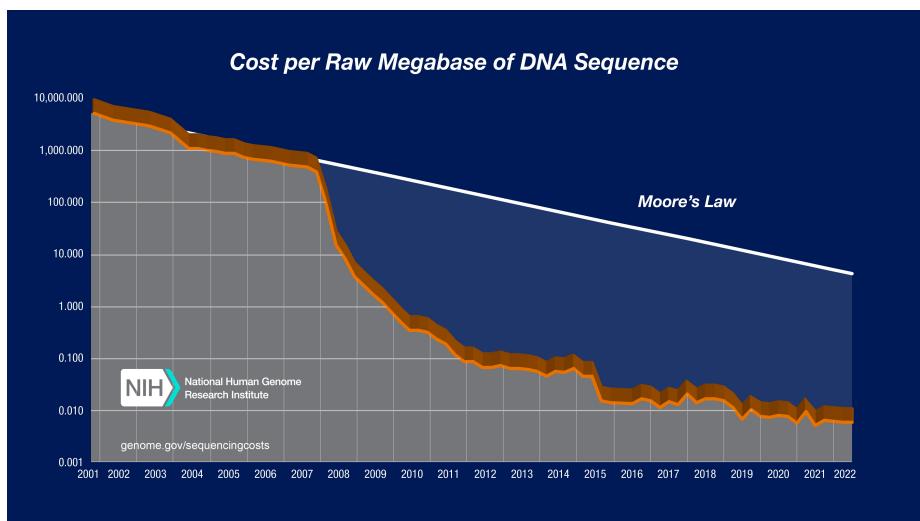
- Псевдогены
- Экспансии повторов
- Большие инделы и транслокации

! NB!

Виды секвенирования

- Таргетное (панельное) - секвенирование ограниченного числа локусов (панели)
- Полноэкзомное - секвенирование всех (почти всех) белок-кодирующих участков генома
- Полногеномное

Стоимость NGS

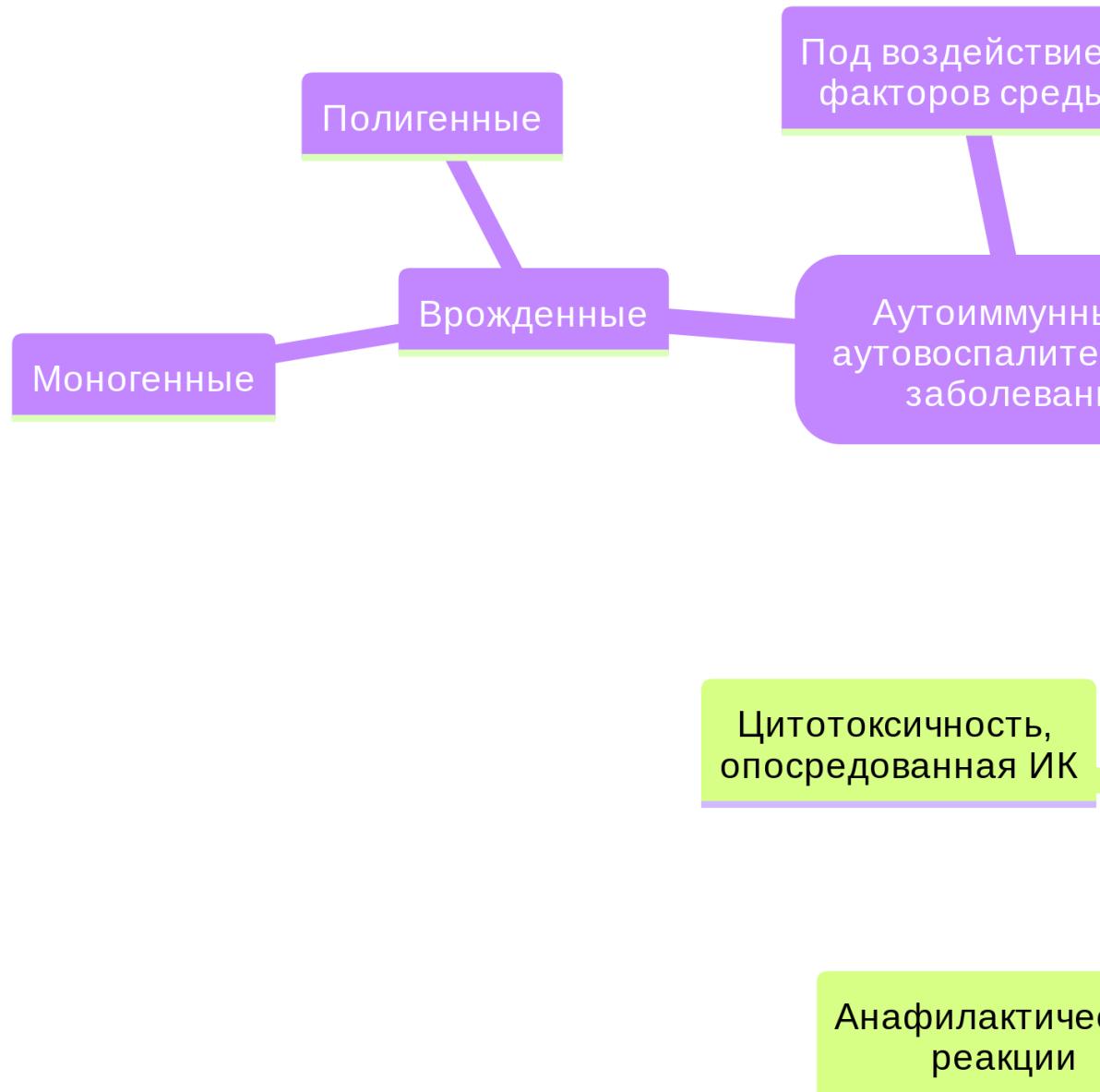


15.61.1	Секвенирование клинического экзома (6640 генов, кровь с ЭДТА) Срок готовности: 50 раб. дн	?	43900 ₽
15.62.1	Секвенирование полного экзома с анализом Срок готовности: 50 раб. дн	?	49900 ₽
15.65.1	Секвенирование полного генома с анализом данных (1 чел.) Срок готовности: 70 раб. дн	?	150000 ₽
15.3	Секвенирование панели генов «иммунодефициты» методом NGS Срок готовности: 50 раб. дн	?	43900 ₽
15.64.2	Биоинформационический анализ данных геномного секвенирования 1 человек Срок готовности: 25 раб. дн	?	22000 ₽

Болезни иммунитета и методы их диагностики

Основные группы заболеваний

< fa thumbs-up >



Первичные иммунодефициты

1. Комбинированные Т- и В-клеточные иммунодефициты;
2. Преимущественный дефицит антител (например, агаммаглобулинемия, селективный дефицит IgA);
3. Иммунодефициты с синдромальной патологией (например, атаксия-телеангиектазия);
4. Генетические нарушения иммунной регуляции (например, X-сцепленный IPEx синдром — иммунодисрегуляция, полиэндокринопатия, энтеропатия; X-сцепленный лимфопролиферативный синдром и др.);
5. Врожденные дефекты фагоцитов (числа, функций или того и другого, например, гипер-IgE-синдром);
6. Дефекты врожденного иммунитета и сигнальных компонентов (например, дефект IRAK4- киназы);
7. Аутовоспалительные синдромы (характеризуются рецидивирующими генерализованным воспалением в отсутствии явных инфекционных или аутоиммунных причин, например, семейная средиземноморская лихорадка);
8. Дефициты системы комплемента;
9. Фенокопии ПИД (классификация 2013 г.), это ненаследуемые ИД, не связанные с мутациями в генах «зародышевой линии». Ассоциированы с соматическими мутациями в генах, которые могут приводить к появлению, например, аутоантител к цитокинам, источнику цитокинов и в дальнейшем к развитию ИД.

Аутоиммунные и аутовоспалительные заболевания

	Аутовоспалительные	Аутоиммунные
Моногенные	CAPS;FMF;HIDS;TEAPS;PAPA;Синдром Блау	AMPs;IPEx;APS-1
Полигенные	Болезнь Кроны;Неспецифический язвенный колит;Псориатический артрит;Ювенильный идиопатический артрит	Целиакия;Болезнь Грейвса;Сахарный диабет 1 типа;Болезнь Аддисона

Промежуточные болезни: псориаз, анкилозирующий спондилит, реактивный артрит, болезнь Бехчета.

Эпигенетика и иммунодефициты

Хроматинопатии - группа наследственных заболеваний, затрагивающая структуру и функцию хроматина, что приводит к мультисистемным нарушениям.

Синдром Кабуки

Варианты в генах KMT2D (гистон-лизин N-метилтрансфераза), KDM6A (лизин-специфичная деметилаза 6A)



Синдром Сотоса

Варианты в гене NSD1 (транскрипционный ко-регулятор гистоновой метилтрансферазы)



NGS диагностика

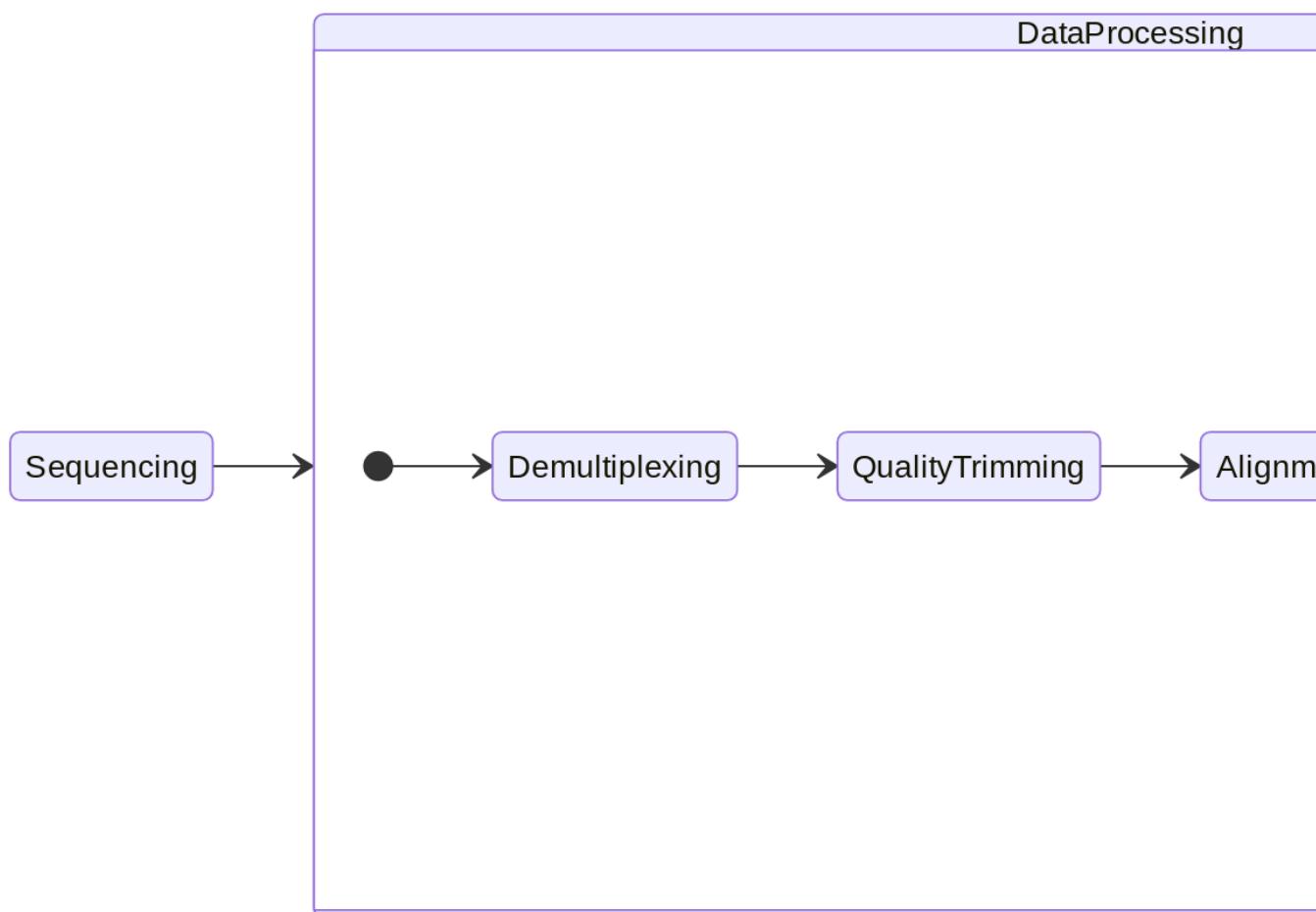
Статистика с Human Phenotype Ontology

24 августа 2025 г.

Термин	Ассоциации с болезнями	Ассоциации с генами
Immunodeficiency	246	223
Cellular	21	39
immunodeficiency		
Humoral	5	3
immunodeficiency		
Combined	44	41
immunodeficiency		

Пайплайн обработки

Общая схема



Формат sam

A

Coor	10	20	30	40
ref	12345678901234	5678901234567890123456789012345		
+r001/1	TTAGATAAAGGATA*CTG			
+r002	aaaAGATAA*GGATA			
+r003	gcctaAGCTAA			
+r004		ATAGCT.....TCAGC		
-r003		ttagctTAGGC		
-r001/2			CAGCGGCAT	

B

@HD VN:1.5 SO:coordinate		Header section		QUAL (read quality: * meaning such information is not available)	
@SQ SN:ref LN:45					
r001	99 ref 7 30 8M214M1D3M = 37 39 TTAGATAAAGGATACTG *				
r002	0 ref 9 30 3S6M1P14M *	0	0 AAAAGATAAAGGATA	*	
r003	0 ref 9 30 5S6M *	0	0 GCCTTAAGCTAA	* SA:Z:ref,29,-,6H5M,17,0;	
r004	0 ref 16 30 6M14N5M *	0	0 ATAGCTTCAGC	*	
r003	2064 ref 29 17 6H5M *	0	0 TAGGC	* SA:Z:ref,9,+,5S6M,30,1;	
r001	147 ref 37 30 9M = 7 -39 CAGCGGCAT			*	NM:i:1

Below the table, a legend explains the SAM fields:

- QNAME** (query template name, aka. read ID) indicates the reference sequence name, e.g. chromosome / transcript id.
- FLAG** indicates alignment information about the read, e.g. paired, aligned, etc.
- RNAME** (reference sequence name, e.g. chromosome)
- POS** (1-based position)
- MAPQ** (mapping quality)
- CIGAR** (summary of alignment, e.g. insertion, deletion)
- RNEXT** (reference sequence name of the primary alignment fragment of the next read, for paired-end sequencing, NEXT read is the paired read; corresponding to the RNAME column)
- PNEXT** (Position of the primary alignment of the next read in the template; corresponding to the POS column)
- TLEN** (the number of bases covered by the reads from the same fragment. In this particular case, it's 45 + 7 = 39, corresponding to the POS column)
- SEQ** (read sequence)
- Optional fields in the format of TAG:TYPE:VALUE**

Подробнее

Спецификация

BAM и CRAM

BAM

1. SAM файл, поля которого сжаты с использованием BGZF
2. Сжатие происходит за счет особой кодировки и компоновки полей
3. Наиболее универсальный и популярный формат хранения выравниваний

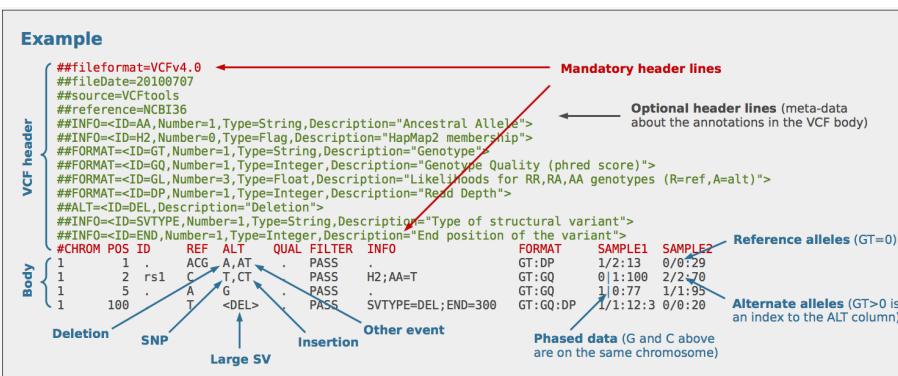
CRAM

1. SAM файл, поля которого сжаты с использованием gzip/bzip2
2. Хранит только различия образца от референсного генома
3. При компрессии/декомпрессии необходим референсный геном
4. По оценкам на 40-50% эффективнее BAM в сжатии без потерь

Совместимость

Все три формата совместимы между собой и поддерживаются основным ПО для медицинской генетики.

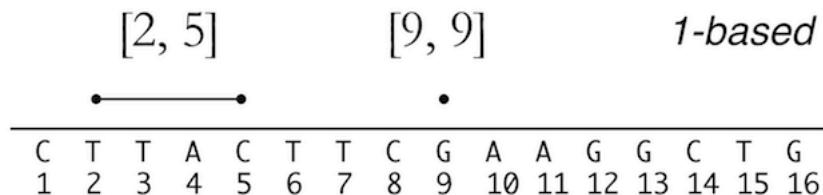
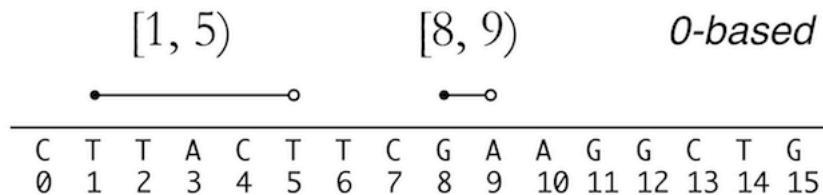
VCF



Подробнее

1. VCF также имеет свою бинарную, сжатую версию - BCF
2. Спецификация

Дьявол координатных систем



Format/library	Type
BED	0-based
GTF	1-based
GFF	1-based
SAM	1-based
BAM	0-based
VCF	1-based
BCF	0-based
Wiggle	1-based
GenomicRanges	1-based

Источник

Нормализация вариантов

1. Приведение к биалльной записи
2. Парсимония - представление вариантов минимальным количеством нуклеотидов без уменьшения длины какого-либо аллеля до 0.
3. Левостороннее выравнивание - запись координаты варианта как можно "левее" (т.е с наименьшим положением на хромосоме).

Фильтрация вариантов (техническая)

⚠ NB!

1. Учитывайте тип образца, тип и технологию секвенирования, версию референса
1. FILTER == PASS
2. Фильтрация по VAF - для герминальных вариантов окрестность 0, 0.5, 1. Промежуточные значения может означать контаминацию
3. Фильтрация по глубине - 1 или 5 процентиль, меньше 10. Малопредставленные варианты могут означать артефакт секвенирования или мозаичный вариант.
4. Фильтрация по чёрным спискам

Аннотация вариантов

Аннотация вариантов Процесс присоединения контекста (клинический, функциональный, популяционной и пр.) к варианту

1. Генетический контекст (ген, транскрипт, положение в гене, классификация варианта)
2. Функциональный контекст (функция гена, задействованные сигнальные и метаболические пути)
3. Популяционный контекст (популяционные частоты аллеля в различных популяциях)
4. Клинический контекст (известные заболевания, синдромы, ассоциированные с геном, вариантом)
5. Вычислительный контекст (предсказание эффекта варианта)

Программы-аннотаторы

1. VEP
2. snpEff
3. ANNOVAR
4. OpenCravat

ℹ NB!

Программы требуют большого места на дисковом хранилище для кэша и предохраненных баз данных

Источники данных

Генетический контекст

- NCBI
 - RefSeq
 - MANE Select
- Ensembl

Функциональный контекст

- GO
- KEGG
- Reactome

Популяционный контекст

- GnomAD
- 1000G
- RuSeq
- RuExac
- База частот ЦСП ФМБА

Клинический контекст

- OMIM
- HGMD

- LOVD
- ClinVar

Вычислительный контекст

- Предикторы, основанные на правилах (SIFT, PolyPhen-2)
- Метапредикторы (Meta-EA, CADD)
- Предикторы, основанные на машинном обучении (SAV-Pred, SpliceAI, AlphaMissense)

Клиническая интерпретация вариантов

1. Рекомендации ACMG
2. Российские рекомендации

Категоризация вариантов

Критерии категоризации

Полезный инструмент

Типы правил

1. Очень сильные - PVS (только P)
2. Сильные - PS (P и B)
3. Средние - PM (только P)
4. Поддерживающие - PP (P и B)
5. Автономные - SA (stand-alone, только B)

Категории правил

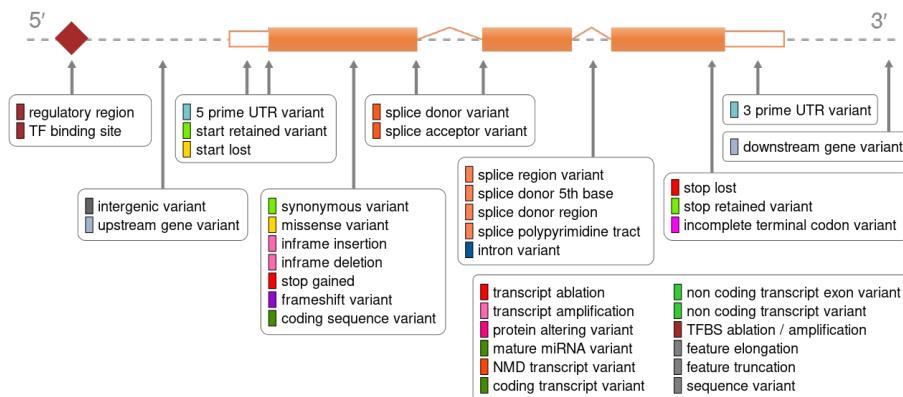
1. Популяционные данные
2. Вычислительные и предсказательные данные
3. Функциональные данные
4. Данные по сегрегации
5. Данные по *de novo*
6. Аллергические данные
7. Данные из других источников
8. Прочее

Правила по категоризации

Классификация	Комбинация критериев
Pathogenic	<ul style="list-style-type: none"> • PVS1 + (≥ 1 PS OR ≥ 2 PM OR 1 PM + 1 PP OR ≥ 2 PP) • ≥ 2 PS • PS1-4 + (≥ 3 PM OR 2 PM + ≥ 2 PP OR 1 PM + ≥ 4 PP)
Likely Pathogenic	<ul style="list-style-type: none"> • PVS1 + 1 PM • PS1-4 + $1-2$ PM • PS1-4 + ≥ 2 PP • ≥ 3 PM • 2 PM + ≥ 2 PP • 1 PM + ≥ 4 PP
Benign	<ul style="list-style-type: none"> • BA1 (Stand-alone) • ≥ 2 BS1-4
Likely Benign	<ul style="list-style-type: none"> • BS1-4 + BP1-7 • ≥ 2 BP1-7
Uncertain Significance	<ul style="list-style-type: none"> • Other criteria not met • Conflicting benign/pathogenic criteria

Фильтрация вариантов (биологическая)

Генетический контекст



Популяционный контекст

Базы данных: gnomAD, 1000G и пр.

- для аутосомно-доминантных заболеваний частота аллеля не должна превышать 0,01%,
- для аутосомно-рецессивных заболеваний — 0,5%,
- для доминантных X-сцеплённых — 0,01%,
- для рецессивных X-сцеплённых — 0,3%.

Собственная база данных: если в вашей собственной БД вариант встречается слишком часто, скорее всего он не является причинным.

Клинический контекст

1. Составление виртуальных панелей, наиболее соответствующих фенотипу пациента (HPO)
2. Поиск вариантов с известной клинической значимостью (HGMD, ClinVar, LOVD).

! NB!

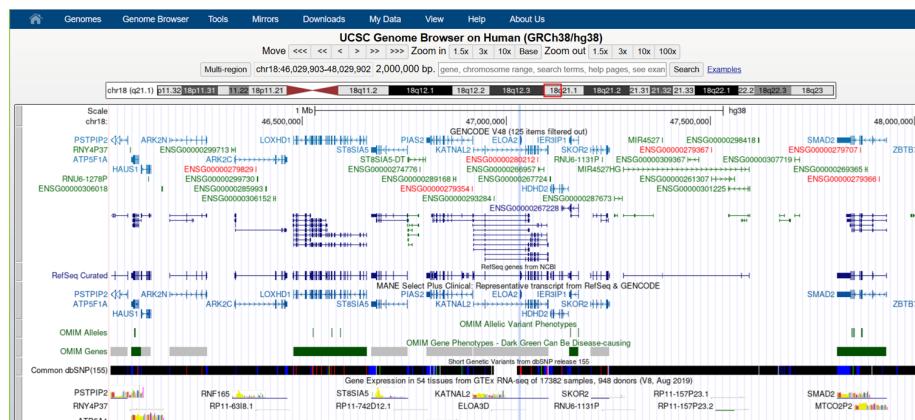
Всегда перепроверяйте уже опубликованную информацию и перечитывайте статьи! Опубликовано может быть что угодно, а ответственность на пациенте на вас.

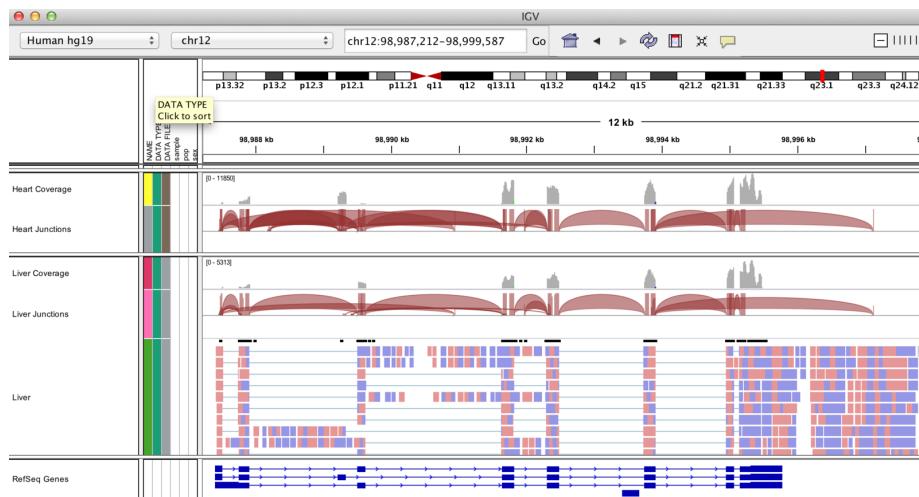
Функциональный и вычислительный контекст

1. Функциональный контекст используется только для обоснования эффектов найденных вариантов
2. Вычислительные данные обладают на данный момент слабой доказательной силой и используются только тогда, когда надо "натянуть" нужный вариант на нужную категорию.

UCSC Genome Browser

ucsc mirror



igv**Заключение по результатам анализа****Обязательные пункты**

1. Патогенные варианты
2. Вероятно патогенные варианты
3. Варианты с неизвестной клинической значимостью

Необязательные пункты

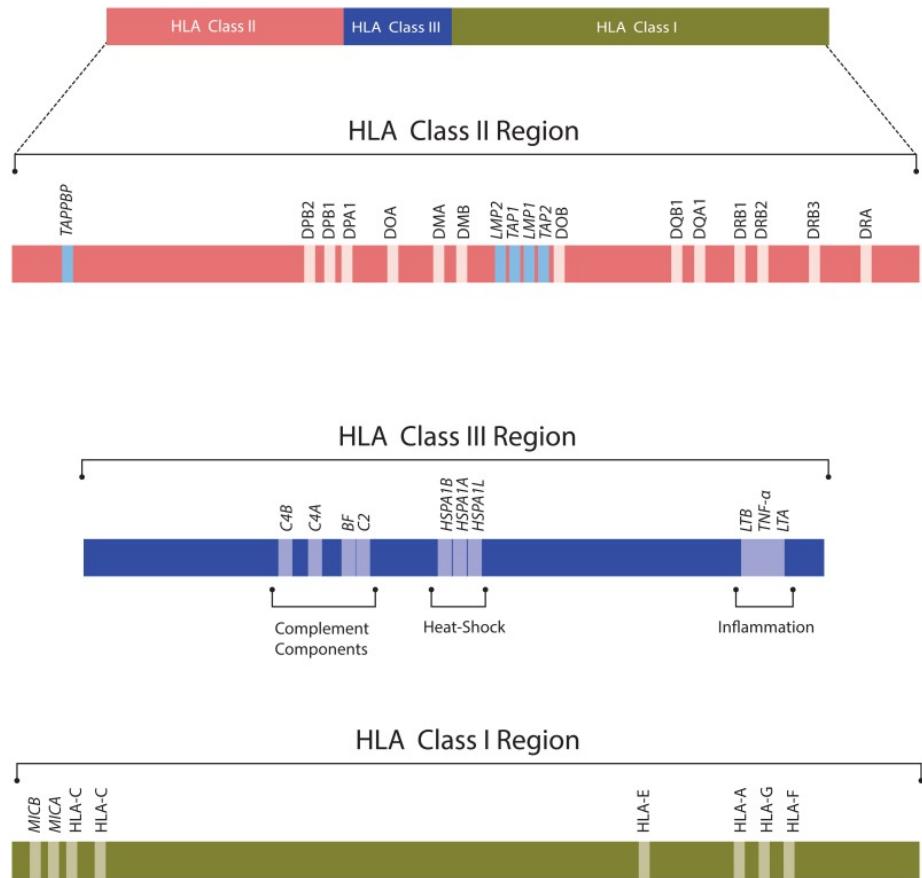
1. Вторичные находки
2. Структурные варианты и варианты числа копий
3. Экспансии повторов

HLA-типирование

процесс определения аллелей главного комплекса гистосовместимости
(для человека)

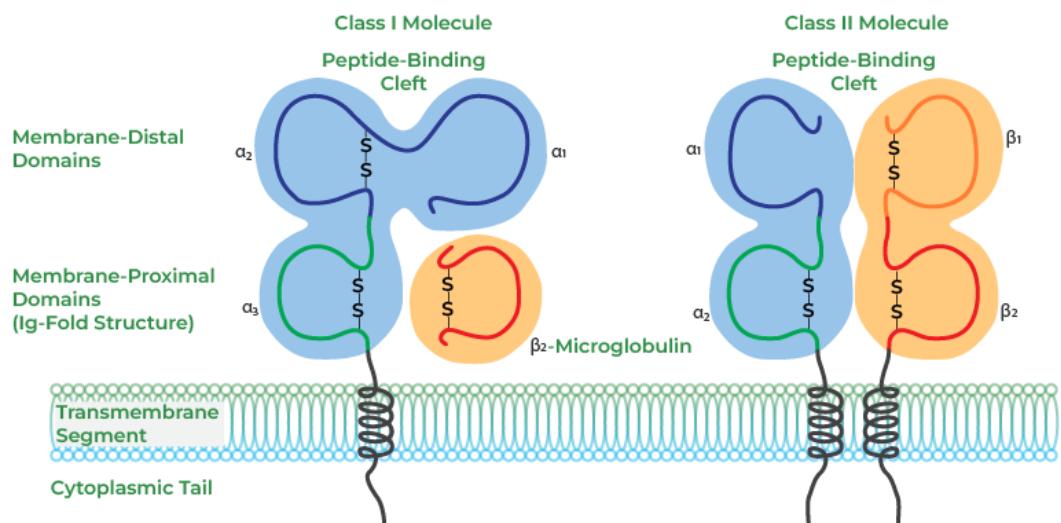
Главный комплекс гистосовместимости

комплекс тесно связанных генетических локусов, расположенных на коротком плече 6-ой хромосомы (бр), а также их белковых продуктов, отвечающих за развитие иммунного ответа и синтез трансплантационных антигенов.



Строение молекул МНС

MHC Class I vs MHC Class II



Источник

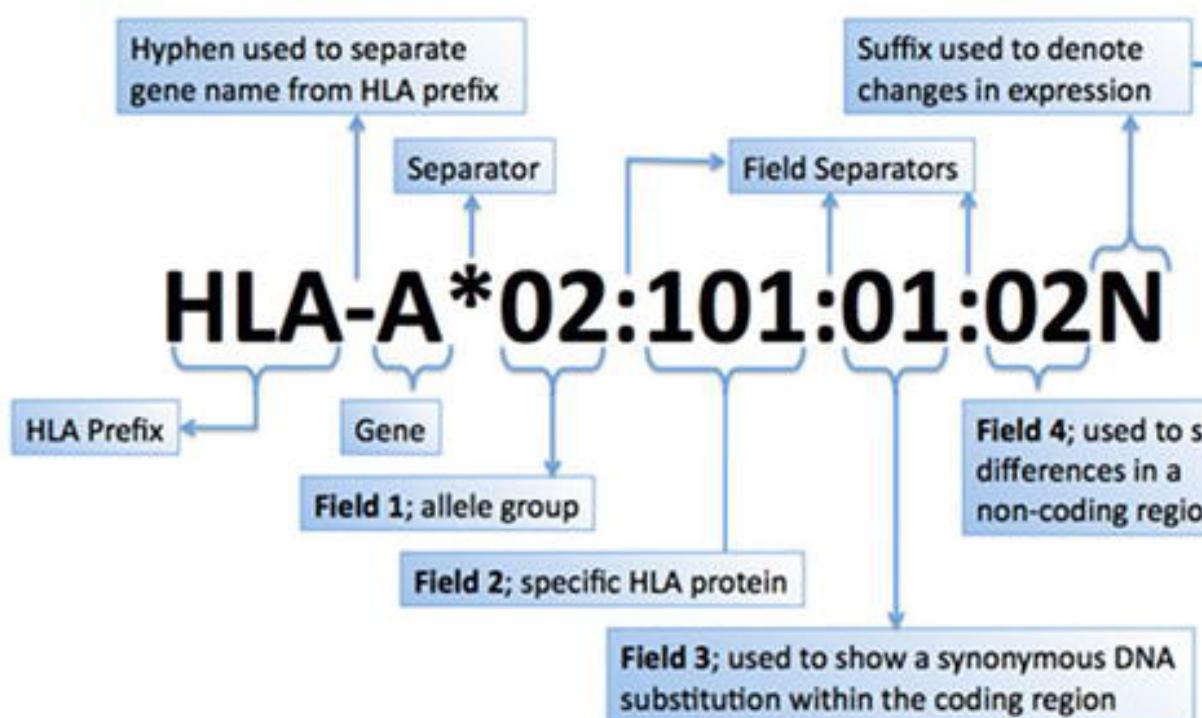
Функция молекул МНС

- Обеспечение процессинга и презентации антигенных пептидов – индукторов и мишени иммунного ответа;
- Обеспечение взаимодействия клеток
- Распознавание собственных, измененных собственных и чужеродных клеток; запуск и реализация иммунного ответа против носителей генетической информации
- Поддержание иммунологической толерантности (в том числе во время беременности);
- Участие в селекции Т-лимфоцитов;
- Создание генетического разнообразия и обеспечение выживаемости вида.

Свойства локуса

- Высокая полиморфность
- Высокая степень гетерозиготности
- Кодоминантное наследование

Номенклатура IGMT



© SGE Marsh

Cui bono?

1. **Трансплантология**
2. Диагностика некоторых аутоиммунных заболеваний (болезнь Бехтерева)

Федеральный регистр доноров костного мозга

Регистр доноров костного мозга

В твоем решении целая жизнь!

Регистр доноров костного мозга РНИМУ им. Н.И. Пирогова, созданный в 2019 году, является частью Федерального регистра доноров костного мозга. В 2019 в Регистр поступали образцы крови в основном студентов Университета. С тех пор к нам присоединилось огромное количество добровольцев – потенциальных доноров из различных организаций Москвы. РНИМУ им. Н.И. Пирогова проводит полный цикл работ в рамках Федерального регистра: это работа рекрутингового центра, работа типирующей лаборатории, центр заготовки костного мозга и центр трансплантации.

Информация о наделении РНИМУ им. Н.И. Пирогова полномочиями Рекрутингового центра Федерального регистра в [Единой системе нормативно-правовой информации](#).

11 083

ПОТЕНЦИАЛЬНЫХ ДОНОРА
В РЕГИСТРЕ ДОНОРОВ
КОСТНОГО МОЗГА

234

ЗАПРОСА НА
ПРЕДВАРИТЕЛЬНУЮ
АКТИВАЦИЮ ДОНОРА

46

В ПРОЦЕССЕ ОБСЛЕДОВАНИЯ

85

СДАЛИ КОСТНЫЙ МОЗГ
И ГЕМОПОЭТИЧЕСКИЕ
СТВОЛОВЫЕ КЛЕТКИ



Методы типирования

1. Серотипирование (антителами)
2. ПЦР-SSO, ПЦР-SSOP
3. Сэнгер
4. NGS
5. Длинные прочтения

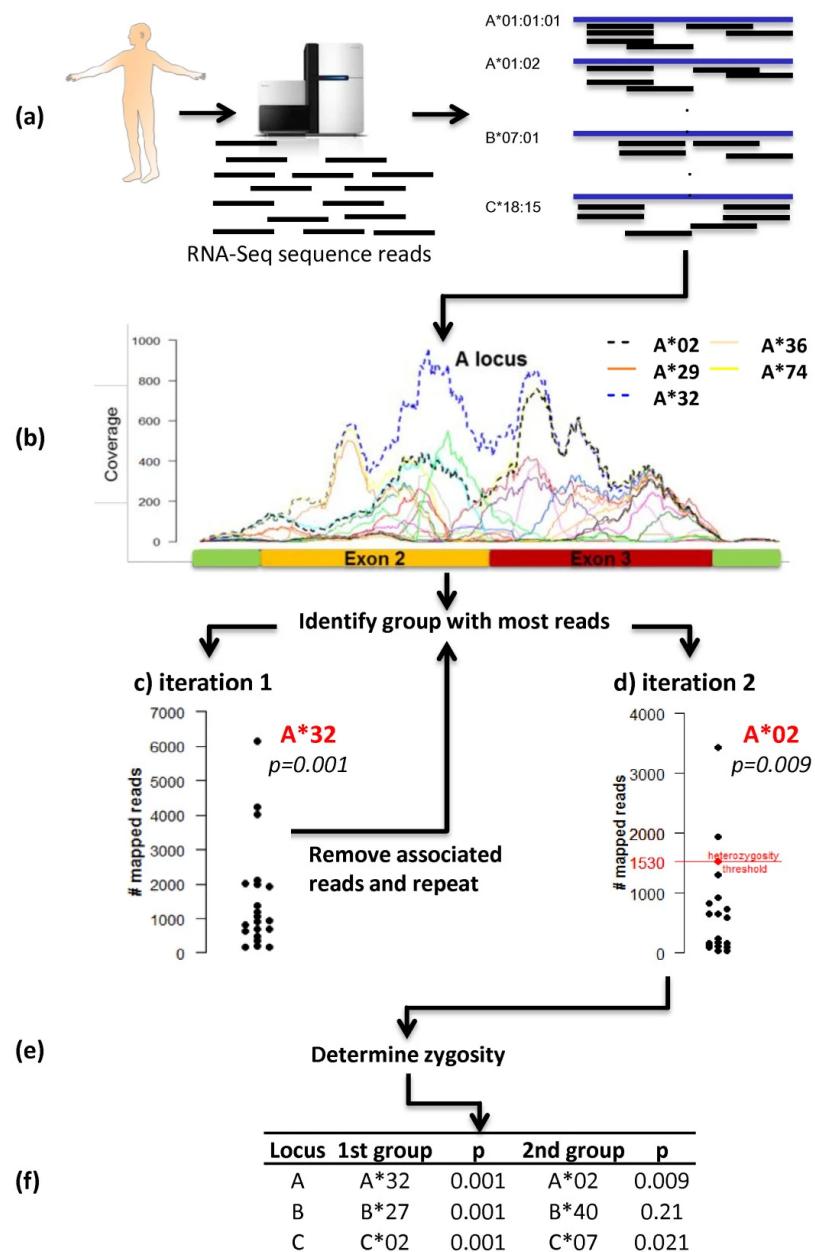
Особенности для биоинформатики

1. Малая степень отличий между аллелями (1-2 нуклеотида)
2. Высокая степень полиморфности

Инструменты

1. seq2HLA
2. Optitype
3. HLA-LA
4. HISAT-genotype
5. ...

seq2HLA



Источник

Optitype

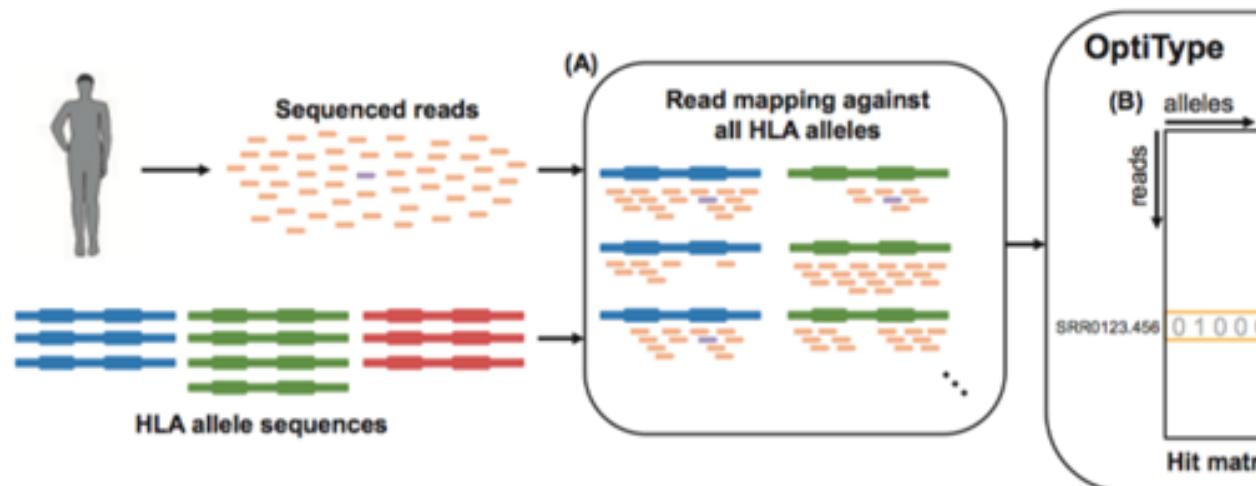
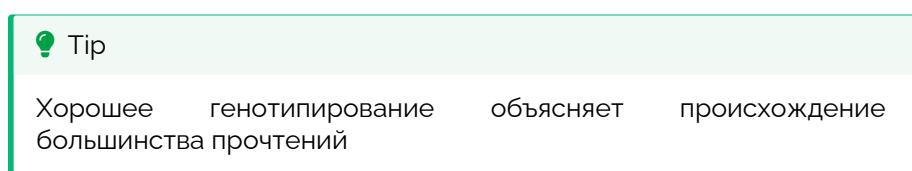
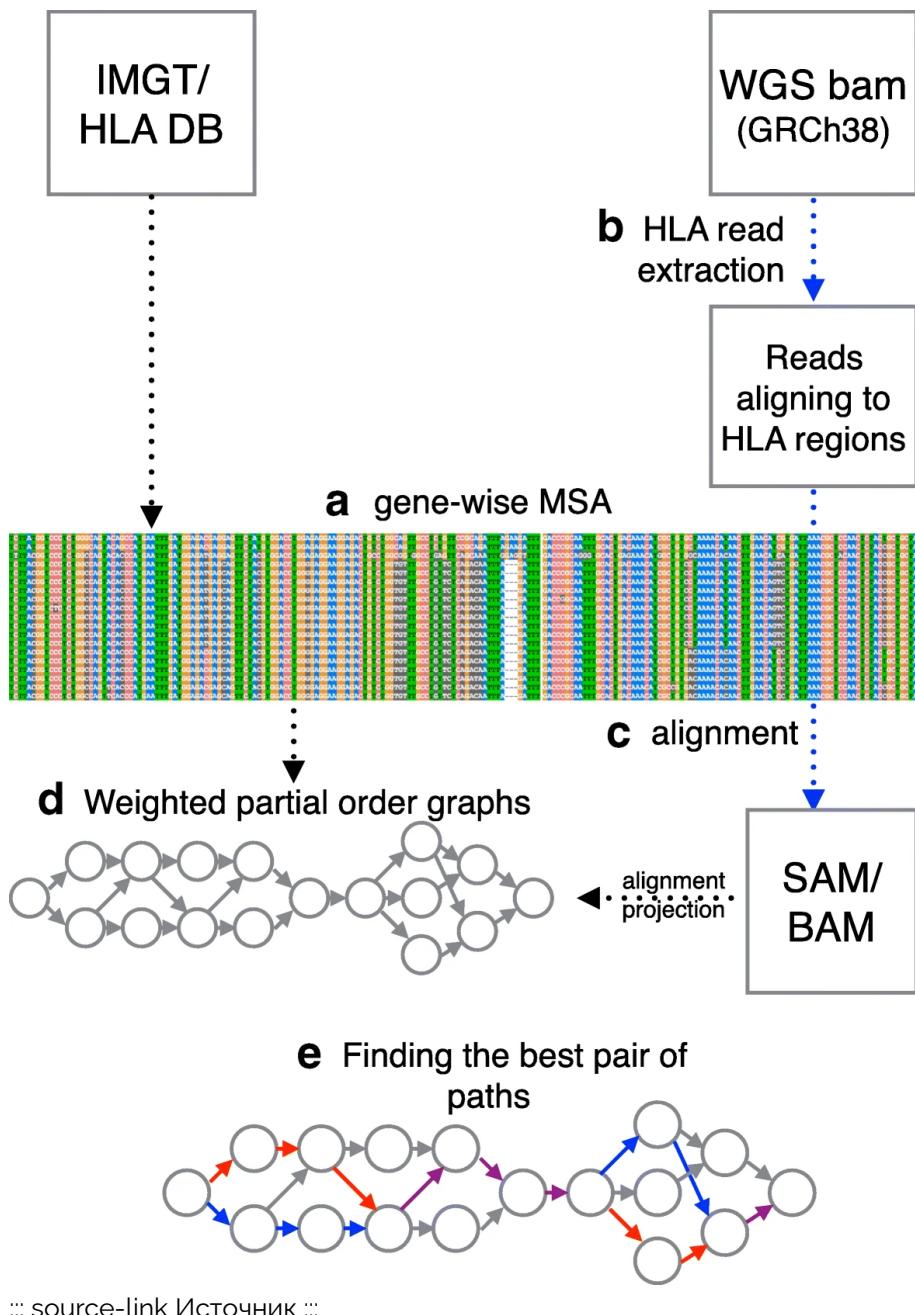


Fig. 1. OptiType's four-digit HLA typing pipeline. Reference libraries for genomic and CDS are generated by extracting flanking intronic regions from the closest relative HLA-I allele. For genomic sequences, flanking intronic regions are also extracted. If some of these regions are missing, they are reconstructed from the closest relative HLA-I allele. NGS reads are mapped against the so-called reference library. From the mapping result a binary hit matrix $C^{R \times A}$ is constructed for all reads $r \in R$ mapping to at least one allele $a \in A$. If a read r could be mapped to allele a ; otherwise, $C_{r,a} = 0$ (B). Based on this hit matrix, an ILP is formulated that optimizes the selection of alleles by selecting up to two alleles (columns of the hit matrix) for each HLA-I locus (C). The selected alleles represent the final HLA type.

Источник

Графовые методы



... source-link Источник ...

Итоги

1. Иммуноинформатика имеет прикладное значение в медицинской генетике
2. Интерпретация результатов NGS осуществляется по специальных рекомендациях
3. HLA-типирование требует особых алгоритмов выравнивания.

