

ЦИФРОВОЙ ФЕНОТИП

Проект по идентификации человека: Parseq Lab.

Отчет по летней практике 2016.

Анна Белоусова, anna.belousova@skolkovotech.ru

ПРОЕКТ ЦИФРОВОЙ ФЕНОТИП

1. Определение фенотипа в криминалистике
2. Текущее положение дел
3. Модели определения фенотипа
4. Идея проекта
5. Основной результат
6. Выводы
7. Планы на будущее

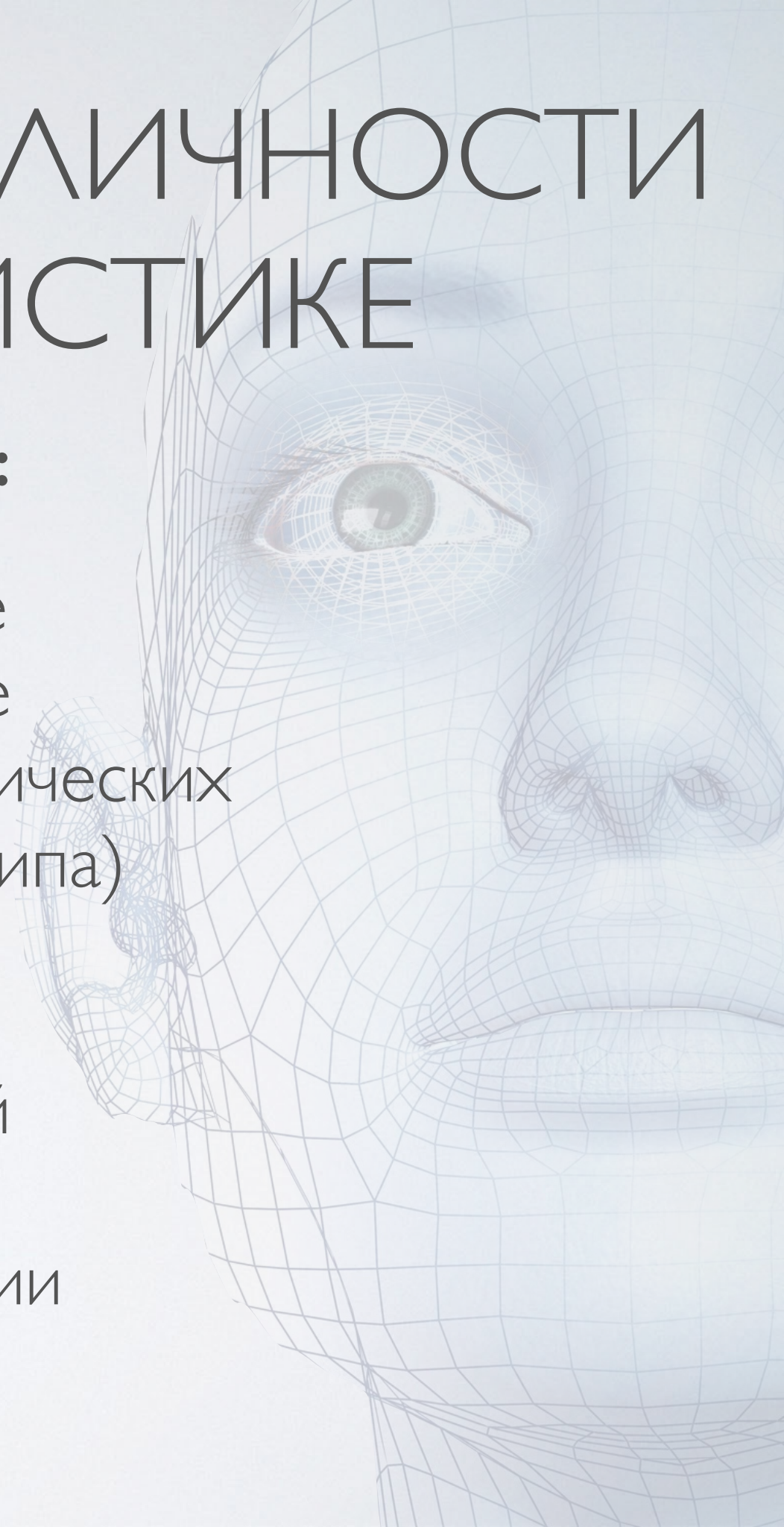
ИДЕНТИФИКАЦИЯ ЛИЧНОСТИ В КРИМИНАЛИСТИКЕ

Цели:

- подтверждение следственной гипотезы
- идентификация пострадавших при бедствии / террористическом акте
- идентификация жертвы насилия
- поиск пропавших без вести
- создание криминалистической базы данных

Задачи:

- объективное определение характеристических черт (фенотипа)
- разработка однозначной системы классификации фенотипа



ОПРЕДЕЛЕНИЕ ВНЕШНИХ ФИЗИЧЕСКИХ ХАРАКТЕРИСТИК

Основная задача: сужение
круга объектов поиска на
начальных этапах анализа.

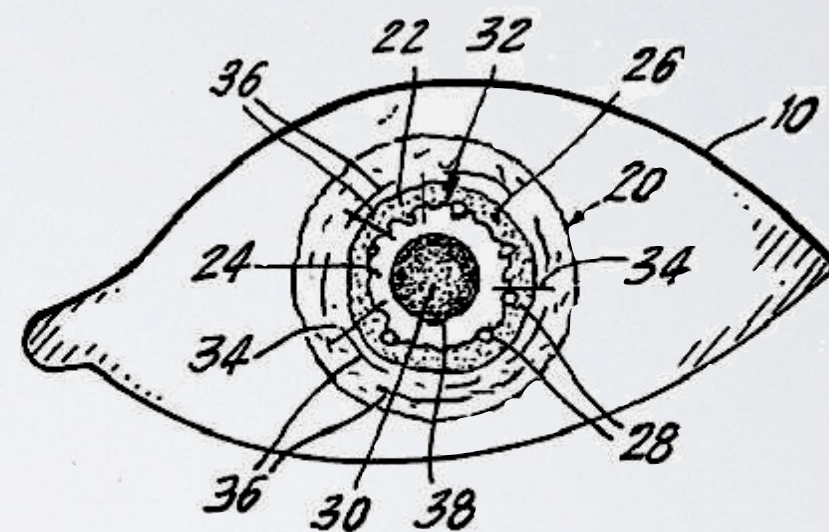


Схема из патента US5291560,
защищающего технологию сканирования радужки,
1991 год.

ОСОБЕННОСТИ ОБРАЗЦОВ ДНК

- следовые количества
- быстрая деградация
- смеси

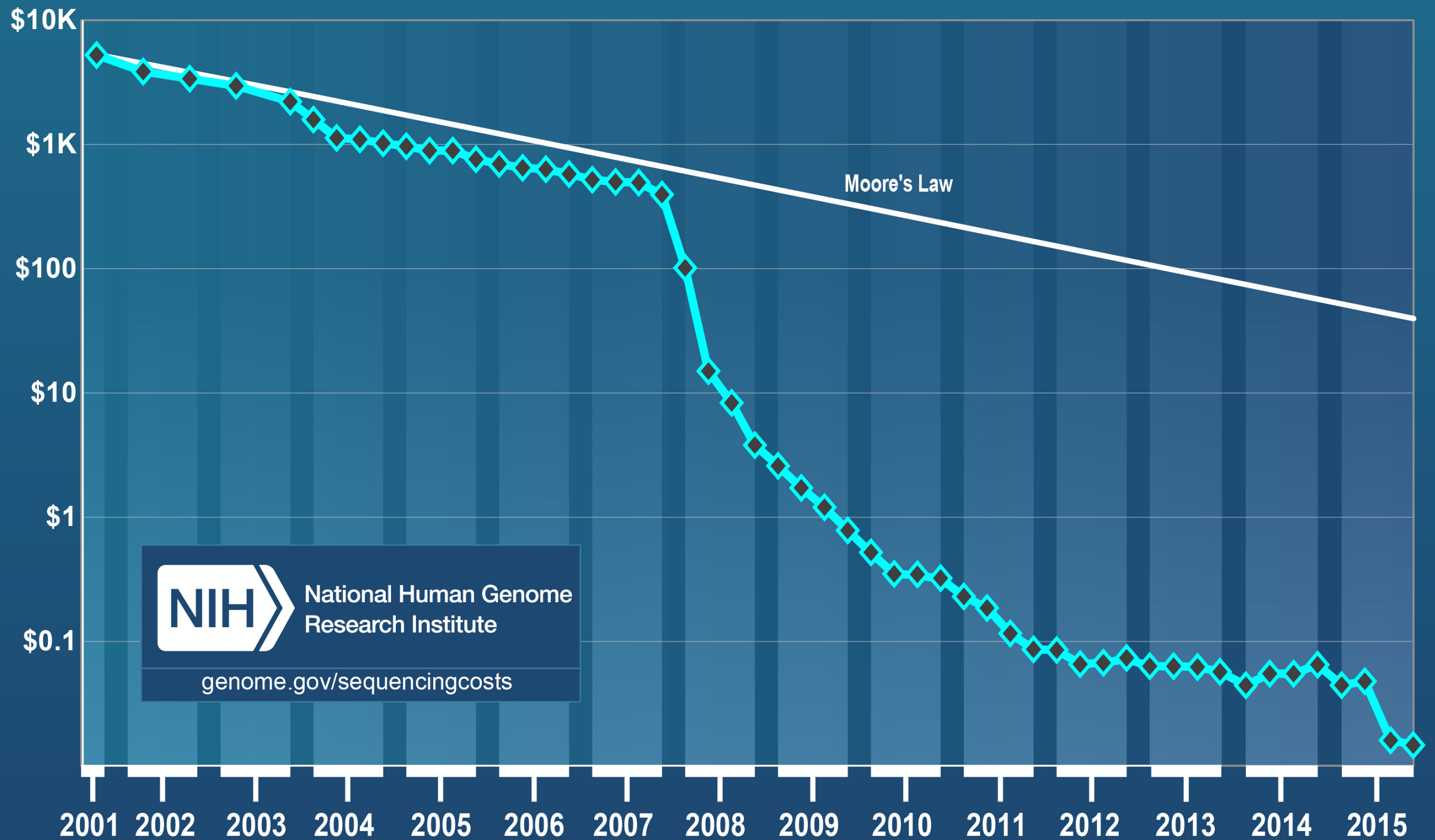
БИОТЕХНОЛОГИЧЕСКИЕ МЕТОДЫ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ

- анализ полиморфизма длин рестрикционных фрагментов (ПДРФ, RLFP)
- методы на основе ПЦР для анализа повторов
- аутосомальные STR (short tandem repeat) маркеры
- X- и Y-STR — повторы на половых хромосомах
- анализ локусов однонуклеотидных полиморфизмов (SNPs)
- анализ митохондриальной ДНК
- данные экспрессии микроРНК из биологических жидкостей

HUMAN GENOME PROJECT

- 1990 - 2003, 40 inst, 8-9x coverage
- + db SNP

Cost per Raw Megabase of DNA Sequence



НОВЫЕ ТЕХНОЛОГИИ СЕКВЕНИРОВАНИЯ

- NGS
- MPS
- (возможности — скорость, выход, покрытие, снижение числа ошибок, стоимость)

МЕТОДЫ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ

- ~~анализ полиморфизма длин рестрикционных фрагментов (ПДРФ, RFLP)~~
- ~~методы на основе ПЦР для анализа повторов~~
- аутоматические STR (short tandem repeat) маркеры
- X- и Y-STR — повторы на половых хромосомах
- анализ локусов однонуклеотидных полиморфизмов (SNPs)
- анализ митохондриальной ДНК
- данные экспрессии микроРНК из биологических жидкостей

МЕТОДЫ АНАЛИЗА STR И Y-STR

- Short tandem repeats — подробнее о повторах

АНАЛИЗ МТ-ДНК

АНАЛИЗ МИКРОРНК

ИСТОРИЯ: ПОДДЕЛКА ОБРАЗЦА ДНК

- Нужно ли вставлять историю, как человек с помощью донорской крови и антикоагулянтов обеспечил себе алиби в деле по изнасилованию?

ВИДЫ АНАЛИЗА SNP

- SNP для идентификации личности: наиболее полиморфные локусы внутри популяции
- SNP для анализа географического происхождения / этнической принадлежности
- SNP для анализа родства: ди / мультиаллельные снипы
- SNP для анализа фенотипа: пигментация глаз, волос, кожи
- SNP для анализа физиологии: связанные с заболеваниями

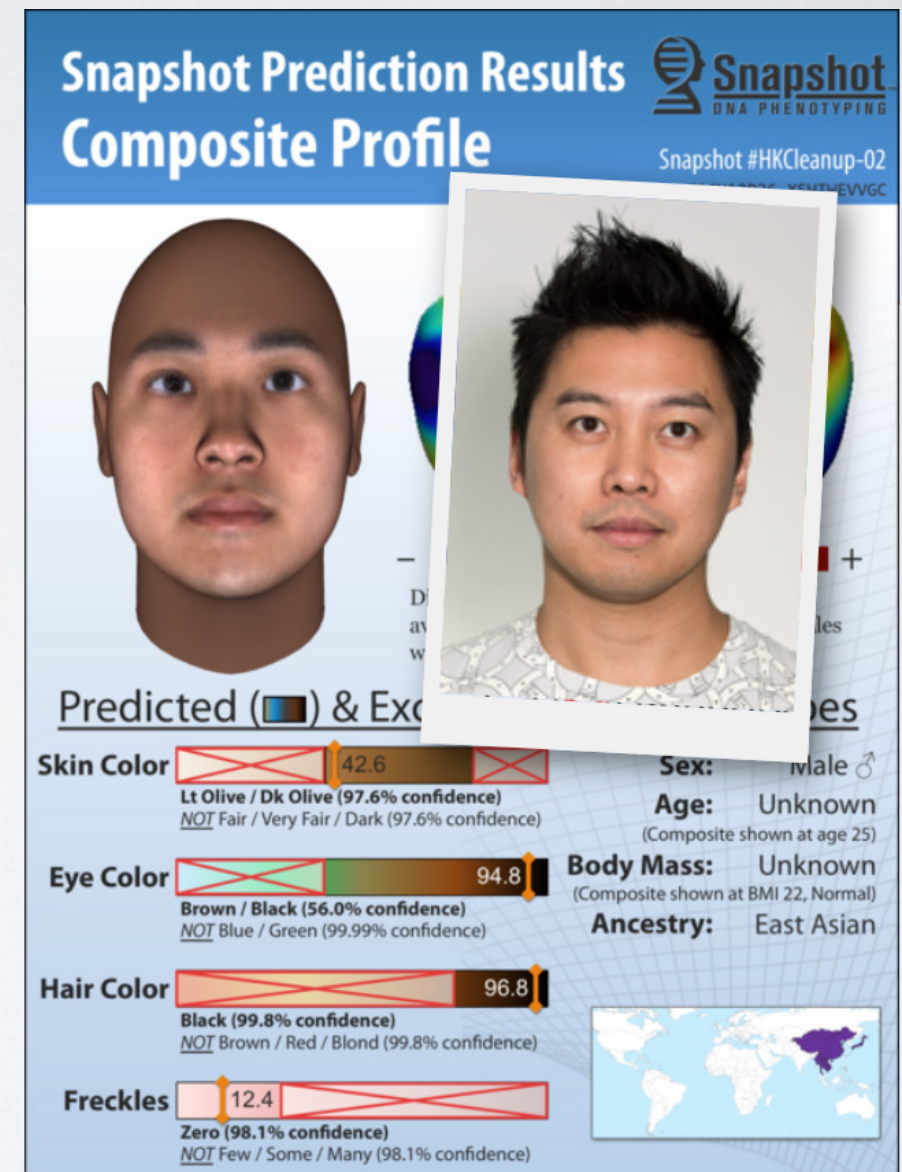
КАК SNP ОПРЕДЕЛЯЕТ ФЕНОТИП

МОДЕЛИ ОПРЕДЕЛЕНИЯ ФЕНОТИПА

- Decision tree
- Linear regression model

СОВРЕМЕННЫЕ ИНСТРУМЕНТЫ ОПРЕДЕЛЕНИЯ ФЕНОТИПА ПО SNP

- IrisPlex, HIrisPlex, HIrisPlex-S
- Snipper
- Parabon Snapshot



ПОДРОБНЕЕ О КАЖДОМ ИЗ ИНСТРУМЕНТОВ

- база данных
- какая модель используется
- есть ли он-лайн версия (какие in, out данные)
- ссылки на публикации

ЦИФРОВОЕ ФЕНОТИПИРОВАНИЕ: ПОСЛЕДОВАТЕЛЬНОСТЬ ДЕЙСТВИЙ

1. Чтение последовательности ДНК,
2. Индексирование, выравнивание и картирование чтений,
3. Выделение областей в чтениях и получение файлов с заменами в локусах,
4. Расчет наиболее вероятного фенотипа в соответствии с заменами на основании выбранной математической модели,
5. Получение сведений о фенотипе из опроса,
6. Чтение и форматирование сведений,
7. Получение изображений фенотипических черт,
8. Обработка изображений, распознавание фенотипа,
9. Формализация фенотипа,
10. Валидация результатов предсказания на основании сведений из опроса / изображений.

Работа с последовательностями ДНК

Чтение ДНК
образцов

Индексирование, выравнивание,
картирование чтений

Отбор локусов,
поиск SNP

расчет наиболее вероятного фенотипа
в соответствии с генотипом

Работа с изображениями

Анализ изображения,

Коррекция изображения,
распознавание области для анализа

Создание
фотографии
для анализа

Предсказание фенотипа

предсказание
фенотипа

формализация
фенотипа

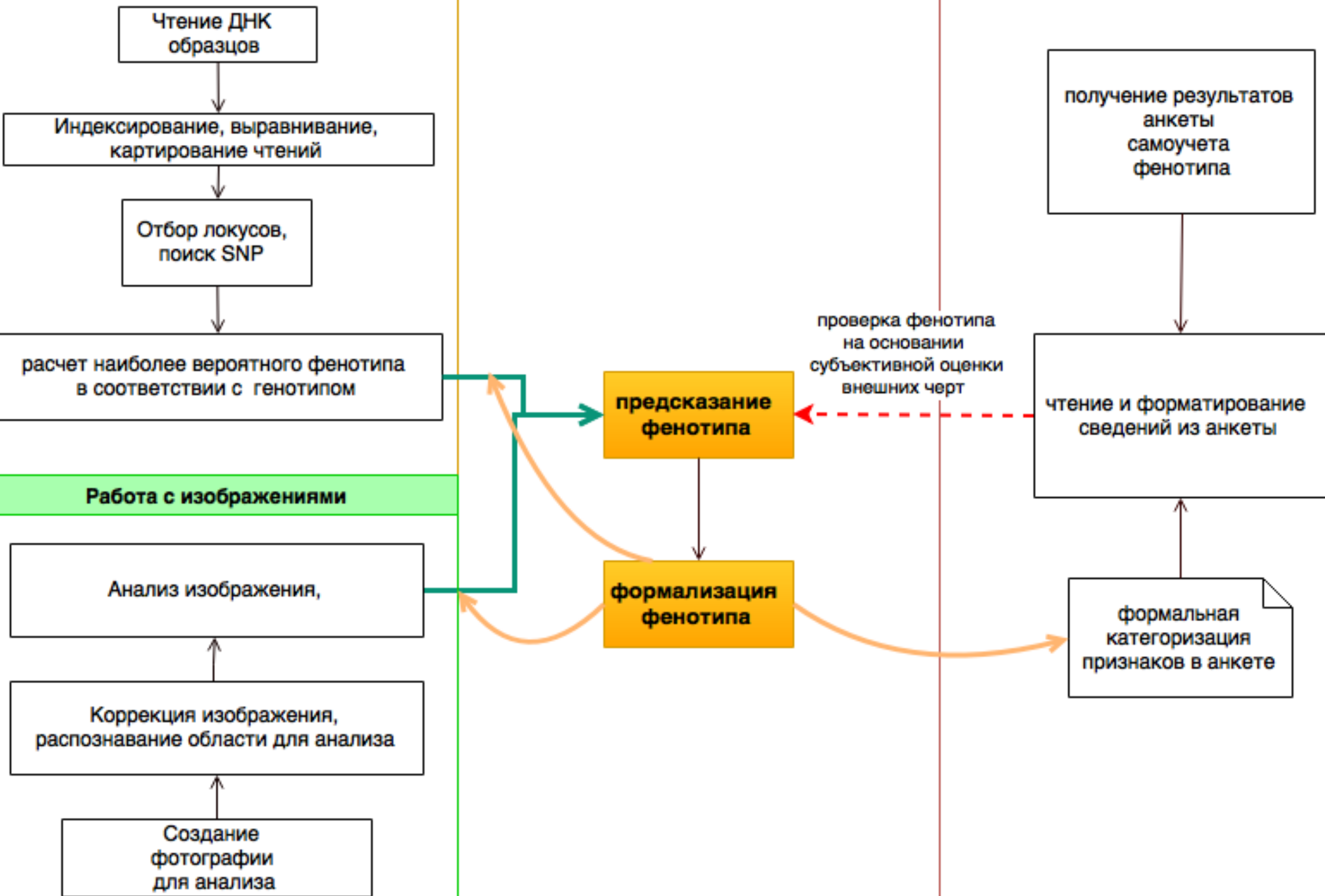
Валидация модели

получение результатов
анкеты
самоучета
фенотипа

чтение и форматирование
сведений из анкеты

формальная
категоризация
признаков в анкете

проверка фенотипа
на основании
субъективной оценки
внешних черт



КОНКРЕТНЫЕ ПРИМЕРЫ

- «рамка»

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

- Основные результаты работы представлены в публичном репозитории (<https://github.com/parseq/digital-phenotyping>)

ВЫВОДЫ

- модель IrisPlex — 60%
- (ИЛИ) субъективный учет фенотипа
недостаточно эффективен

ПЛАНЫ НА БУДУЩЕЕ

- нарисовать H-S пространство для пигментации глаз