# مشروع Zenome : ورقة العمل
# النظام الجينومي القائم على بلوكشين

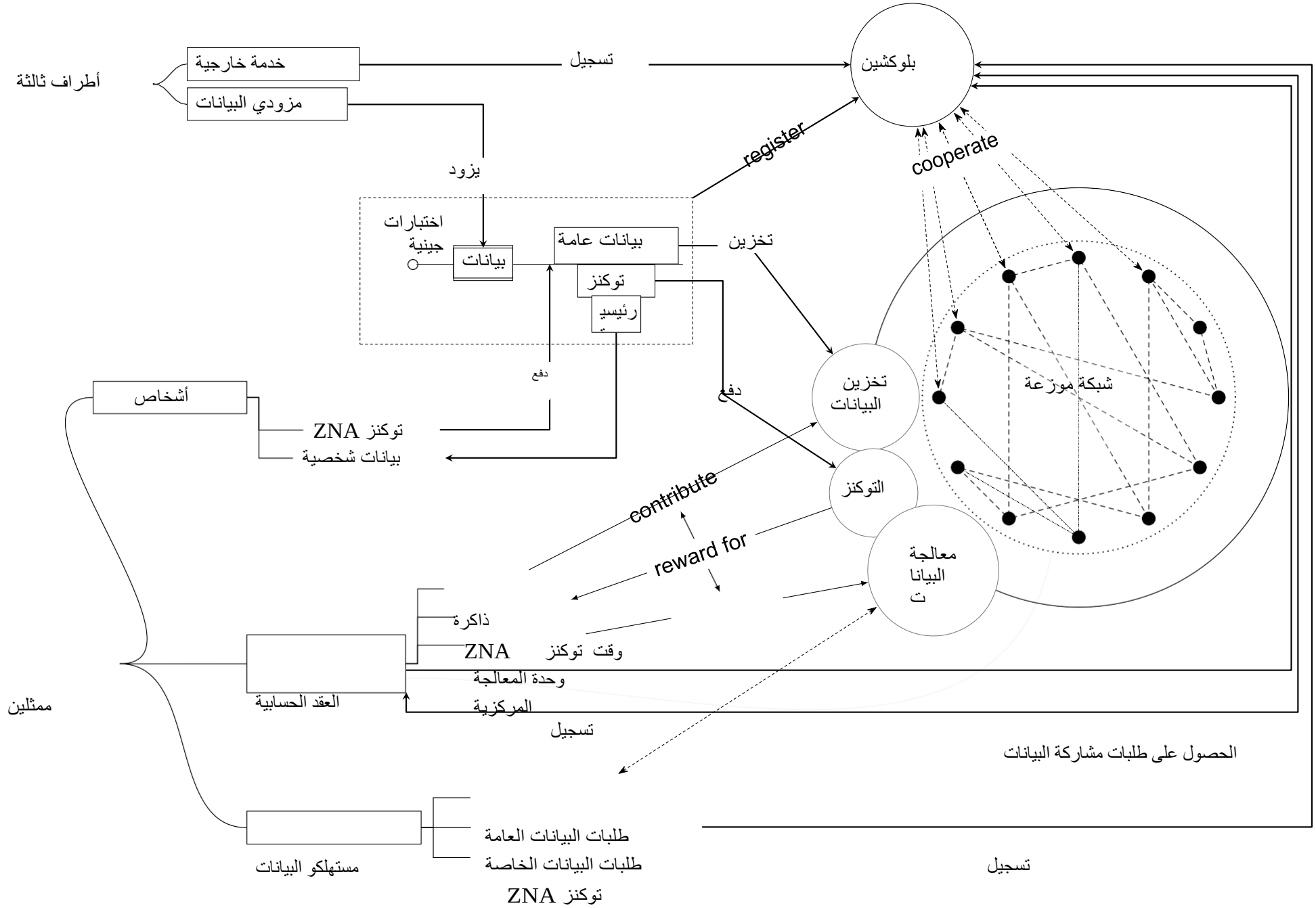
Nikolay Kulemin Sergey Popov Alexey Gorbachev

6 أكتوبر 2017

## ملخص


الصناعة 4.0 تعتبر عنوان للاتجاه الحالي للأتمتة والتحجيم وتبادل البيانات في تقنيات التصنيع.
يشمل الذكاء الاصطناعي ، والواقع الافتراضي ، وإنترنت للأشياء وتحليل البيانات الكبيرة ..Genomics ممثل قوي لصناعة 4.0 يتطلب حل العديد من المشاكل العاجلة ، مثل تخزين البيانات الكبيرة وتحليلها مع الحفاظ على الوصول العام للباحثين والخصوصية للناس.
حاليا هناك مشكلة عدم المساواة في صناعة الجينوم. وهذا يعني أن الجزء الرئيسي من البيانات الجينومية الشخصية كان يتركز في مراكز بيانات الشركات الجينية والحكومية والمؤسسات العلمية والطبية وشركات الأدوية. علاوة على ذلك ، هناك مشكلة تتعلق بالحدود القانونية للوصول إلى البيانات الجينومية الشخصية ، فضلاً عن عدم وجود إمكانية لإدارة البيانات الجينية ومشاركتها. هذا الاحتكار للبيانات الجينية يثبط بشكل كبير التطور في عدد من المجالات العلمية والطبية.
يؤدي تطوير تقنيات العملات الرقمية المبنية على بلوكشين إلى تحول كبير في العديد من المجالات الاقتصادية. سيكون تطبيق نهج بلوكشين الأساس مما يسمح لرفع مستوى تطوير الجينوميات الشخصية. سيجعل كل شخص مالكًا لبياناته الوراثية.
مشروع Zenome يعتبر قاعدة بيانات لامركزية تحركها مجموعة من المعلومات الجينومية. تدعم هذه المنصة إمكانية إدارة بياناتك الجينية مع الحفاظ على الخصوصية والقدرة على جني الربح من بيع الوصول إلى أجزاء مختلفة من الجينوم. سيحدد شروط متساوية لتطوير العقاقير ولتقدم التقنيات العلمية والطبية.
Zenome بيئة اقتصادية جديدة تعتمد على البيانات الجينومية وتقنية بلوكشين .إن تنفيذ نموذجنا المفاهيمي سوف يحل الصعوبات التالية:

- إنشاء بنية تحتية لتخزين البيانات الجينومية الضخمة باستخدام قاعدة البيانات الموزعة
- الوصول المفتوح إلى ملايين الجينوم البشري في جميع أنحاء العالم مع حماية الخصوصية
- إمكانية مشاركة كل شخص في البحث العلمي والطبي وتحقيق الربح من ذلك
- تحفيز تعزيز العلوم الجينومية في البلدان النامية وإزالة احتكار البيانات الجينومية في البلدان المتقدمة


أطراف ثالثة
خدمة خارجية
مزودي البيانات
تسجيل
بلوكشين
يزود
register
cooperate
اختبارات جينية
بيانات
بيانات عامة
توكنز رئيسي
تخزين
دفع
دفع
أشخاص
توكنز ZNA
بيانات شخصية
تخزين البيانات
شبكة موزعة
التوكنز
contribute
reward for
معالجة البيانات
ذاكرة
ZNA
وقت توكنز
وحدة المعالجة المركزية
ممثلين
العقد الحسابية
تسجيل
الحصول على طلبات مشاركة البيانات
طلبات البيانات العامة
طلبات البيانات الخاصة
مستهلكو البيانات
تسجيل
توكنز ZNA

# الفريق

## Alexey Gorbachev

المؤسس
علم الأحياء الجزيئي وخبير بلوكشين
شهادة الدكتوراة في البيولوجيا الجزيئية والكيمياء الحيوية

الجامعة: جامعة موسكو الحكومية ، قسم البيولوجيا الجزيئية.

يتمتع بخلفية علمية كبيرة وخبرة في إدارة الأعمال وإدارة المشاريع. في Zenome هو المسؤول عن الرؤية الشاملة وتطوير الأعمال.

البريد الإلكتروني : alex@zenome.io
لينكيدين https://www.linkedin.com/in/alexey-gorbachev-24b5305b/

## Nikolay Kulemin

المؤسس
شهادة الدكتوراة في المعلوماتية الحيوية

الجامعة: معهد موسكو للفيزياء والتكنولوجيا ، قسم الفيزياء الجزيئية والبيولوجية.

متخصص في المعلوماتية الحيوية والبيولوجيا الرياضية ، لديه خبرة في البحوث الأكاديمية والتطورات التطبيقية في صناعة الجينوم. نيكولاي هو مؤسس شركة تقوم بتطوير خوارزميات جديدة للتحليل الجينومي.

البريد الإلكتروني :nick@zenome.io
لينيكيدين : https://www.linkedin.com/in/nikolay-kulemin-50ab47a9/

## Vladimir Naumov

عالم البيانات الجينومية

عالم البيانات في علم الجينوم البشري
ي الجامعة: جامعة بيرغوف الروسية الطبية البحثية الوطنية
متخصص المعلوماتية الحيوية وتحليل البيانات. رئيس سابق في iBinom ، عالم المعلوماتية الحيوية في GERO ، **8** سنوات في صناعة الجينوميات. يعمل على إنشاء خطوط الأنابيب وتطوير طرق جديدة في تحليل ومرئيات البيانات الجينومية.

البريد الإلكتروني : vov@zenome.io
لينكيدين :https://www.linkedin.com/in/vladimir-naumov-8285a25b/

## Sergey Popov

مطور

،النظم الموزعة ، الرياضيات البحتة بلوكشين

الجامعة: معهد موسكو للفيزياء والتكنولوجيا ، قسم الفيزياء العامة والتطبيقية

لديه خبرة في الفيزياء النظرية ، المعلوماتية النظرية ، الرياضيات البحتة ، تطوير الأنظمة الموزعة والعقود الذكية النامية.

البريد الإلكتروني :sp@zenome.io

## Dmitry Kwon

مستشار

شهادة الدكتوراة في البيولوجيا الجزيئية ، مدير تطوير الأعمال

الجامعة: جامعة موسكو الحكومية ، قسم البيولوجيا الجزيئية

متخصص في علم الوراثة الجزيئية. يتمتع بخلفية علمية كبيرة ، ويمتلك تقنيات تحليل وراثي ، وأسواق الجينوم والتشخيص ، ولديه خبرة تجارية ناجحة في مجال التكنولوجيا الحيوية ، وشركات Dx العالمية.

البريد الإلكتروني :dk@zenome.io
لينكيدين :https://www.linkedin.com/in/dmitry-kwon-2763b119/

# المحتويات

## الفريق

## اقتصاد توكن **GENOME**

## الجزء الأول

# GENOMICS


**خلفية.....7**

*Genome* (ص 7). مشروع الجينوم البشري الدولي (ص 8). تطوير تحليل الجينوم (ص 8). التأثير على العلوم الأخرى (ص 9). قواعد البيانات الوراثية (ص 10)

**نظرة عامة على السوق الجينومية.....10**

ديناميات نمو سوق الجينوم (ص 10). مجالات المنتج (ص 12)

**تحديات علم الجينوم ..... 12**

انخفاض توافر تحليل الجينوم (ص 13). المساومة الخاصة مثل أسعار المشاركة (ص 14). الدراسات العلمية والطبية المتعددة المراكز (ص 14). البنوك الحيوية التي تخزّن المواد من مختلف الأفراد (ص 14). تعديل البرامج التحليلية لتطبيق واسع النطاق (صفحة 15). تفسير البيانات الجينومية (ص 15). تفسير البيانات الجينومية (ص. 16). تخزين آمن لبيانات كبيرة جدا (ص 17). قاعدة بيانات تحتوي على استبيانات مستكملة باستمرار (الصفحة 18). المستقبل القريب للجينوميات (ص 20)

**القضايا الأخلاقية لعلم الجينوم الشخصية .....20**

الخصوصية (ص 21). دعاية (ص 21). الحق في امتلاك البيانات الجينومية؟ (ص 22). الحق في الوصول إلى المعلومات الجينومية (ص 23)


## الجزء 2 - المفهوم


**مشروعZENOME .....24**

عرض فلسفي (ص 24). النظام الاقتصادي لمنصة جينوم (ص 25). نظرة عامة على بنية النظام (ص 25). سوق الخدمات الوراثية (ص 27)

**أدوار في منصةZENOME .....28**

من وجهة نظر نقاط الموارد (ص 28). الشخص / المستخدم (ص 29). مستهلك البيانات (ص 30). مزود الخدمة (ص 30)

**بيانات جينية.....31**

أنواع بيانات *omics*(الجينوميات) (ص 31). معالجة البيانات الوراثية (ص 31 .مشكلة البيانات المخادعة (ص 32). مشكلة تحديد المستخدم على أساس البيانات الجينومية (ص. 32). تخزين البيانات الجينومية (ص 33)

**الملفات الشخصية .....33**

مواصفات الاستبيان (ص 34). استعلامات للنظام. (ص 34). نقل آمن للبيانات الشخصية (ص. 35)

**نظام التقييم ..... 36**

**حالات استخدام..... 36**

مستخدم فردي (ص 36). الصحة (ص 37). شركة (ص 38). المجتمع العلمي (ص 38)

## الجزء 3 - الجزء التقني

**الأهداف الموزعة ..... 39**

مفهوم النظام الفرعي الموزع (ص 39). العمليات الداخلية (ص 40)

**توزيع أنظمة فرعية الرئيسية للمنصة ..... 40**

**يطبقة بينية منخفضة المستوى ..... 41**

شبكة بين الأنداد الموزعة (ص 41). خصائص توزيع الشبكة الموزعة (صفحة 41). تبادل الرسائل في بيئة موزعة (ص. 42)

**بلوكشين..... 42**

المعلومات الأساسية (ص 42). نظام فرعي للعمل مع بلوكشين( ص. 43). التوكنز الداخلية (ص. 43). مفهوم الحساب (ص 44)

**شبكة تخزين بيانات موزعة ..... 44**

مبادئ تشغيل الشبكة الموزعة (ص 44). ضمان موثوقية تخزين البيانات (صفحة 45). ضمان خصوصية تخزين البيانات (صفحة 45). ميزات محددة لتخزين المعلومات الجينومية (ص 45)

**بيانات المستخدم الشخصية..... 46**

مواصفات الاستبيان (ص 46)

# مقدمة: اقتصاد توكن Genome

في ما يلي سيتم تقديم وصف عام لمنصة*Zenome* .وتناقش الحاجة إلى التوكنز والفوائد للمستثمر المحتملين.

---

بالنسبة للجزء الأكبر ، يتم تخزين المعلومات الجينومية في قواعد البيانات ، التي تمولها الحكومات أو الشركات الكبرى. على نحو فردي ، تحتوي كل قاعدة بيانات على بيانات غير كافية لجعل القفزة الكمية تتجه نحو عصر الجينوم والطب الدقيق. في الوقت نفسه ، تحتوي كل قاعدة بيانات على الكثير من المعلومات التي يستحيل على شركة واحدة معالجتها.
يبدو أن تبادل المعلومات الوراثية له أهمية حاسمة. يجب أن تضمن السوق الوراثية المرتقبة الحماية من سوء الاستخدام المحتمل والتمييز الجيني بشكل خاص. من المهم بشكل خاص الحفاظ على الشفافية والوصول المتساوي إلى هذه السوق.
يجب أن يتناول التبادل العالمي للمعلومات الجينية المسائل التالية:

- تجزئة البيانات الجينية.
- محدودية وصول العلماء والمسعفين والشركات إلى البيانات الوراثية.
- انخفاض القدرة على تحمل تكاليف الاختبارات الجينية.
- عدم خصوصية أولئك الذين وافقوا على فتح باب المشاركة في بياناتهم الجينومية.
- موارد الحوسبة غير الكافية.

يهدف Zenome إلى إنشاء البنية التحتية الشخصية للجينوم ، والتي من شأنها تمكين المشاركين من:

- تحميل المعلومات الجينية والسيطرة عليها

- تخزين آمن للمعلومات الوراثية الخاصة.
- تحقيق الربح عن طريق بيع الوصول إلى البيانات الوراثية أو جزء منها.
- الخضوع للاختبارات الجينية في مقابل الحق في استخدام المعلومات الجينية.
- الحصول على توصيات غذائية فردية أو برنامج تدريب شخصي على أساس التركيب الوراثي.
- الاستفادة من الخدمات الوراثية الأخرى.

العملاء الرئيسيون للمعلومات الجينية هم من الشركات المهتمة بالاستهداف الجيني مثل غوغل و الفيسبوك Unilever وشركات الأدوية.

داخل منصة Zenome تتشابك أنواع المعلومات المختلفة ، وهي البيانات الجينومية والشخصية والمالية ، بشكل لا يتم تفريقه. تحدد الطبيعة المحددة لكل نوع طريقة تخزين المعلومات من هذا النوع. يتم تخزين البيانات المالية ، والتي تشمل سجلات المعاملات ، على بلوكشين .يتم تخزين البيانات الجينومية مجهولة المصدر على الشبكة الموزعة. يتم الاحتفاظ بالبيانات الشخصية للمشاركين على أجهزة الكمبيوتر الخاصة بهم فقط. معالجة البيانات من نوع مختلف يوفر خصوصية وكذلك قابلية النظام.

وبما أن جميع معاملات البيانات، بما في ذلك بيع وشراء البيانات، تحكمها عقود الذكية، الأمر الذي يعكس الطبيعة اللامركزية للمنصة، فالتفاعلات يمكن أن تشمل فقط الأرصدة المخزنة على بلوكشين .استخدام أي من التوكنز الموجودة مسبقًا لهذا الغرض سيؤدي إلى اعتماد غير معقول على تقييم التوكن الخارجي أو العملة. بالتالي ، يجب إصدار توكن نفعي منفصل لتشغيل و إدارة التفاعلات الاقتصادية على المنصة. هذا ، على وجه الخصوص ، يعني أنه لا يمكنك شراء البيانات الوراثية بالمال «العادي» ، يجب عليك الحصول على التوكنز أولاً.

Zenome DNA (ZNA)هو توكن في منصةZenome .يرتبط تقييم ZNA بنجاح المنصة.

في ورقة العمل هذه سنناقش بمزيد من التفصيل المشاكل الأكثر إلحاحًا في مجال علم الجينوم ، بالإضافة إلى حل منصة Zenome .

# الجزء 1

# Genomics

## خلفية

في هذا القسم ، يتم تعريف "الجينوم" و "الجينوميات". يتم النظر في تاريخ أول جهود للتسلسل الجيني وظهور التكنولوجيا التسلسلية عالية الإنتاجية*(NGS)* . الشركات المصنعة الرئيسية للكواشف والمعدات المستخدمة تم استخدامها للحصول على بيانات الجينوم. يتم النظر في القضايا المتعلقة بتراكم البيانات الجينية والحد من تكلفة التحليلات. يتم توفير مراجعة لقواعد البيانات الجينومية الحالية.

---

## Genome

الجينوم هو مجموعة كاملة من التعليمات الجينية الموجودة في الخلية [1]

يحتوي الجينوم على المعلومات البيولوجية اللازمة لتطوير وعمل الكائن الحي. يتكون الجينوم البشري من جزيئات DNA مزدوجة حلزونية خطية منظمة في 22 زوجا من الكروموسومات بالإضافة إلى اثنين من الكروموزومات الجنسية X -وY .يتم تشفير جميع المعلومات الواردة في الجينوم باستخدام الكود الرباعي من خلال سلسلة من 4 النيوكليوتيدات المعينة A و T و C وG .يعني مصطلح "قراءة جينوم" "تحديد تسلسل نوكليوتيدي بعملية متسلسلة "[2]

يحدد التسلسل الفردي للجينوم مجموعة متنوعة من السمات العضوية ، بما في ذلك المظهر ، والقابلية للإصابة بأمراض معينة ، والقدرة الرياضية ، والتمثيل الغذائي ، والأفضليات الغذائية ، والتوافق مع الشركاء الجنسيين (القدرة على الحمل بالأطفال) ، وغيرها الكثير.

## مشروع الجينوم البشري الدولي

تم إطلاق مشروع الجينوم البشري الدولي الأول تحت إشراف المعاهد الوطنية للصحة (NIH) في عام 1990 لتحديد التسلسل الكامل للجينوم البشري الفردي. كان قائد المشروع الأولي أحد الباحثين عن بنية الحمض النووي ، الحائز على جائزة نوبل ، جيمس واتسون

تم الانتهاء من مسودة تسلسل الجينوم البشري في منتصف عام 2000 ونشرت في بداية عام 2001 في مجلة نيتشر. بلغت تكلفة هذا المشروع الدولي بتمويل عام حوالي 3 مليارات دولار. في عام 1998 ، انضمت شركة خاصة ، هي Celera Genomics ، إلى سباق تسلسل الجينوم البشري. كان قائد المشروع الخاص ، الذي تم تطويره بالتوازي مع المؤسسات الحكومية ، عالمًا مشهورًا ورائدًا تجاريًا كريج فنتر ، الذي نجح في جمع 300 مليون دولار من الاستثمارات الخاصة لمشروع سيليرا. باستخدام نهج التسلسل الجديد للأسلحة النارية والأساليب الحسابية الأكثر إنتاجية ، تم نشر تسلسل جينوم كريج فنتر في وقت واحد تقريبًا مع البيانات التي أنتجها الكونسورتيوم الدولي في عام 2001 [3] في مجلة ساينس. تم نشر الجينوم البشري "الكامل" في عام 2007 ، ولا تزال بعض المناطق الجينومية البشرية التي يصعب ترتيبها غير معروفة.

## تطوير تحليل الجينوم

لقد قدمت الاستثمارات الواسعة ، وعدد كبير من المشاركين البارزين من المجتمع العلمي ، والمنافسة بين المؤسسات الخاصة والعامة زخما كبيرا لتطوير تكنولوجيات تحليل الجينوم. نتيجة لذلك ، ظهرت تكنولوجيات التسلسل الحديثة مثل NGS (الجيل التالي من التسلسل) 2 مع فرع جديد من العلوم يسمى المعلوماتية الحيوية ، وهو حقل بحثي جديد يجمع بين الرياضيات وتكنولوجيا المعلومات والبيولوجيا ، التي تقوم بتطوير التقنيات والخوارزميات لتحليل مجموعات البيانات البيولوجية الكبيرة في الطرق الإنتاجية والحسابية الفعالة.

أدى ظهور تكنولوجيات التسلسل من الجيل الثاني والثالث (NGS) إلى انخفاض كبير في تكلفة تحليل الجينوم. في حين أنه حتى في عام 2009 كانت تكلفة التحليل الكامل لجينوم حوالي 100000 دولار ، فإن متوسط السعر لنفس التحليل انخفض في الوقت الحالي إلى أقل من 1000 دولار تقريبًا (انظر الشكل 1 والجدول 3)

[1]https://en.wikipedia.org/wiki/Human_Genome_Project

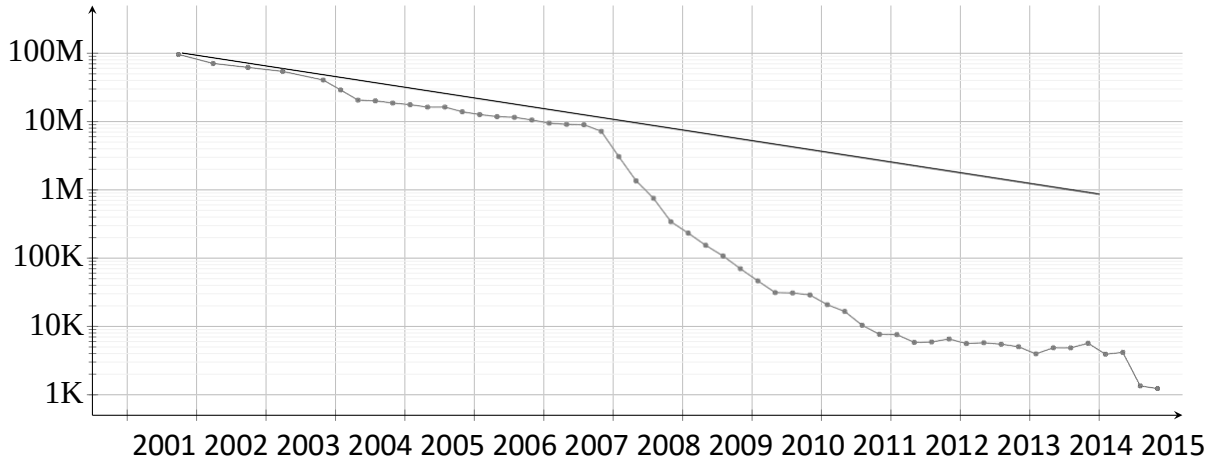
[2]https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-you-will-learn/what-next-generation-dna-

https://en.wikipedia.org/wiki/Moore's_law

الجدول 1: مطورو المعدات وموردي الكاشف للتسلسل الجينومي. القيمة السوقية المأخوذة منYahoo Finance .

| الشركة | المنتجات | رأس مال | الدولة |
|---|---|---|---|
| Illumina | الأجهزة والكواشف والمواد الاستهلاكية والبرمجيات | 28.06 B | USA |
| Thermo Fisher Scientific | الأجهزة والكواشف والمواد الاستهلاكية والبرمجيات (جزء من الأعمال) | 68.98 B | USA |
| Oxford Nanopore Technologies | الأجهزة والكواشف والمواد الاستهلاكية | 534.41 M | UK |
| Pacific BioScience | الأجهزة والكواشف والمواد الاستهلاكية | 436,93 M | USA |
| Roche | الأجهزة والكواشف والمواد الاستهلاكية والبرمجيات (جزء من الأعمال) | 213.44 B | Switzerland |
| Agilent Technologies | الأجهزة والكواشف والمواد الاستهلاكية والبرمجيات (جزء من الأعمال) | 19.32 B | USA |

تكاليف التسلسل ، بالدولار




الشكل 1: قانون مور وخفض تكاليف التحليل الجينومي. يعود الانخفاض الكبير في الأسعار في عام 2008 إلى ظهور تكنولوجيات التسلسل في الجيل التالي.[4] (NGS)

## التأثير على العلوم الأخرى

أدى تطور علم الجينوم (مجال العلوم الذي يدرس مختلف الجينوم) إلى تحول العديد من المجالات العلمية ، من علم الأحياء وعلم الإنسان إلى الطب وحتى العلوم الاجتماعية. حدد عدد من الشركات التجارية الكبرى مثل غوغل و آبل و آي بي أم و أمازون وعلي بابا هدف استخدام الجينوميات لتعديل منتجاتها وخدماتها وفقًا للملفات الجينومية لعملائها. ستسمح هذه التعديلات لهذه الشركات بضبط العلاقات مع المستخدم والتنبؤ باحتياجات عملائها وأنشطته المحتملة [3].

### قواعد البيانات الجينية

أدى انخفاض تكلفة التسلسل إلى زيادة هائلة في البيانات الجينومية المتاحة. على سبيل المثال ، يمكن أن يمثل الجينوم البشري الكامل في نسق يسمى "البيانات الخام" 50 جيجابايت إلى 2 تيرابايت من البيانات (حسب عمق التسلسل المطلوب). لتخزين مثل هذه الكمية الكبيرة من البيانات الجينومية ، تم إنشاء قواعد بيانات خاصة للجينوم تحتوي على أنواع مختلفة من البيانات ، مثل البيانات الخام التي تم الحصول عليها من متسلسلات الجينوم ("القراءات" أو "القراءات") ، متسلسلات الجينات والبروتينات ومجموعات مناطق الترميز. من جينوم يسمى exomes ، وحتى تسلسل الجينوم الكامل (السقالات) ؛ تحتوي بعض قواعد البيانات هذه على معلومات ذات صلة طبية بالإضافة إلى العلاقات بين الصفات الوراثية والأمراض. تتم إدارة معظم قواعد البيانات هذه بشكل مركزي وتمويلها من قبل الحكومات أو الشركات الكبرى. ويشارك العلماء في جميع أنحاء العالم في إضافة بيانات جديدة إلى قواعد البيانات هذه ، مما يتيح التحديث السريع والتزامن. في الجدول [2] ، يتم وصف بعض قواعد البيانات المعروفة.

تستضيف غالبية قواعد البيانات هذه في البلدان المتقدمة وتتم إدارتها مركزياً وتسيطر عليها الحكومات. الوصول إلى بعض قواعد البيانات هذه مقيد ، حتى بالنسبة للمجتمع العلمي ، أو محدود من خلال الاشتراكات التجارية. على الرغم من أن مؤسسي قواعد البيانات الجينومية يزعمون أنهم يخزنون البيانات الجينومية بشكل آمن ومجهول ، إلا أن البيانات المخزنة هي في الواقع مجرد اسم مستعار ، كما في بعض الحالات ، تم التعرف على الأفراد استنادًا إلى معلوماتهم الجينومية [5]

## نظرة عامة على السوق الجينومي

في هذا القسم ، يتم تقديم تحليل موجز لسوق تقنيات الجينوم. يتم وصف أمثلة من المنتجات الجينومية الأكثر شهرة والشركات التي تقدم خدمات الجينوم المختلفة.

---

### ديناميات نمو سوق الجينوم

إن سوق التكنولوجيات الجينومية ينمو بسرعة ويعد واعداً للغاية. في الوقت الحالي ، يبلغ إجمالي حجم السوق حوالي 25 مليار دولار مع نمو يقارب عشرة أضعاف ، من 5.9 مليار دولار في عام 2010 إلى 60 مليار دولار في عام 2020 (متوقع)

[3]https://www.smeal.psu.edu/fcfe/documents/innovations-in-medical-genomics-pdf

الجدول 2: قواعد البيانات الجينومية

| | |
|---|---|
| GenBank | http://exac.broadinstitute.org |

المالك: NCBI-NIH, USA
المنتج: تسلسل قاعدة البياناتGenome
التخزين: أكثر من 199341377 تسلسل جينوم مختلف

| | |
|---|---|
| ExaC | www.ncbi.nlm.nih.gov/genbank |

المالك: معهد واسع من معهد ماساتشوستس للتكنولوجيا وجامعة هارفارد ، الولايات المتحدة الأمريكية ، رخصة قاعدة البيانات المفتوحة(ODBL) ODCL
المنتج: اتحاد جمعExome
التخزين: 60706عينات / تسلسل exome البشري

| | |
|---|---|
| UniprotKB | www.ebi.ac.uk/uniprot/ |

المالك: EMBL-EBI, SIB, PIR, UK, Switzerland, USA
المنتج: فتح قاعدة المعرفة. دليل خبير التنظيم. البروتينات وسلسلة الجينات
التخزين: أكثر من 555100 سجل تم مراجعته يدويًا ووضع تعليقات توضيحية.

| | |
|---|---|
| ClinVar | https://www.ncbi.nlm.nih.gov/clinvar/ |

المالك: NCBI-NIH, USA
المنتج: المحفوظات المتاحة بحرية لتفسيرات الأهمية الطبية لمتغيرات الجينومية:
158000 < تفسيرات مقدمة ، يمثل> 125000 متغير.

| | |
|---|---|
| HGMD | http://www.hgmd.cf.ac.uk/ac/index.php |

المالك: QiaGen
المنتج: قاعدة بيانات تجارية
التخزين: 208368 سجلات طفرة بشرية مع ملاحظات

| | |
|---|---|
| SNPedia | https://www.snpedia.com/index.php/SNPedia |

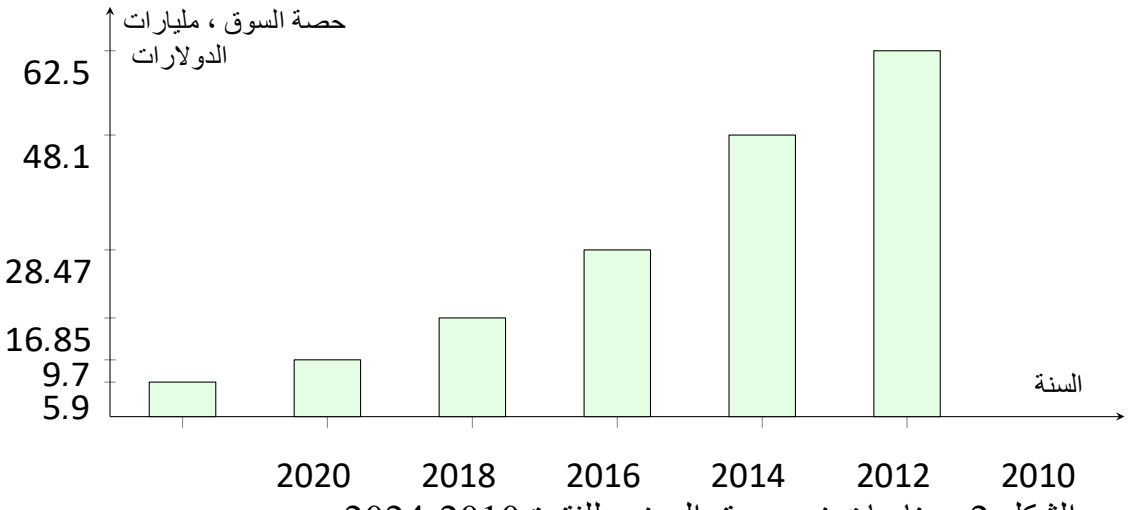
المالك: قاعدة بيانات مفتوحة
المنتج: SNPediaهو wiki يحقق في علم الوراثة البشرية
التخزين: 073 SNP،s107 والسجلات المرتبطة

| | |
|---|---|
| 1000 Genomes Project | www.1000genomes.org |

المالك: EMBL-EBI, Wellcome Trust
الهدف: ي العثور على معظم المتغيرات الجينية مع ترددات على الأقل 1 ٪
التخزين: أكثر من 2504 عينة / تسلسل جينوم

| | |
|---|---|
| 100000 Genomes Project | http://www.genomicsengland.co.uk/ |

المالك: NHS, Government of UK
المنتج: قاعدة بيانات حكومية بريطانية تحتوي على سلسلة من 100000 جينوم
التخزين: 32،642 تسلسل جينوم كامل




الشكل 2: ديناميات نمو سوق الجينوم للفترة 2010-2024..

## مجالات المنتج

r كمكون أساسي في الصناعة 4.0 ، تمتلك الجينوميات مجموعة واسعة من التطبيقات المحتملة في جميع المجالات الاقتصادية تقريبًا. مجالات المنتجات الرئيسية لسوق الجينوميات الحالي:

- **NIPT (اختبار ما قبل الولادة غير الجراحي)**

| | |
|---|---|
| وصف | استنادا إلى الحمض النووي من أصل الجنين المنتشرة في دم الأم. يمكن للاختبار تحديد عدم توازن الصبغية الجنيني [6] والجنس الجنيني في وقت مبكر من ستة أسابيع من الحمل. |
| حصة السوق | 4مليارات دولار |
| قادة السوق: | Illumina, Natera, Ariosa, Sequenom. |

- **PGS (**فحص جيني قبل الزراعة**)**

| | |
|---|---|
| الوصف : | استنادا إلىDNA المايكروي أو NGS التنميط الوراثي المتسلسل للأجنة قبل الزرع أثناء إجراء التلقيح الاصطناعي (بالتخصيب في المختبر) |
| حصة السوق | 336.4 مليون دولار |
| قادة السوق: | Illumina, Agilent Technologies |

- **DTC (**مباشرة للمستهلك**)** ، اختبار جيني-**SNP**التنميط الجيني

| | |
|---|---|
| الوصف | اختبار وراثي يعتمد على التنميط الوراثي لـ **DNA-microarray SNP** للحصول على بعض التوصيات التالية: النسب ، احتياجات التغذية ، حالة الناقل ، التمرين الأمثل.. |
| حصة السوق: | 2مليار دولار |
| قادة السوق | AncestryDNA, 23andMe, DNAfit, deCode genetics. |

- **oncogenomics)** بما في ذلك**)** شركة تشخيص

| | |
|---|---|
| الوصف: | أنواع مختلفة من إجراءات التشخيص ، تشمل علم الوراثة. تسلسل لجينات الجينات ، Exomes، الجينوم الكاملة ، أخذ عينة سائل |
| حصة السوق | 16 مليار دولار أمريكي |
| قادة السوق : | Pathway Genomics, Human Longevity, Inc, Laboratory Corporation of America, Quest Diagnostics. |

# تحديات للجينوم

في هذا القسم ، يتم النظر في المشاكل الرئيسية في علم الجينوم المعاصر. يجب حل هذه المشاكل لتطبيق مفهوم*Genomics 2.0* التوسع والتطبيق في كل مكان للتكنولوجيات الجينومية الشخصية.:

## انخفاض توافر تحليل الجينوم

يتم سرد تكاليف أنواع مختلفة من تحليل الجينوم في الجدول 3 ، وبشكل عام ، يبدأ من 100 دولار. مقياس السعر هذا منخفض للغاية نسبيًا بالنسبة لتكلفة الجينوم قبل 5 أو 10 سنوات [7]

ومع ذلك ، فإن سعر التسلسل والتفسير الحيوي المعلوماتي لا يزال مرتفعا بما فيه الكفاية ، مما يجعل من الصعب على عامة الناس الوصول إلى التحليل الجينومي ، خاصة في البلدان النامية.

الجدول 3: أنواع تحليل الجينوم.

### DNA microarray

| | |
|---|---|
| الوصف: | تحليل 1-5 مليون SNPs المحددة مسبقا |
| نطاق السعر: | $100-500 |
| كمية المعلومات | 0.033٪ من الجينوم الكامل |
| مقدمي الخدمة | 23andMe, AncestryDNA, DNAfit. |

### Genes panels

| | |
|---|---|
| نظاف السعر : | $100-2000 |
| كمية المعلومات | 0.001-1% من الجينوم الكامل |
| مقدمي الخدمة | Pathway Genomics, CeGaT. |

### Exome sequencing

| | |
|---|---|
| الوصف: | تسلسل ترميز جزء من الجينوم(exome) |
| نطاق السعر : | $250-3000 |
| كمية المعلومات | 2%من الجينوم الكامل |
| مقدمي الخدمة | BGI, CeGaT. |

### Whole Genome Sequencing

| | |
|---|---|
| نطاق السعر: | $600-10000 |
| كمية المعلومات | 80-98%من الجينوم الكامل |
| مقدمي الخدمة | BGI, FullGenomes, Human Longevity. |

ملاحظة: في الواقع ، لن يتم تسلسل جزء من الجينوم. حجمه يعتمد على عمق التسلسل وتفاصيل عملية تحضير العينة. لذا فإن مصطلح "تسلسل الجينوم الكامل" يعني في الواقع الحصول على تسلسل أكثر قليلاً من 80٪ من الجينوم.

إن الحصول على كمية كبيرة من المعلومات الجينومية (متواليات الجينوم مع الخصائص المظهرية) لسكان الدول النامية أمر مهم للغاية من منظور الحصول على أوسع تنوع ممكن للمعلومات الجينومية ، والتي بدورها ستحفز بشكل كبير تطور الجينوم الكبير سوق البيانات. حتى في البلدان الأكثر تطوراً ، خضع أقل من 2٪ من السكان لأي تحليل جيني (ميكروأري أو تسلسل إكسوم أو تسلسل جينوم كامل)

## التسوية الخاصة كثمرة للمشاركة في التحقيقات الطبية الحيوية المفتوحة.

تعتبر الموافقة المسبقة على جمع ومعالجة البيانات الشخصية قضية رئيسية في كل دراسة بحثية طبية حيوية. يبدأ أي مشروع يتضمن دراسات الجينوم البشري بجمع التواقيع من الأفراد الذين يشهدون على حقيقة أنهم يفهمون النتائج ويتفقون مع شروط الدراسة البحثية. يختلف شكل هذه الموافقة اعتمادًا على المشروع وقد يشمل إعطاء الإذن لاستخدام البيانات في المشروعات المستقبلية ، والتي يمكن أن تكون عواقبها غير متوقعة.

في مشروع الجينوم الشخصي ، الذي تشرف عليه كلية الطب بجامعة هارفارد ، يوافق المشاركون طواعية على أن بياناتهم وعيناتهم من المواد الجينية يمكن استخدامها عدة مرات ويمكن إتاحتها للمختبرات الأخرى. يتم إخطار المشاركين في هذا المشروع على وجه التحديد بأن هويتهم قد تكون مجهولة الهوية وأن بياناتهم الخاصة يمكن أن تصبح عامة. يهدف هذا المشروع إلى ضمان أن تكون البيانات الجينومية من أكبر عدد ممكن من الناس متاحة بشكل مفتوح لتحفيز الأبحاث والتطوير في مجال علم الجينوم. يعتقد مؤلفو المشروع أنه إذا لم نوفر الوصول المفتوح إلى البيانات الجينومية وتبادل المعلومات ، فنحن في خطر أن ينتهي بنا المطاف مع الآلاف من مجموعات البيانات الجينومية المعزولة والمخزنة بشكل خاص (من شركات الأدوية والشركات الجينومية والمراكز العلمية). ) ، لكن كل من قواعد البيانات المنفصلة هذه لن تحتوي على بيانات كافية لتمكين اكتشافات الاختراق.

## إجراء دراسات علمية وسريرية دولية متعددة المراكز.

لإجراء البحوث الجديدة والتطوير في مجال علم الجينوم ، من الضروري إجراء دراسات علمية وسريرية على عينات كبيرة من مجموعات سكانية مختلفة. في الوقت الحالي ، من الصعب جمع العينات الجينومية من أفراد ذوي خلفيات عرقية مختلفة ، حيث يتعين إنشاء مشاريع خاصة ، والقيام بحملات استكشافية ، والحصول على إذن من الهيئات التنظيمية المحلية.

اليوم ، تحاول المشاريع الناشئة فقط معالجة هذه المشكلة /https: // www. dnasimple.org مقابل رسوم رمزية ومع وعد عدم الكشف عن هويته.

## إنشاء بنوك حيوية لتخزين المواد من مختلف الأفراد

تعمل البنوك الحيوية كوسيط بين المتبرعين للمواد البيولوجية (عينات الدم ، عينات نخاع العظم ، إلخ) والباحثين عن طريق معالجة المواد التي تم الحصول عليها وتخزينها للاستخدام في المستقبل. بشكل عام ، البنوك الحيوية هي أداة رئيسية ل

تقدم الطب (الدقيق) التشخيصي وتطوير الأدوية. واحدة من أهم وظائف البنوك الحيوية هي جمع المواد المانحة لاستخدامها في المستقبل ، بما في ذلك الدم ونخاع العظام ، وحتى الخلايا الجرثومية.

في الوقت الحالي ، يجري تطوير البنوك الحيوية بنشاط في العديد من البلدان. ومع ذلك ، هناك احتمال أن أكثر العينات قيمة سوف تنتمي إلى البنوك الحيوية الأكثر أمناً من الناحية المالية ، على سبيل المثال ، البنوك الحيوية التي تنتمي إلى شركات الأدوية الكبرى ، والتي ستؤدي إلى عدم المساواة في الوصول إلى المواد الحيوية لمختلف فئات الباحثين. ولذلك ، هناك حاجة لإنشاء بنوك حيوية في أكبر عدد ممكن من البلدان والمدن حتى حدود الاحتياطي البيولوجي الفردي لكل فرد.

## المعالجة: معدل تحليل الجينوم محدود بسبب معالجته الحيوية. تعديل البرامج التحليلية لتطبيق واسع النطاق.

حاليا ، سرعة الحصول على البيانات الجينومية باستخدام تقنيات تحليل الجينوم (التسلسل) عالية وتفوق سرعة معالجة هذه البيانات. إذا أخذنا في الاعتبار أي دراسة علمية واسعة النطاق لعدد كبير من الجينوم ، فإن الخطوات التجريبية للحصول على المعلومات الجينومية لا تمثل أكثر من 20٪ من مدة الدراسة ، بينما تشتمل معالجة البيانات على جزء أكبر بكثير من المشروع. هنا ، معالجة البيانات تعني الخطوات المتخذة من الحصول على بيانات التسلسل الأولية إلى تفسير النتائج والبحث عن مختلف الجمعيات.

مشكلة أخرى مع البرمجيات الحديثة هي أن برامج التحليل الجينومي كانت مصنوعة من قبل العلماء ، وبالتالي تتطلب تعديلات يمكن تطبيقها على نطاق واسع بين الأطباء والمستهلكين بشكل عام. بما أن محللات الجينوم موجودة بالفعل بحجم جهاز USB ، فإن استخدام جهاز التسلسل الشخصي بنفس طريقة استخدام الكمبيوتر الشخصي يمكن اعتباره ليس خيالًا علميًا ، بل واقعيًا في المستقبل القريب.

## تفسير البيانات الجينومية: النماذج الرياضية للأمراض التي تطور المخاطر

يتم استخدام نماذج وخوارزميات مختلفة لتصنيف مخاطر الإصابة بالأمراض على أساس البيانات الوراثية. تعتمد الأنواع الأولية من هذه النماذج على نوع الميراث الذي تم دراسته: أحادي المنشأ ، متعدد الجينات ، متعدد العوامل. يتطلب تقييم مخاطر تطوير أمراض متعددة العوامل أو معقدة حساب تأثير وتدخل العديد من الجينات بالإضافة إلى العوامل البيئية. للحصول على وصف أكثر تفصيلاً للطرق المتاحة لتقييم مخاطر الأمراض المعقدة ، انظر [8 ، 9]

لتطوير نموذج جديد لتقييم مخاطر تطوير مرض ، يعتبر أمر

إشراك عدد كبير من العلماء مهم. من الضروري أيضًا التحقيق في العديد من المنشورات المتعلقة بالمرض الجاري تحليله ، لتحديد نوعه من الميراث ، لتحديد الأشكال المتعددة والطفرات التي تسهم في تطور المرض ، وتطوير "الجبر الجيني" ، أي مجموعة من القواعد. لتقدير المخاطر. عندما يتم إنشاء نموذج والتحقق منه حاسوبياً، يجب إجراء الدراسات السريرية لتقييم إمكانية تطبيقه. هذا النهج هو الأكثر دقة في الوقت الحالي ، ولكنه مكلف من حيث مقدار الوقت والجهد المطلوبين.

## تفسير البيانات الجينومية: تطبيق التعلم الآلي.

إن استخدام خوارزميات التعلم الآلي لتقييم مخاطر الأمراض متعددة العوامل يتم بحثه على نطاق واسع ، ولكن حتى الآن ، ونظراً لعدم وجود عدد كافٍ من عينات التدريب ، فإن النماذج الرياضية الحالية التي طورها علماء البيولوجيا تتفوق على أساليب التعلم الآلي.

ومع ذلك ، يتم استخدام التعلم الآلي للتنبؤ ببعض الخصائص المعقدة لجسم الإنسان. مثال على ذلك هو ظهور التنبؤ في أعمال كريج فنتر وزملائه. شمل جوهر عملهم تحليل الجينوم ونحو 30000 نقطة لبيانات الوجه من عدة آلاف من المتطوعين. استنادا إلى البيانات التي تم الحصول عليها ، تم بناء عينات التدريب لخوارزميات التعلم الآلي وتم تحديد التسلسل بين السمات الجينية والمظهر الفردي. بسبب هذا العمل ، تعلمت الآلات استعادة مظهر الشخص بدقة استنادًا إلى بياناته الجينية [10 ، 11]

تتيح نتائج هذا المشروع التنبؤ بظهور مجرم أو طفل لم يولد بعد أثناء المراحل المبكرة من الحمل. من خلال الحصول على عينة دم من امرأة حامل واستخراج DNA الجنين من الدم ، يمكن التنبؤ بدقة بظهور طفل لم يولد بعد في عيد ميلاده الثامن عشر.

لتنفيذ هذا المشروع ، قام Craig Venter بتوظيف واحد من أفضل المتخصصين في التعلم الآلي من غوغل ، وهو فرانز أوتش ، وهو عالم نجمي كمبيوتر يُعرف بالمهندس الرئيسي لترجمة غوغل.[12]

في الوقت الحالي ، لا يستخدم التعليم الآلي على نطاق واسع في الأمراض ، حيث أن هناك حاجة إلى عينات كبيرة جدًا ومصممة بشكل صحيح للتدريب. إن إنشاء قاعدة بيانات شاملة عن الجينوم البشري ، فضلاً عن توافر استبيانات تفصيلية تعكس الحالة الصحية للأفراد ، يمكن أن يحفز تطوير التدريب الحسابي في علم الجينوم ، وسيؤدي إلى دقة تنبؤية عالية في تحديد مخاطر تطور المرض. وفي الوقت نفسه ، ستكون هذه البيانات عامة ومتاحة لجميع مستخدمي النظام ، باستثناء إمكانية احتكارهم. وهذا التوافر بالغ الأهمية ، لأن تركيز كميات كبيرة من البيانات في قواعد بيانات الشركات سيؤدي إلى احتكارات في مجال تعلم الآلة الجينومية.

## تخزين آمن لبيانات كبيرة جدًا

أمن البيانات الشخصية مهم جدا. كلنا نحاول منع سرقة بيانات بطاقة الائتمان ، ومعلومات التأمين ، وأرقام الحسابات المصرفية ، والمعلومات الطبية من أي نوع. قد تبدو سرقة المعلومات الجينومية غير مهمة بالنسبة للكثير من الناس اليوم. ومع ذلك ، يمكن أن يؤدي ذلك إلى عواقب وخيمة جداً يصعب التنبؤ بها ، على سبيل المثال ، إمكانية تجميع جزء من التسلسل الجينومي للفرد وزرعه في موقع جريمة أو عمل إرهابي.

تتضمن الحلول الحالية لهذه المشكلة تخزينًا مشفرًا على خادم مركزي ، مثل [13] https://www.pathway.com/أو https://www.23andme.com/ أو www.humanlongevity.com//.هذا النوع من تخزين البيانات "المغلقة" هو نسبيًا آمن ، لكنه يحظر إمكانية تبادل البيانات وتوفير الوصول إلى العلماء من جميع أنحاء العالم ، وهو شرط بالغ الأهمية لتطوير علم الجينوم الحديث.

تتعلق مشكلة أخرى مرتبطة بالتخزين بحجم كبير للجينوم نفسه ، والنمو المتسارع لمقدار البيانات الجينية المتاحة ، حيث يتعرض المزيد والمزيد من الأفراد إلى تسلسل الجينوم. في إحدى الدراسات [14] ، بحلول عام 2025 ، من المتوقع أن يصل الحجم الإجمالي للبيانات الجينومية المخزّنة (نظرًا لحجم جينوم كامل هو 100 غيغابايت) إلى 40 إكسابايت في العام ، ومن المتوقع أن يصبح التخزين الجيني أحد أكبر المستهلكين لقدرات التخزين ومعالجة المعلومات.

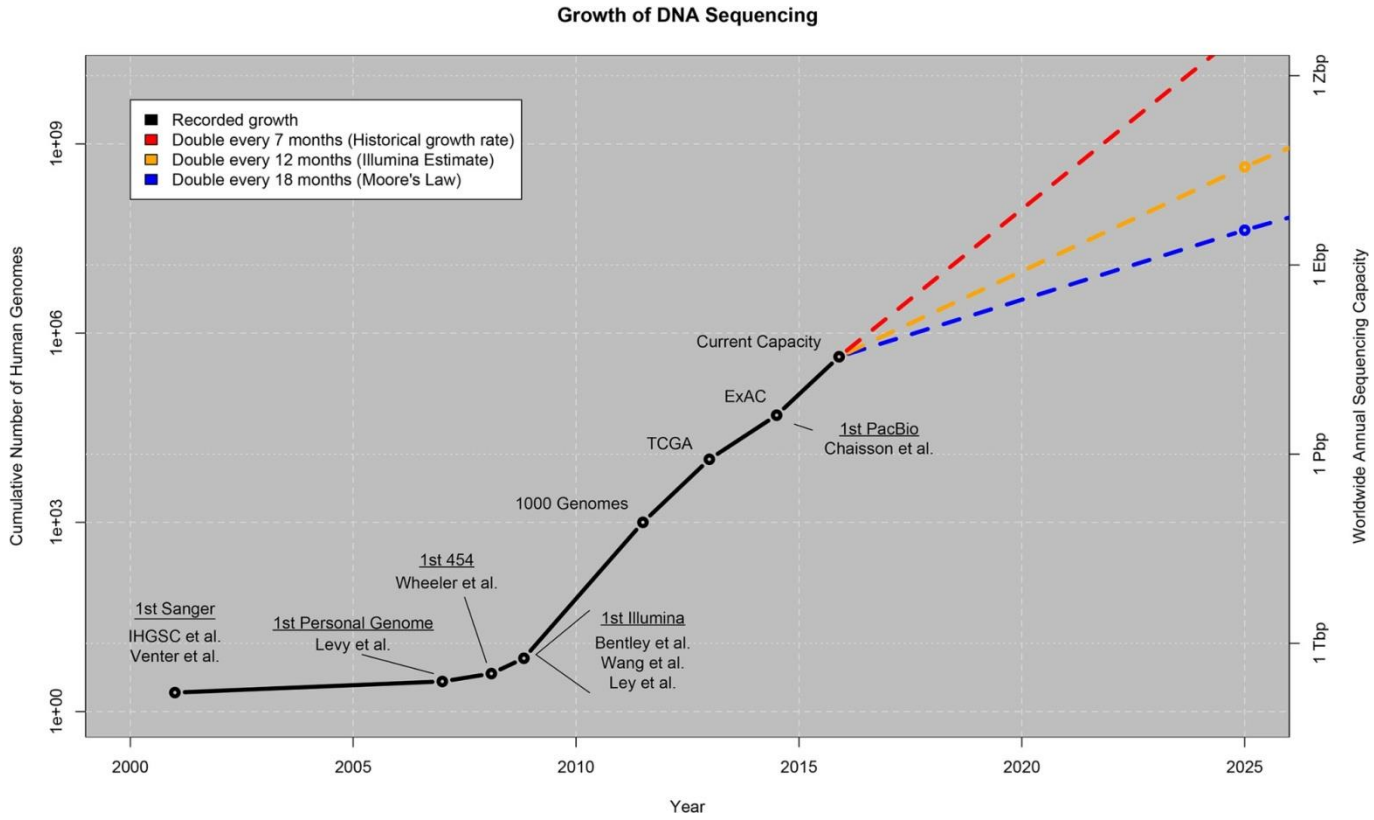
الجدول 4: أربعة مجالات للبيانات الضخمة في عام 2025.

| | | |
|---|---|---|
| **الفلك** | | |
| | الاكتساب | 25 zetta-bytes/year |
| | التخزين | 1 EB/year |
| | تحليل | تقليل البيانات ؛ المعالجة في الوقت الفعلي ؛ أحجام ضخمة |
| | التوزيع | |
| تويتر | | |
| | الاكتساب | 0.5-15 billion tweets/year |
| | التخزين | 1-17 PB/year |
| | تحليل | تحليل ميتا داتا |
| | التوزيع | |
| يوتيوب | | |
| | الاكتساب | 500-900 million hours/year |
| | التخزين | 1-2 EB/year |
| | تحليل | متطلبات محدودة |
| | التوزيع | |
| علم الجينوم | | |
| الاكتساب | 1 zetta-bases/year | |
| التخزين | 2-40 EB/year | |
| تحليل | البيانات غير المتجانسة والتحليل. استدعاء متغير ، ∽ 2 تريليون ساعة من وحدة المعالجة المركزية (CPU) ؛ تحالفات الجينوم لجميع الأزواج ، ∽ 10،000 تريليون ساعة وحدة المعالجة المركزية. | |
| التوزيع | العديد من عمليات نقل البيانات الضخمة (10 ميغابايت / ثانية) وعدد أقل (10 تيرابايت في | |

## الحاجة إلى قاعدة بيانات مشتركة مع استبيانات مستكملة باستمرار.

في الوقت نفسه ، يمكن أن يؤدي عدم وجود قاعدة بيانات عامة يتم تنفيذها استنادًا إلى مفهوم التخزين الموزع واستخدام برامج مفتوحة المصدر إلى الهيمنة الكلية لمنطقة السوق هذه من قِبل شركات مثل Google أو Amazon نظرًا لخوادمها القوية [4]. إذا أصبحت الشركات أو الشركات الصيدلانية محتكرًا في مجال المعلومات الجينومية ، فسوف نلاحظ التطور البطيء والمكلف للدواء مع استمرار النهج العلاجي الحالي بدلاً من المستقبلي الذي يمكن منع المرض فيه حتى قبل تطوره أو في المراحل الأولى من مظهره.

[4]https://cloud.google.com/genomics/




شكل 3: نمو بيانات تسلسل الحمض النووي. تظهر هذه البيانات نمو تسلسل الحمض النووي على حد سواء في العدد الإجمالي لجينوم الإنسان المتسلسل (المحور الأيسر) وسعة التسلسل السنوية العالمية (المحور الأيمن Tera-basepairs (Tbp) :، Peta-basepairs (Pbp) ، Exa-basepairs (Ebp) ، و Zetta-basepairs (Zbp). تستند القيم حتى عام 2015 على سجلات النشر التاريخية مع معالم بارزة في التسلسل (أول تسلسل سانجر من خلال أول جينوم بشري PacBio تم نشره) وكذلك ثلاثة مشاريع نموذجية باستخدام التسلسل الواسع النطاق: 1000 مشروع جينومز ، الذي جمع مئات الجينوم البشري بحلول عام 2012؛ Cancer Genome Atlas (TCGA)، الذي جمع عدة آلاف من أزواج الجينوم / الورم الطبيعي ؛ وجمعية Exome Aggregation Consortium (ExAC) ، التي جمعت أكثر من 60،000 من الأكسيدات البشرية. وتمثل العديد من الجينومات المتسلسلة حتى الآن exomes بالكامل بدلاً من الجينومات الكاملة ، لكننا نتوقع أن تكون هذه النسبة متحيزة بشكل متزايد نحو الجينوم بأكمله في المستقبل. تمثل القيم إلى ما بعد 2015 توقعاتنا تحت ثلاثة منحنيات نمو ممكنة ، كما هو موضح في النص الرئيسي. مأخوذة من [14]

## مستقبل الجينوميات القريب

في الختام لهذا الجزء ، من الجدير وصف بعض وجهات النظر والمخاطر الافتراضية والفنية التي يمكن أن تواجهها الصناعة الجينية والمجتمع بشكل عام:

- تخفيض تكلفة تحليل الجينوم وتصغير أجهزة التسلسل الجيني إلى مستوى المكونات الإضافية للهاتف الخلوي ؛
- النمو الهائل للبيانات الجينومية وتخزينها وظهور "المتسللين الجينومين" وقضايا الخصوصية (حماية البيانات) ؛
- التوزيع الشامل للطب الجينومي والرعاية الصحية عن بعد ؛
- التغييرات في صناعة الأغذية التي تنطوي على تنفيذ التغذية الشخصية على أساس الجينوم ؛
- تطوير العلاج بالعقاقير الشخصية ؛
- يؤرخ على أساس التوافق الجينومي.
- تحديد الشخصية باستخدام المعلومات الجينومية ، بما في ذلك القدرة على سداد المدفوعات والحصول على الخدمات ؛
- زيادة متوسط العمر الافتراضي في جميع البلدان ، وإطالة العمر الافتراضي النشط ، والتحيز القوي نحو كبار السن ، والحمل المتأخر ، وانخفاض معدل المواليد ؛
- تطوير تقنيات تحرير الجينوم ؛
- تصميم المظهر لأطفال المستقبل والتطورات الأخرى التي يصعب تخيلها. فعلى سبيل المثال ، توجد بالفعل إمكانية فنية لاختيار أطفال أصحاء وتحديد مظهرهم المستقبلي سواء في مرحلة الأجنة أو بالحصول على DNA الجنين من دم المرأة الحامل [10]

# القضايا الأخلاقية للجينوميات الشخصية

في هذا القسم ، يتم النظر في المشاكل الأخلاقية المرتبطة بالتطوير والتوزيع المنتشر للتكنولوجيات الجينومية ، مثل الخصوصية ، وقواعد البيانات العامة ، والوصول المفتوح للباحثين ، وإساءة الاستخدام المحتملة والتهديدات للحريات الشخصية بسبب التوسع في علم الجينوم.

NOTE

في ورقة العمل الحالية ، نقدم فقط مراجعة موجزة للمشاكل والرؤى الرئيسية للصناعة الجينومية. للاطلاع على مناقشة أكثر تفصيلاً لمشاكل الخصوصية وأمان البيانات ، انظر https://www.smeal.psu.edu/fcfe/ documents/innovations-in-medical-genomics-.pdf

## الخصوصية

معلومات الجينوم الشخصية حساسة جدا لكثير من الناس. ومع ذلك ، فإن العديد من الناس لا يفهمون بشكل كامل أنه ، استنادًا إلى المعلومات الجينومية الخاصة بهم ، من الممكن تحديد عمرهم ، والميل إلى اتخاذ قرارات عاطفية (تلاعب في اتخاذ القرار) ، واحتمال تطوير أمراض عقلية مختلفة وخطر الموت المفاجئ بسبب ، على سبيل المثال ، عدم انتظام ضربات القلب.

قد تكون هذه المعلومات غير مناسبة للتوظيف ، والمشاركة في الانتخابات ، وتسعير التأمين الطبي. هناك أيضا احتمال أن يكون الفاعل السيئ الذي يعرف تسلسل جينوم يمكن أن يترك أجزاء من الحمض النووي مماثلة لهذا الجينوم ، على سبيل المثال ، في موقع عمل إرهابي لإطار شخص ما أو اتهامه بطريقة غير مشروعة. يمكن أن يحرم المرء من العلاج الطبي (أو يُطلب منه دفع رسوم أعلى) أو يُحرم من الحصول على وظيفة مرغوبة.

يمكن للشركات والحكومات أن تؤثر عمداً على قرارات ومشتريات المرء باستخدام معرفته ب "نقاط الضعف" في المعلومات الجينومية لأحدنا. وبالتالي ، فإن حماية خصوصية البيانات الجينومية ضرورية لحماية الحقوق المتساوية لمختلف فئات الأشخاص.

وفي الوقت نفسه ، أجريت بعض الدراسات التي تسمح بتحديد هوية الأفراد استنادًا إلى جينوماتهم المجهولة [5 ، 15]

علاوة على ذلك ، تمتلك بعض الشركات (http://www.humanlongevity.com/media/) خوارزميات تستند إلى التعلم الآلي يمكنها أن تعيد بناء مظهر الفرد بدقة باستخدام بياناته الجينومية فقط [16]

## العلنية

سيمكن هذا النهج من تطوير الطب الوقائي الذي يسمح من خلاله تحليل كميات كبيرة من البيانات بالتنبؤ بتطور المرض (قبل حدوثه) ، وتمكين الإجراءات التي تزيد من العمر وتحسين نوعية الحياة [17] بالإضافة إلى تحديد المانحين في جميع أنحاء العالم لتلبية الاحتياجات الطبية المختلفة. للأسف ، لا توجد حاليًا حلول صالحة للاستخدام العام للمعلومات الجينومية مع الحفاظ على الخصوصية الفردية. ومع ذلك ، ينبغي ذكر بعض الشركات الناشئة العاملة في هذا المجال باستخدام تقنية بلوكشين.

وصفت Encrypgen ، وهي شركة ناشئة أجرت مؤخرًا حملة إيكو ، المشكلات القائمة والعلاقة بين الخصوصية والدعاية باستخدام بلوكشينز .[18] ومع ذلك ، تفتقر ورقة العمل المرتبطة بهذا المشروع إلى وصف للتنفيذ الفني الذي سيحل مشكلة الخصوصية والتوفر.

هناك مشروع آخر في هذا المجال هو مشورع [19] (DNAbits) ، الذي قام مؤسسه ، Dror Samuel Brama، بتسجيل براءة اختراع على منهج عام لتخزين البيانات ونقلها باستخدام تقنيات بلوكشين .[20] ومع ذلك ، لم تنفذ هذه الشركة من الناحية الفنية مفهومها خلال السنوات الثلاث الماضية.

## S الحق في امتلاك البيانات الجينومية؟

لا يوجد حاليا أي تعريف تشريعي للحق في امتلاك المعلومات الجينية الخاصة به. في بعض البلدان المتقدمة ، بما في ذلك الولايات المتحدة وألمانيا والنمسا ، لا يتمتع المواطنون بالحق في الوصول إلى بياناتهم الوراثية وحيازتها في سياق تفسيرها [21]. هناك حاجة إلى وكيل ، يمثله طبيب أو مركز طبي مع الحق في تقديم مثل هذه المعلومات. يتم استخدام هذا المسار من قبل الشركات Genways Pathway في الولايات المتحدة الأمريكية و CeGaT في ألمانيا.(http://www.cegat.de/en/)

لإجراء تحليل وراثي ، هناك حاجة إلى نصيحة الطبيب الذي يمكن أن يكون مزودًا للاختبار الجيني ، وهذا الطبيب فقط هو الذي له الحق في تفسير المعلومات التي يقدمها التحليل الجيني.

في الولايات المتحدة الأمريكية ، هناك مزوّدون للخدمات في مجال "علم الوراثة من أجل المتعة" ، مثل شركتي andMe23 و Ancesty.com ، اللتان تبيعان الاختبارات الجينية مباشرة إلى العميل النهائي ، ولكن هذه الشركات لا يمكنها سوى تقديم معلومات عن الأصل العرقي وبعضها الخصائص الصحية (على سبيل المثال ، الخصائص الرياضية) وتفتقر إلى الإذن بتقديم معظم المعلومات القيمة طبيا. هذه القيود ، التي يفرضها المنظمون مثل FDA ، لا تعوق قدرة andMe23على بيع الوصول إلى البيانات الوراثية لشركات الأدوية الكبيرة. ومن المعروف أن بعض هذه الصفقات معروفة: أبرمت صفقة مع Genentech (قسم فرعي من شركة الأدوية العملاقة روش) مقابل 60 مليون دولار [22] لإجراء دراسة لمرض باركنسون ، وتم التوصل إلى اتفاق مع شركة أدوية كبيرة أخرى هي شركة فايزر ، لدراسة أمراض الأمعاء الالتهابية (على سبيل المثال ، مرض كرون) [23]. كما تدعي بعض التقارير أن 23 andMe أجرت مفاوضات مع شركة Novartis حول دراسة مرض الزهايمر [24]

وبالتالي ، فإننا نمنح الشركات الكبرى حاليًا الحق في إدارة معلوماتنا الجينية وتخزينها والاستفادة منها. الشركات ، وراء غطاء النوايا الحسنة ، تحتكر البيانات الضخمة الجينومية ، ولا يمكننا التنبؤ بكيفية تأثير هذا الاحتكار على أسعار الأدوية المستقبلية والاكتشافات الطبية..

## الحق في الوصول إلى المعلومات الجينومية

يجب مناقشة قضية أخلاقية أخرى. أعلاه ، لاحظنا أن انتهاكات الخصوصية والوصول إلى المعلومات الجينومية يمكن أن تستخدم بشكل غير قانوني ، على سبيل المثال ، من قبل صاحب العمل. يمكن طرد أو رفض أحد من ترقية الوظيفة على أساس المعلومات الجينية. بالنسبة لأولئك الذين لديهم وظائف تتعلق بسلامة الأشخاص والأنظمة ، مثل سائق الشاحنات أو الحافلات ، أو الطيارين ، أو مشغلي محطات الطاقة الذرية ، أو الأشخاص ذوي المهن المماثلة الأخرى ، فإن الحالة الصحية مهمة للغاية ، ويمكن للمعلومات الجينومية أن تمنع وقوع حادث أو حتى كارثة. في بعض المهن ، يمكن أن يهدد خطر محتمل العامل وليس المتفرجين ، مثل عامل مناجم الفحم أو الماس الذي يعاني من مشاكل في الرئة. في هذه الحالات ، من الضروري إجراء مناقشة تضم المهنيين والخبراء بالإضافة إلى عامة الجمهور لوضع معايير قانونية تنظم استخدام المعلومات الجينومية من قبل أصحاب العمل.

# الجزء 2

# المفهوم

## مشروع Zenome

### عرض فلسفي

يظل الوعي العام بالطب الجينومي منخفضًا جدًا في الدول المتقدمة وحتى أسوأ في الدول النامية. وهذا يعني أن الناس ، بشكل عام ، ليس لديهم سوى القليل من الفهم للفوائد المحتملة لعلم الجينوم ، فضلاً عن المخاطر المحتملة المرتبطة به. وفي العديد من البلدان ، أدى إلى تطوير إجراءات مؤسسية معقدة للغاية لحماية المعلومات الوراثية من سوء الاستخدام المحتمل الذي يعيق التقدم العلمي من جهة أخرى.

تعمل منصة زينوم على زيادة الوعي حول الطب الجيني ، بحيث يمكن للمستخدمين اتخاذ قرارات واعية فيما يتعلق ببياناتهم. لضمان ذلك ، تستند منصة زينوم إلى المبادئ الأساسية التالية:

**الملكية الفردية** للمعلومات الجينومية الشخصية كل مشارك لديه كل الحقوق للبيانات الجينومية الشخصية.

**حرية الاختيار** يقرر كل مشارك كيف يجب استخدام المعلومات الجينية الفردية. قد يقرر المرء ما إذا كان يشارك في الأبحاث العلمية / الطبية أم لا.

**الحق في المشاركة** يجوز للمشارك منح حق الوصول إلى المعلومات الجينية إلى طرف ثالث بطريقة تقيد نسخ البيانات.

**الخصوصية** يجعل تشفير البيانات الخاصة من المستحيل الوصول إلى المعلومات الجينية الفردية دون إذن صريح من المستخدم.

**تخزين البيانات الموزعة** توفر بنية قاعدة البيانات الموزعة توفرًا عاليًا وتحملاً للخطأ من خلال القدرة على النسخ والتكرار.

**معالجة البيانات الموزعة** تتم معالجة البيانات على العديد من عقد الشبكة في نفس الوقت. يمكن لأي مستخدم أن يصبح عقدة من خلال توفير مساحة القرص ووقت وحدة المعالجة المركزية للشبكة.

قابلية التطوير تتيح بنية المنصة إمكانية كبيرة للتطوير والمرونة في النظام.

### نظام Zenome الاقتصادي

على منصة Zenome مستخدم[1] يشارك في العديد من أنواع التفاعلات المختلفة في جميع أنحاء النظام. تحدث هذه التفاعلات على مستويات مختلفة من النظام ، ولا تتداخل مع بعضها البعض وتتضمن أنماطًا مختلفة من التفاعلات. وبالتالي ، ينبغي تمثيلها ككيانات متميزة لها أدوار مختلفة.

الأدوار التالية متوفرة على منصةZenome :

(حساب / تخزين) العقدة التي توفر التخزين والطاقة وحدة المعالجة المركزية لمكافأة.

الشخص الذي قام بتحميل بيانات وراثية فردية إلى المنصة وربما يستخدم خدمات وراثية.

محلل مهتم بتحليل المعلومات الجينية على المنصة. قد تمثل: عالم بيانات ، منظمة علمية وما إلى ذلك.

مزود الخدمة الذي ينفذ خدمة جينية للمساحة المستخدمة (قد تكون مدفوعة) على المنصة. في الأساس ، إنها منظمة تستخدم البيانات الجينية كجزء من نشاطها التجاري.

باختصار

**ينخرط كل مستخدم في عدد من التفاعلات من نوع مختلف ، ويتولى أدوارًا مختلفة. البعض منهم ، مثل مقدم الخدمة والتحليل ، يحتاجون إلى معرفة خاصة ، لكن لا علاقة للعقدة والشخص.**

سيتم مناقشة كل دور بالتفصيل في وقت لاحق.

## نظرة عامة على بنية النظام

منصة Zenome عبارة عن تطبيق موزع يتكون من 3 طبقات رئيسية.

[1]User in a broader sense, i.e. a person or software that runs on behalf of that person.

**طبقة الشبكة (والوصول إلى البيانات):** توفر مستوى من التجريد الذي يغطي تفاعل الشبكة ويوفر واجهة لبيئة موزعة على الطبقات العليا.

| DHT Kademlia | بلوكشين | |
|---|---|---|
| منخفض | متوسط | تكلفة تخزين البيانات |
| خاطئة | صحيح | ثبات البيانات |
| متوسط | منخفض | الأداء |
| خاطئة | صحيح | النتيجة الحتمية |

تتكون هذه الطبقة من نظامين موزعين لطبيعة مختلفة تمامًا:

**دفتر الأستاذ الموزع** (على أساس بلوكشين) الذي يسجل المعاملات بين المشاركين بطريقة يمكن التحقق منها ودائم. للوصول إلى برنامج عقدة بلوكشين يتم إدارة عميل الايثريوم بشكل مدمج.

**توزيع شبكة السرعة الموزع** (بناءً على بروتوكول Kademlia) الذي يقوم بتجميع العقد المادية في شبكة تغطية ويمكن تمرير الرسائل بين العقد وتخزين البيانات الموزعة.

ROLE

(الحساب / التخزين) تعمل العقدة في طبقة الشبكة.

تحتوي البرمجيات الوسيطة على إدارة الحساب وواجهة متناسقة لميزات الأمان للطبقة الأساسية وواجهات برمجة التطبيقات عالية المستوى للبرامج التي تعمل على مستوى التطبيق.

ROLE

**مزود الخدمة** يعمل على البرمجيات الوسيطة. تُمكِّن واجهة برمجة تطبيقات المنصة للخدمات الخارجية الجهات الخارجية من بدء الخدمات الجينية على المنصة.

**تطبيق طبقة Zenome** يتميز واجهة المستخدم النهائي المتقدمة التي تترجم إجراءات المستخدم إلى الوسيطة بطريقة متسقة للغاية. إن الواجهة قابلة للإمتداد بالتصميم بحيث يمكن تشغيل الخدمات الوراثية أصلاً على مجموعة البرامج.

## سوق الخدمات الوراثية

يتم تطوير سوق الخدمات الوراثية حاليًا في المجالات التالية:

1. اعتماد البحوث والتكنولوجيا في السوق.
2. توفير خدمات التشخيص الجيني.
3. التصديق الحكومي على التقنيات الوراثية.
4. تطوير إطار قانوني. على وجه الخصوص ، التدابير التشريعية لحماية المعلومات الجينية ؛

NOTE

هيكل السوق معقد جدا. في الواقع ، يتطور بعض اللاعبين في عدة اتجاهات ليجدوا مكانهم في السوق في مواجهة احتياجات المستهلكين المتزايدة بسرعة.

يتم تمثيل اللاعبين التاليين في سوق الخدمات الوراثية:

تعمل الشركات العلمية على اكتشاف واعتماد تقنيات جديدة في السوق. وتشمل هذه:

- الشركات الصيدلانية وشركات التكنولوجيا الحيوية والتشخيص ، مثل شركة Pfizer وشركةMyriad
- الشركات التي تقوم بتطوير وبيع جميع المواد الكيميائية التكميلية الضرورية (مثل تقنيات الحياة)

تشارك شركات تكنولوجيا المعلومات الحيوية في اختراع وتطوير أساليب معالجة البيانات الحاسوبية. لا يزال اللاعبون في هذا القطاع يكافحون للتعامل مع أنواع وأحجام البيانات التي تم الحصول عليها.

تلعب المراكز العلمية والطبية دوراً رائداً في توفير وتطوير خدمات التشخيص الجيني.

توفر المختبرات التجارية خدمات تشخيص وراثية سريعة وفعالة ورخيصة نسبيا. وهم يملكون قدرات مالية وموارد كبيرة.

وتزيد شركات التشخيص الوراثي المباشر إلى المستهلك من اهتمام السكان بالتشخيص الجيني. هذا الجزء صغير الحجم حاليًا ، ولكن في المستقبل ، يمكن أن ينمو إلى جزء من السوق الرائد ويمكن اعتماده في الممارسة السريرية.

الجدول 5: مقارنة مع المنتجات المماثلة في السوق

| | We | GeneCoin | Encrypgen | 23andMe | Pathway Genomics | Snpedia (Promethease) | Human longevity |
|---|---|---|---|---|---|---|---|
| لامركزية | ✓ | ✓ | ✓ | – | – | – | – |
| مناسبة للكائنات غير البشرية | ✓ | ✓ | ✓ | – | – | – | – |
| العميل هو مالك بياناته | ✓ | ✓ | ✓ | – | – | ✓ | – |
| إمكانية تحميل البيانات الخاصة بك | ✓ | ✓ | ✓ | – | – | ✓ | – |
| البيانات غير الخاصة المفتوحة | ✓ | ✓ | – | – | – | – | ✓ |
| يقوم بتحليل البيانات الخاصة به | ✓ | ✓ | – | ✓ | ✓ | ✓ | ✓ |
| يقدم تقريرا للعملاء | ✓ | – | – | ✓ | ✓ | ✓ | – |
| يستخدم الذكاء الاصطناعي والتعلم الآلي | ✓ | – | – | – | – | – | ✓ |
| تقاسم دون نقل البيانات الضخمة | ✓ | – | ✓ | – | – | – | – |
| كسب باستخدام البيانات الخاصة بك | ✓ | – | – | – | – | – | – |
| مفتوح للعلماء منصة لأدوات أخرى | ✓ | – | – | – | – | ✓ | – |

# أدوار في منصة Zenome

**من وجهة نظر نقاط الموارد**

النظام من منظور العقدة الحسابية

العقدة - أحد المشاركين يوفر موارد جهاز الكمبيوتر الخاص به (التخزين ووقت المعالج) لغرض التخزين الموزع ومعالجة البيانات الوراثية لمكافأة في توكنزZNA .

لتصبح عقدة حسابية ، يقوم المستخدم بتشغيل برنامج Zenome على جهاز الكمبيوتر الخاص به ويقوم بتنشيط دور عقدة باستخدام واجهة مستخدم رسومية. يجب تشغيل البرنامج باستمرار في الخلفية.

ملحوظة

يتم أيضًا توفير إصدار خط أوامر لبرنامج Zenome والذي يمكن استخدامه تحديدًا لبدء العقدة الحسابية.

يمكن تكوين سياسة تخصيص موارد النظام وإدارة المهام بشكل مرن بواسطة مالك العقدة:

- حجم التخزين الأقصى المسموح باستخدامه
- سياسة استخدام الموارد الحاسوبية:

**عدد استخدام وحدة المعالجة المركزية / وحدة معالجة الرسومات الثابتة** من النوى والتحميل الأقصى (في المائة) لكل نواة.

**يتم تخصيص موارد استخدام CPU / GPU الديناميكية** للبرنامج بطريقة لا تتداخل مع تطبيقات المستخدم.

- اختر العملية الحاسوبية (بواسطة id) التي ستكون لها أولوية تنفيذية.
- أغلق العقدة مؤقتًا: اطلب من الشبكة نقل البيانات غير المتوفرة في أماكن أخرى إلى عقد أخرى. انتظر حتى اكتمال نقل البيانات وقطع الاتصال.
- قم بإيقاف تشغيل العقدة تمامًا: اطلب من الشبكة نقل جميع البيانات إلى عقد أخرى ، وانتظر حتى اكتمال نقل البيانات وحذف البيانات.

**الشخص / العضو**

الشخص - الشخص الذي يرغب في تقديم معلوماته الوراثية إلى نظام Zenome من أجل تحقيق الربح من بيع بياناته الشخصية أو استخدام الخدمات القائمة على الوراثة المتوفرة على المنصة.

يقوم مستخدم بتثبيت برنامج Zenome للعمل مع المعلومات الجينية الشخصية. باستخدام واجهة المستخدم الرسومية ، يستطيع المستخدم ما يلي:

•إنشاء حساب شخصي

- إدارة التوكنز: الاستلام ، النقل ، الاستخدام ، الدفع مقابل تخزين البيانات ، الإنفاق على الخدمات الوراثية المدفوعة.
- تحميل البيانات الوراثية (غالباً ما يتم الكشف عن تنسيق الملف تلقائياً)
- إدارة البيانات الشخصية: تقديم البيانات الشخصية ، وملء الاستبيان ، وعرض قائمة من الاستبيانات الأكثر شعبية.
- العمل مع العروض المستهدفة باستخدام النظام الفرعي اقتراحات آمنة.
- استخدام الخدمات الوراثية وتكوين مستوى الخصوصية لكل خدمة. بشكل فردي.

**عند تحميل بيانات الوراثة ، يمكن اختيار مستوى الخصوصية بين:**

**الخصوصية الكاملة** في هذه الحالة ، يتم تخزين البيانات في شكل مشفر ، ويتم تحميل السعر الكامل لتخزين هذه البيانات.

**الخصوصية القياسية** يتم تخزين البيانات الوراثية كأجزاء مما يجعل من المستحيل تحديد المستخدم. يتم تخزين كل جزء في النظام بشكل علني. المعلومات المتعلقة بمطابقة الأجزاء إلى أرقام تعريف المستخدم خاصة. في هذه الحالة ، يكون التخزين أرخص بسبب الإعانات.

يتم تخزين بيانات الوصول العام بشكل علني. التخزين مجاني بسبب الإعانات.

NOTE

انتباه! على الرغم من عدم وجود قيود فنية ضد زيادة مستوى الخصوصية ، تتأثر البيانات المستقبلية فقط. لن تكون البيانات التي تم نشرها عامًا خاصة مرة أخرى أبدًا. ضع في اعتبارك ذلك عند تحميل البيانات لأول مرة.

## مستهلك البيانات

**s** مستهلك البيانات - عالم أو شركة تجارية أو منظمة علمية أو مشارك آخر في المنصة ، يهتم بتحليل البيانات الجينية باستخدام قدرات المنصة. يستطيع مستهلك البيانات تقديم طلبات للمستخدمين وتعيين استرداد الأموال التي سيتم دفعها للمستخدمين المستجيبين

ملاحظة

ملاحظة: توجد بعض القيود على الاستعلامات المصممة لمنع إزالة إخفاء هوية مستخدم النظام وإساءة استخدام المنصة الأخرى. هذه القيود تعتمد أيضا على تصنيف المستهلك في النظام.

## مزود الخدمة

**مزود الخدمة** - منظمة تستخدم البيانات الجينية في أعمالها ، وبالتالي تنفذ خدمة مستخدم على المنصة

ملاحظة

يمكن للمستخدم اختيار نفسه أو البيانات التي يريد مشاركتها مع مزود الخدمة. سيتم إشعار المستخدم إذا كان من الممكن استخدام البيانات المطلوبة لتحديد هويته.

ملاحظة

على سبيل المثال ، لا يستطيع مزود الخدمة رؤية أكثر من 70٪ من قائمة الطفرات في شكل واضح باستخدام الاستعلام المباشر أو للحصول على المعلومات التي تطابق البيانات الأولية (مثل ملفاتfastq )مع استبيان المستخدم

# البيانات الجينومية

## أنواع بيانات omics(الجينوم)

في إطار هذا المفهوم ، يمكن تمييز 3 أنواع من البيانات *omics* (الجينومية) تبعاً للدفع وقيمة المعلومات:

- البيانات المفتوحة ، وهي غير قيمة لأصحابها ، ولكنها مهمة للعلماء.
مثال *Genome* :من سلالة البكتيريا البكتيريةHelicobacter .
- البيانات المفتوحة ، التي تُعد قيمة لكل من مالكها وللاتحاد.
مثال *Genomes* :لغالبية المشاركين في الشبكة.
- البيانات المقيدة المخزنة ببساطة داخل الشبكة.
للمشاريع المقيدة لمختلف المؤسسات التجارية والعامة.

## معالجة البيانات الوراثية

يتكون عادة التعامل مع البيانات الجينية NGS (ومعظم مجموعات البيانات omics الأخرى) من خطوتين مستقلتين:

1. المعالجة الأولية للبيانات "الخام."
2. تحليل مخصص للمتسلسلات الجينية لوضع توصيات شخصية أو في إطار البحث.

NOTE

**جينوم مرجعي** - هو مجموعة بيانات الحمض النووي الرقمية مجمعة كمثال تمثيلي لمجموعة من الجينات**.**

تتضمن المعالجة الأولية للبيانات الجينومية في NGS الخطوات التالية:

1. رابط انسجام NGS مع الجينوم المرجعي.
2. البحث عن الطفرات والاختلافات الأخرى من الجينوم المرجعي ، وحفظ قائمتهم بتنسيقgVCF ..

NOTE

ملاحظة: البروتوكول هو نفسه بالنسبة لبيانات الكائنات غير البشرية. بالطبع ، في هذه الحالة ، هناك حاجة إلى جينوم مرجعي مناسب.

إذا كانت مجموعة البيانات الشخصية تمثل نتيجة تقنية التنميط الجيني microarray (نوع الملف andMe23)، فيمكن أيضًا تحميلها إلى المنصة كـ gVCF نظرًا لأن صيغ البيانات الخاصة بأنواع الملفات هذه متشابهة.

## مشكلة البيانات الوهمية

إذا تم تحميل البيانات الجينية للكائنات الأخرى (غير البشرية) بدلاً من الجينوم الصحيح (عن طريق الخطأ أو بشكل متعمد) ، فسيتم الكشف عن ذلك أثناء المعالجة المسبقة للبيانات الخام وسيتم إبلاغ المستخدم. إذا قام المستخدم عن قصد بتحميل بيانات وراثية مشوهة (مزيفة) إلى النظام ، يمكن اكتشاف ذلك باستخدام العديد من طرق التحقق المعروفة قبل حفظها في التخزين.

NOTE

الإقناع الاقتصادي للامتناع عن تحميل البيانات المزيفة هو أن الدفع مقابل تخزين البيانات يجب أن يكون مقدما لسنة كاملة.

## مشكلة تحديد المستخدم على أساس البيانات الجينومية

إن الوصول المفتوح إلى المعلومات الجينية يثير مشكلة تحديد المستخدمين من خلال بياناتهم الجينية وغيرها. إذا قرر أحد المستخدمين عدم الكشف عن معلوماته الجينية بشكل كامل ، فيجب اتخاذ التدابير المناسبة في كل خطوة من خطوات معالجة البيانات وتخزينها. لحل مشكلة تحديد هوية المستخدم ، يجب أن يتم تصميم التفاعل بطريقة بحيث لا يمكن لأي عقدة تحديد ملكية المادة الوراثية من قبل فرد معين ، أو حتى المدينة التي يعيش فيها هذا الفرد.

NOTE

تشكل الاختلافات بين المقيمين في مدينة واحدة حوالي 0.01٪ من التسلسل.

في كل مرحلة ، يتم تحقيق الهدف أعلاه بطرق مختلفة:

1. في مرحلة المعالجة المسبقة - بتقسيم الملف المصدر إلى أجزاء بحيث يكون متوسط التغطية أقل من عتبة الثقة (6 نسخ)
2. في مرحلة التخزين - حساب تخزين مجزأ حسب الطول.

## تخزين البيانات الجينومية

توجد البيانات الجينومية في شبكة موزعة على أساس بروتوكولDHT Kademlia .يتلقى المشاركون الذين يوفرون الموارد لتشغيل هذه الشبكة (راجع التفاصيل المتعلقة بدور العقدة) الدفعات الخاصة بها في توكنزZNA .من أجل الحصول على السداد ، يجب عليهم أن يثبتوا للشبكة أنهم يقومون بالفعل بتخزين هذه البيانات. يعتمد إجراء هذا التدقيق على استخدام بلوكشين .يتم استخدام التشفير عند الضرورة.

NOTE

ملاحظة: سيتم أيضًا تنفيذ التكامل مع عقد *Storj* و*FileCoin* .

كما سبق ذكره ، يمكن أن تكون بيانات خام ويتم معالجتها.

| نوع البيانات | خام | معالَجة |
|---|---|---|
| الميزات | | |
| التنسيق | fastq / bam | (lists of mutations) gvcf / vcf + bed / 23me(txt) |
| الحجم, Gb | 50 | 2 |
| القيمة | لتحسين تقنية التسلسل والمعالجة (لسوق تطوير المعدات) | لإجراء البحوث ، فضلا عن تقديم تقرير. |

| شروط التخزين | | |
|---|---|---|
| عدد النسخ | أقل شي ثلاث عقد مستقلة | أقل شيئ 5 عقد مستقلة |

يتم تخزين البيانات الجينومية مقسمة إلى أجزاء بطريقة لا يسمح بها طول الجزئ بالتعرف بشكل لا لبس فيه على أن هذا الجزء ينتمي إلى فرد معين.

NOTE

المعلومات المتعلقة بأجزاء الجينوم التي تشكل جينوم المستخدم هي أيضًا معلومات خاصة ولا يمكن الحصول عليها إلا بإذن المستخدم.

## الملفات الشخصية

ملء الاستبيانات الشخصية يزيد بشكل كبير من تطبيق البيانات الجينومية. المستخدمين يملئوا الاستبيان باستخدام واجهة رسومية.

NOTE

إذا أصبحت بعض الاستبيانات شائعة ، فإن التطبيق يطالب المستخدم بملئها. يجوز لكل محلل إنشاء استبيان خاص به ووضعه في المنصة.

## مواصفات الاستبيان

يمكن لعدد من الاستبيانات أن تكون ضخمة ؛ لذلك ، فمن الضروري إدخال مفهوم مواصفات الاستبيان.

**مواصفات الاستبيان** - هو وصف كامل لجميع مجالات الاستبيان والقيم المسموح بها لهذه الحقول.

NOTE

بشكل رسمي ، تحتوي مواصفات الاستبيان على مرجع للمؤلف ، وصف ومجموعة من السجلات مرتبة ، كل منها يتوافق مع حقل واحد من الاستبيان.

قد تحتوي الحقول على عدة أنواع:

الحقل الرقمي قيمة الحقل عبارة عن عدد صحيح.

اختيار متعدد إن قيمة الحقل هي عدد من الإجابات.

NOTE

سيتم وضع الإجابات على هذه الأنواع من الأسئلة في وضع مفتوح (بدون أي إشارة إلى المستخدم) ، لأنها لا تهدد الخصوصية.

**حقل السلسلة** قيمة الحقل عبارة عن سلسلة. إنه حقل خاص ، لأنه يسمح بتهديد هوية المستخدم استنادًا إلى إجابة محددة.

**كتلة خاصة** تتيح عمل أي مجموعة من الحقول خاصة ، بغض النظر عن نوعها الحقيقية.

## استعلامات للنظام.

ستكون البيانات الإحصائية المتعلقة ببعض الحالات الجينومية مفتوحة للجمهور إذا لم يقرر المالك تشفيرها. بالإضافة إلى ذلك ، فإن المعلومات المتعلقة بالإجابات المتاحة على الاستبيانات مفتوحة أيضًا. لذلك ، الجميع يعرف ، على سبيل المثال ، عدد مستخدمي الشبكة الذين أعمارهم 25 سنة أو وجود طفرة rs6025 (عامل تجلطV )

تجعل بنية النظام من المستحيل استخراج قاعدة البيانات الكاملة:

- أثناء إنشاء استفسارات ترابطية ، لا يحصل العميل على بيانات أولية.
- تتضمن الرسوم الأساسية عددًا محدودًا من الطلبات يوميًا. إن رسوم الاستعلامات الإضافية المضمنة في الحصة تنمو باطراد خلال يوم واحد.
- إذا كانت نتيجة الاستعلام التجميعي تحتوي على أقل من 100 مستخدم ، فلن يتم تقديم النتيجة إلى العميل.

إذا قام مستخدم ما بتشفير بياناته بالكامل ، فإنه يقرر بنفسه بياناته الشخصية ، بما في ذلك الأجزاء التي يتكون منها جينومه ، والتي يُسمح بنقلها. لن يتمكن أي محلل من معرفة أنواع البيانات المشفرة.

**نقل آمن للبيانات الشخصية**

يجب أن تمتلك عملية نقل البيانات الشخصية بين المشاركين في النظام الخصائص التالية:

1. يجب أن تكون البيانات الكاملة التي يتم إرسالها متاحة فقط للمشتري والبائع.
2. يجب أن يتم إرسال الرموز فقط إذا تم نقل البيانات بنجاح.
3. يجب تحديد ومحاولة بيع بيانات غير صحيحة.
4. ينبغي الكشف عن محاولة اتهام البائع زورا عن عمد في بيع بيانات خاطئة.
5. لا ينبغي الوثوق بنقل البيانات إلى طرف ثالث.

ملاحظة

سيتم استخدام تقنية بلوكشين لنقل البيانات بأمان. ومع ذلك ، ينبغي اعتبار أن تخزين (ونقل) كميات كبيرة من المعلومات إلى بلوكشين هو استهلاك للموارد. وبالتالي:

1. يُسمح بنقل كمية صغيرة فقط من البيانات عبر بلوكشين .يمكن نقل البيانات المتبقية من خلال قناة اتصال مشفرة بسيطة.

# نظام التقييم

ستقوم المنصة بإنشاء تصنيف لـ:

- بشكل منفصل لكل جزء وراثي أو كتلة من البيانات الشخصية
- المنظمات
- مقدمي الخدمة

موفرو البيانات (مثل معامل تسلسل الحمض النووي)

ملاحظة

لن يكون هناك تصنيف مستخدم فردي ، حيث إنه يمثل فعليًا مجموع تصنيفاته الجينية ، وهو أمر غير متاح للجمهور. إذا تم تحميل البيانات الوراثية باستخدام حساب المؤسسة ، فسيتم تلقائيًا زيادة التقييم الأولي حسب تصنيف المؤسسة.

العوامل التي تؤثر على تصنيف الأجزاء الجينية

**يزيد التأكيد من المختبر** تصنيف البيانات التي تم تحميلها بما يتناسب مع تصنيف المختبر الذي تم الحصول عليه. قد يمثل التأكيد توقيعًا رقميًا للمختبر ، أو تحميل البيانات بواسطة المختبر نفسه بناءً على طلب العميل.

**التحقق من المقبولية** يتيح التحقق من المعلومات الوراثية للمستخدم باستخدام نماذج إحصائية محددة مسبقًا لترددات متعددة الأشكال والارتباط الوراثي. هذه الوحدة قيد التطوير.

**المشاركة في البحث** يزداد التصنيف مع عدد الأبحاث الناجحة التي تتضمن الجزء. إذا اعتبرت نتيجة استقصاء المعلومات غير قابلة للتصديق ، ينخفض التقييم.

# حالات استخدام

## مستخدم فردي

بالنسبة إلى المستخدمين الفرديين ، هناك فرصة للحصول على معلوماتهم الجينومية وتحويلها إلى مصدر للدخل. إن الجمع بين الجينوم وتفاعله مع البيئة مورد معلومات قيّم. وستسمح منصتنا للمستخدم بإدارة هذا المورد بأمان.

توفر المنصة القدرة على تخزين المعلومات الوراثية ومشاركتها بشكل آمن ، مما يسمح للمستخدمين بتلقي مجموعة متنوعة من الخدمات الوراثية. وفيما يلي بعض الأمثلة على ذلك:

- تقارير وتوصيات بشأن التغذية ، ومخاطر الأمراض ، والتجميل ، والنظام الغذائي ، واللياقة البدنية
- البحث عن الأقارب وتوضيح النسب
- خدمات المواعدة
- اختيار فردي للملابس والأحذية ، وإعداد المناخ المنزلي ، وجهات السفر ومناطق الإقامة
- اختلافات مختلفة في التقارير الجينية لمجموعة من الأفراد ، على سبيل المثال الفرق الرياضية أو مجموعات العمل
- يتأثر كل جانب من جوانب الحياة البشرية تقريبًا بالوراثة ؛ لذا دعنا نرى ما يمكن أن تقدمه الشركات الجديدة باستخدام المنصة

إلى جانب تعدد الخدمات ، يُمنح المستخدم فرصة لتحقيق الربح من تفرده الوراثي من خلال توفير بيانات استبيان للشركات لأغراض البحث. وهكذا تصبح البيانات الفردية مع المعلومات الجينية تناظرية للسلع أو الموارد المعدنية

## الصحة

لا يمكن تخيل الرعاية الصحية الحديثة والطب الشخصي دون استخدام تقنيات الجينوم. ستتيح المنصة للمرضى مشاركة المعلومات الوراثية بشكل آمن في العيادة مع الموظفين الطبيين:

- الجرعة الفردية والتعصب للأدوية (على سبيل المثال ، الجرعات الفردية لوارفارين مضاد للتخثر على أساس الخصائص الجينية)
- النطاقات الشخصية المقبولة من معلمات الجسم البيوكيميائية (على سبيل المثال ، علامةPSA )
- الاستعداد الوراثي للأمراض المختلفة (على سبيل المثال ، ارتفاع مخاطر التنكس البقعي والحاجة إلى مزيد من البحوث والوقاية)
- زرع الأعضاء والتبرع بها. يمكن للمستخدمين بشكل آمن مشاركة المعلومات المتعلقة بنوع مستضدات HLA التي تحدد التوافق بين الأفراد أثناء الزرع. وبالتالي ، سيكون من الممكن إنشاء قاعدة بيانات آمنة من المتبرعين والمتطوعين لإنقاذ الأرواح من خلال الزرع.

## الشركة

هناك نوعان من الشركات ، أولهما يوفر للمستخدمين خدمات تستند إلى بيانات الجينوم ، بينما يهتم الثاني بالحصول على البيانات الجينية من المستخدمين لإجراء أبحاثهم الخاصة.

النوع الأول موضح في منطقة حالات استخدام المستخدم. النوع الثاني يمكن وصفه بأنه شراء بيانات الجينوم للمستخدمين لإجراء أبحاثهم الخاصة ولتحسين خصائص المستهلك للمنتجات ، والاستهداف الجيني للمنتجات والإعلان ، وبعض الأمثلة منها:

- على سبيل المثال ، تخطط شركة أدوية لإطلاق دواء جديد يعمل ضد بروتين سرطان متحولة. قد تجد الشركة مستخدمي النظام ، الذين نجوا من المرض ، للدفع لهم مقابل البيانات الوراثية والحصول على تواتر الطفرات في الجين الذي يرمّز البروتين الذي يستهدفه المادة الفعالة
- تخطط شركة المستهلك لدخول سوق جديدة وتحتاج إلى اختبار كيفية إدراك المستخدمين لمذاق المنتج. من المعروف أن بعض النكهات تسبب الاستياء بين الحاملات من متغير جيني معين من جين مستقبلات الذوقية. ترسل الشركة عبر شبكة عرضاً لدراسة ناقلات هذا النوع الجيني وتلتقط نكهة أخرى ، أو تكتشف تسلسل هذا البديل الجيني في الأسواق المختلفة وبالتالي تستهدف المنتج في الأسواق المختلفة.

## المجتمع العلمي

بالنسبة للمجتمع العلمي ، يفتح النظام إمكانية التخزين والمشاركة وأداء البحث باستخدام بيانات جينية مختلفة. نظرًا لأن المنصة لا تقتصر على العمل مع الجينوم البشري ، يمكن استخدامه للتخزين الآمن للبيانات الجينومية ومعالجتها ، على سبيل المثال ، ذات الصلة بالزراعة (النباتات والحيوانات والكائنات الحية الدقيقة)

بشكل عام ، يؤدي وجود النظم البيئية إلى إثراء المجتمع العلمي من خلال الوصول إلى البيانات السكانية العامة ، حتى بدون الرجوع إلى الاستبيانات الفردية. بالإضافة إلى ذلك ، وبموافقة المستخدمين ، يمكنهم أيضًا أن يصبحوا جزءًا من البحث العلمي.

كما يوفر النظام الأساسي القدرة الحسابية الموزعة ، والتي تتيح إمكانية الوصول إلى معالجة كميات كبيرة من البيانات الجينية (على غرار AWS التي تم تكييفها للعمل مع البيانات الجينية)

# الجزء 3

# الجزء التقني

## الأهداف الموزعة

**مفهوم النظام الفرعي الموزع**

**النظام الفرعي (الموزع)** - عبارة عن مجموعة من بعض عناصر النظام (النظام الأساسي) والعمليات التي يمكن تمثيلها من المنظور الموجه للكائن ككيان له هوية مميزة ويظهر سلوكًا واضحًا واضحًا خارجيًا.
لإعطاء خاصية تفصيلية لنظام فرعي ، يجب وصف الجوانب التالية له:

1. **الهيكلية**: العناصر والعمليات التي تشكل النظام.
2. **السلوك الخارجي**: تفاعل النظام الفرعي ككل مع المشاركين الآخرين. خاصه:
   - **الواجهات**: مجموعة من الاستعلامات الممكنة لنظام فرعي ككل.
   - **الإجراءات**: الإجراءات التي اتخذها نظام فرعي بشأن المشاركين الآخرين في النظام.
3. **الدولة الداخلية**: الحالة الداخلية الإجمالية للنظام الفرعي.

قد يتم تمثيل النظام الفرعي على أنه quasiobject ، والذي يكون المشاركون الآخرون قادرين على التفاعل معه.

NOTE

تعني البادئة «quasi-» أن التفاعل الحقيقي يحدث مع عناصر النظام الفرعي ، والتي بدورها لها تفاعلات معقدة مع بعضها البعض كجزء من هذا التفاعل الحقيقي ، بحيث يمكن اعتبار كل هذه الأشياء تفاعلًا

مع بعض الكائنات التجميعية.

ملاحظة

أدناه سنقوم بإسقاط الفوارق بين النظام الفرعي وتمثيله باعتباره*quasiobject* .

**يمكن تمثيل التفاعلات مع النظام الفرعي كتسلسل لعدد صغير نسبيًا من العمليات الأساسية:**

- واجهة النظام الفرعي ، أي إجراءات المشاركين الآخرين فيما يتعلق بالنظام الفرعي.
- الإجراءات المتعلقة بالمشاركين الآخرين في المنصة.
- العمليات الداخلية التي تغير الحالة الداخلية للنظام.

**العمليات الداخلية**

**العمليات الداخلية بالمعنى الواسع** - تمثل مجموعة من العمليات الداخلية لكل عنصر فرعي والتفاعلات بين هذه العناصر.

**العمليات الداخلية (بالمعنى الضيق)** - تمثل العمليات داخل النظام الفرعي ككل والتي تغير حالتها الداخلية. يحتوي الوصف الكامل للعمليات الداخلية على كافة سلوك النظام الفرعي باستثناء مشكلات التنفيذ الخاص به.

## d النظم الفرعية الموزعة الرئيسية للمنصة

تحتوي المنصة على مستويات المؤسسة التالية:

- طبقة النظام الأساسية (البنية التحتية الحرجة)
- تخزين البيانات ومعالجتها المستوى
- التفاعلات عالية المستوى

يتم تضمين الأنظمة الفرعية التالية داخل المنصة:

1. طبقة النظام الأساسية

النظام الفرعي البيني منخفض المستوى لتبادل الرسائل بين عقد الشبكة. كما يتيح إنشاء جداول التجزئة الموزعة.

**التراخيص** البنية التحتية للحسابات وإدارة الوصول إلى المعلومات الخاصة.

2. تخزين البيانات ومعالجتها المستوى

مستوى المخازن التجريدية للوصول إلى نظام الملفات الموزعة.

تجهيز البنية التحتية للحسابات الموزعة.

3. التفاعلات عالية المستوى

الاستعلامات الآمنة نظام فرعي يتيح تكوين العروض لشراء البيانات الجينومية التي يمكن عرضها للمستخدمين المناسبين فقط.

عمليات البيانات المفتوحة يوفر أدوات للتشغيل مع البيانات المفتوحة (وليس الخاصة)

توفر المنصة للخدمات الخارجية واجهة برمجة التطبيقات لتوصيل الخدمات الخارجية التي تتم إدارتها مباشرة إلى المنصة. يرتب نقل البيانات بأمان باستخدام بروتوكولات الويب التقليدية.

## طبقة التشغيل المتداخل ذات المستوى المنخفض

### شبكة P2P الموزعة

**أساس المنصة** هو شبكة P2P على أساس بروتوكولKademlia .تتضمن الشبكة أجهزة الكمبيوتر الخاصة بالمستخدمين الذين قاموا بتثبيت برنامجZenome .

العقدة - هي عقدة من شبكة P2P الموزعة التي تمثل كمبيوتر المستخدم مع برنامج Zenome المثبت عليه.

NOTE

بالتالي ، يتم إنشاء شبكة تراكب بين الأجهزة المشاركة في الشبكة. وهي تمثل شبكة افتراضية يتم فيها تخصيص "NodeId" لكل مشارك ، ولا علاقة له بعنوان IP الحقيقي للجهاز. تخزن كل عقدة قائمة بالعقد "القريبة" ، حيث يتم حساب المسافة بين العقد استنادًا إلى العقد "NodeId" ولا ترتبط بالمسافة المشتركة.

تقوم العقد بتخزين البيانات باستخدام جداول التجزئة الموزعة.

### خصائص تنفيذ الشبكة الموزعة

يتم استخدام مواصفات بروتوكول معدلة. الاختلافات الرئيسية هي ما يلي:

يمكن للعقد تبادل الرسائل العشوائية بين بعضها البعض. تستطيع العقدة نقل رسالة إلى عقدة أخرى تعرف فيها "NodeId" فقط.

- قد توجد عدة جداول تجزئة داخل الشبكة. يتم تحديد جدول التجزئة بواسطة مفتاح السلسلة.
- يمكن تعيين سياسات مختلفة من تخزين القيمة وحذفها لجداول مختلفة.
- يتم تشفير البيانات المنقولة من عقدة إلى أخرى (انظر أدناه)

## تبادل الرسائل في بيئة موزعة

بالتالي ، يمكن للمشاركين في شبكة نظير إلى نظير تبادل الرسائل التالية:

| | |
|---|---|
| PING | تحقق من أن العقدة لا تزال على قيد الحياة |
| STORE(T,K,V) | في العقدة التي تتلقى الرسالة "T" في الجدول "K" بواسطة المفتاح "V" قم بتخزين القيمة |
| FIND_NODE(N) | بين العقد التي "N" سترسل العقدة التي تتلقى الرسالة البيانات المتعلقة بالعقد الأقرب إلى العقدة تعرفها. |
| FIND_VALUE(T,K) | ، ثم أرسل البيانات المتعلقة "V" في عقدة استلام ، فأرسل القيمة "(K ،V)" إذا تم تخزين الزوج بالعقد المعروفة "الأقرب" إلى الملف. |
| SEND(M,N,D?) | يتم استخدام "N". إلى العقدة "D" التي يمكن أن تحتوي أيضًا على البيانات "M" يرسل الرسالة لتحديد موقع العقدة "FIND_NODE". |

NOTE

\ملاحظة: مستوى التفاعل هذا هو مستوى نقل المنصة.

# بلوكشين

## المعلومات الأساسية

تستخدم المنصة بلوكشين إيثريوم التي تمثل جهاز ظاهري لامركزي واحد(EVM) .يمكن تنفيذ منطق النظام المطلوب باستخدام العقود الذكية.

QUOTE

العقد عبارة عن مجموعة من الأكواد (وظائفه) والبيانات (حالته) التي توجد في عنوان محدد في بلوكشين إيثريوم.

-مقدمة في العقود الذكية (دليل قوتها)

بالتالي ، فإن العقود الذكية قادرة على تخزين البيانات. بالنسبة للبيانات الموجودة في بلوكشين ، لا يزال نظام نوع الملخص الموضح أعلاه صالحًا (ولكن مع بعض القيود)

## d نظام فرعي للعمل مع بلوكشين

ملاحظة

انتباه: قد تختلف تفاصيل التنفيذ اعتمادًا على منصة المستخدم. الوصف أدناه مناسب لمنصة الكمبيوتر الشخصي.

لتوفير الوصول إلى بلوكشين ، يتضمن برنامج العقدة التنفيذ الكامل لعقدة الإيثريوم الذي يوفر إمكانية الوصول باستخدام "JSON-RPC 2.0".

يتم تخزين المفاتيح السرية في التخزين المشفر. تتم مطالبة المستخدم بتعيين كلمة المرور عند تشغيل التطبيق الأول.

ملاحظة

على الرغم من أنه من المجدي تقنياً استخدام حساب موجود ، فمن المستحسن إنشاء حساب جديد* .

تسمح واجهة التطبيق بـ:

1. إنشاء حساب جديد
2. استيراد حساب موجود
3. قم بإعداد نسخة احتياطية من التخزين الخاص

ملاحظة

انتباه: من المستحسن إعداد النسخ الاحتياطي السحابي لأن هذا سيسمح بالاحتفاظ بحق الوصول إلى الحساب حتى في حالة عدم توفر جهاز الكمبيوتر فعليًا.

ملاحظة

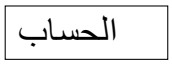
يسمح البرنامج بالحصول على خط أوامر عقدة إيثريوم .هذه الوظيفة مخصصة في المقام الأول من أجل التصحيح. لا ينصح باستخدام خط الأوامر إذا لم تفهم أغراضه.

### التوكنز الداخلية

يتم توفير التفاعلات الاقتصادية داخل النظام باستخدام توكنز ZNA الداخلية. وهي تمثل توكنز إيثريوم صالحة ويمكن شراؤها وبيعها في أسواق التداول.

### مفهوم الحساب

يسمح **الحساب** في المنصة للمستخدم بالتفاعل مع النظام عن طريق لعب العديد من الأدوار بشكل متزامن. كل دور في الحساب يتوافق مع عقد ذكي منفصل في بلوكشين**.**




**الشكل 4**: حساب المستخدم الذي قدم جينومه (دور الشخص) إلى المنصة ، الذي يعمل كمعلم معلوماتي حيوي ويدعم عقدتين حاسوبيتين موجودتين في المنزل وفي العمل.

**الدور** - هو سمة من سمات مشارك النظام فيما يتعلق بنوع معين من التفاعلات التي يشارك فيها. قد يلعب كل مشارك عدة أدوار في وقت واحد.

NOTE

النظام الفرعي منفصل هو المسؤول عن توزيع الأدوار.

## شبكة تخزين البيانات الموزعة

### مبادئ تشغيل الشبكة الموزعة

يتم استخدام الشبكة الموزعة القائمة على بروتوكول DHT Kademlia لتخزين بيانات الأنواع الاعتباطية. يتم تمثيل عقد هذه الشبكة من قبل مزودي الموارد المسجلين ، والتي تكسب مكافأة لتخزين البيانات وأداء الحسابات اللازمة. يدفع المشارك الذي قام بتحميل البيانات مكافأة مقابل تخزين البيانات ، ولكن في عدد من الحالات المهمة عمليًا ، يتم دعم هذه الرسوم جزئيًا أو كليًا من قبل النظام. على سبيل المثال ، توجد آلية لدعم تخزين البيانات ذات القيمة العلمية.

يتم تمثيل وحدة التخزين من خلال كتلة بيانات عشوائية ومفتاح فريد يمكن من خلاله الوصول إليه.

NOTE

يقتصر حجم كتلة البيانات لمنع سوء الاستخدام

**دعم الخدمات الخارجية**. كما سيتم تطبيق إمكانية الدمج مع العقد Storj و FileCoin ، وكذلك مع المخازن المدارة مركزيا**.**

## ضمان موثوقية تخزين البيانات

لضمان موثوقية تخزين البيانات ، في ضوء حقيقة أن العقد يمكن أن تتصل بحرية بالشبكة وفصلها ، يتم تخزين البيانات بشكل مستقل على العديد من العقد في وقت واحد. يحدد نوع البيانات المخزنة عدد العقد التي يتم تكرار المعلومات عليها.

الجدول 6: المعلمات النموذجية للبيانات الجينية.

| | خام | معالجة |
|---|---|---|
| التنسيق | fastq/bam | gvcf/vcf + bed/23me(txt) |
| الحجم | 50Gb | 2Gb |
| العقد | ≥ 3 | ≥ 5 |

عندما يتغير عدد العقد ، يتم إعادة توزيع البيانات المخزنة لتلبية متطلبات الحد الأدنى لعدد العقد.

نظرًا لأن كل شخص يريد أن يصبح نقطة اتصال في الشبكة الموزعة ، فليس من الآمن تصديق أن العقدة تخزن الملف بالفعل عندما تدعي ذلك. وبالتالي ، يتم إجراء فحص استرجاع البيانات بشكل دوري وفقًا للبروتوكول المقابل من طبقة الأمان. يتم تقديم البيانات على نتائج التحقق إلى بلوكشين ويمكن أن يكون السبب في إصدار مكافأة أو تغيير تصنيف العقدة.

## ضمان خصوصية تخزين البيانات

تصبح خصوصية تخزين البيانات في مثل هذه الشبكة الموزعة ممكنة بسبب تطبيق التشفير غير المتناظر. يتم تحديد الأسلوب الفعلي للتشفير بواسطة بروتوكول الأمان المطابق.

ملاحظة

تجدر الإشارة إلى أنه لا يمكن تطبيق آلية الدعم لتخزين البيانات المشفرة.

## ميزات محددة لتخزين المعلومات الجينومية

يمكن تحديد انتماء بعض أجزاء الجينوم للفرد بشكل لا لبس فيه إذا كان الجزء كبيرًا إلى حد ما. هذا هو السبب في تخزين الجينوم في الشبكة الموزعة على هيئة أجزاء قصيرة.

يتم إجراء تجزئة الجينوم دائمًا استنادًا إلى الجينوم المرجعي. بالنسبة لكل جينوم مرجعي (على سبيل المثال ، نسخة أخرى من الجينوم المرجعي البشري أو جينوم كائن حي من أنواع مختلفة) ، يتم اختيار التجزئة مرة واحدة فقط ، ويتم تعيين كل جزء بمعرف فريد لهذا الجينوم المرجعي.

# البيانات الشخصية للمستخدم

سجل - هو وحدة الحد الأدنى لتبادل البيانات. ليس من الممكن نقل (على سبيل المثال ، لبيع مكافأة) جزء من المعلومات من سجل. من الممكن إنشاء سجل جديد يحتوي فقط على جزء من البيانات ، ولكن في مثل هذه الحالة سيكون هذا السجل الجديد أقل إثارة للاهتمام بالنسبة للمشترين المحتملين بسبب انخفاض تصنيفه.

يحدد مخطط البيانات المعلومات التي يجب أن يحتويها التنسيق. يعد استخدام النظام أداة مرنة لتوحيد تنسيق تبادل البيانات بين المشاركين في الشبكة. قد يكون معرف مخطط البيانات عبارة عن أي سلسلة تتيح للفرد العثور على وصفه في الشبكة ، على سبيل المثال ، عنوان URL أو عنوان العقد الذكي مع وصف. يجب أن يكون مفهوما أن الهدف الأساسي من مخطط البيانات هو توحيد مفهوم ما يتم تضمينه في البيانات بين المشترين والبائعين لهذه البيانات.

## مواصفات الاستبيان

ملء الاستبيانات الشخصية يزيد من قيمة البيانات الجينومية بشكل كبير. قد يكون هناك عدد هائل من الاستبيانات ، لذا من الضروري إدخال مفهوم مواصفات الاستبيان.

مواصفات الاستبيان - يمثل وصفا كاملا لجميع مجالات الاستبيان والقيم المسموح بها لهذه الحقول.

بشكل رسمي ، تحتوي مواصفات الاستبيان على مرجع المؤلف ، وصف وقائمة السجلات المطلوبة لكل منها حقل واحد من الاستبيان. الحقل الرقمي يجب أن تكون إجابة المستخدم ضمن نطاق القيمة المسموح به [ أ ، ب ] يتم ترميز الإجابة بواسطة عدد صحيح غير موقعة ، حيث يبدأ العدد من الحد الأيسر للنطاق:

NOTE

الإجابة على هذا النوع من الأسئلة متاحة للجمهور لأنها لا تهدد الخصوصية.

**اختيار من متعدد** يتم تشفير إجابة المستخدم على هيئة عدد صحيح غير موقعة يمثل الرقم التسلسلي للإجابة في قائمة. إذا كانت القيمة تساوي الصفر ، فإن المستخدم فضل عدم الإجابة على هذا السؤال.

NOTE

انتباه: الإجابة على هذا النوع من الأسئلة متاحة للجمهور لأنها لا تهدد الخصوصية.

**إجابة السلسلة** يتم تخزين إجابة المستخدم كسلسلة**.**

NOTE

الإجابة على هذا النوع من الأسئلة هي معلومات خاصة.

**الاستبيان المعبأ - عبارة عن بنية بيانات تحتوي على إجابات المستخدم على أسئلة الاستبيان.**

NOTE

هذه البنية مشفرة بالكامل إذا لزم الأمر. يتم تخزينها على الأقل على كمبيوتر المستخدم. يمكن للمستخدم تحميل النسخة الاحتياطية المشفرة إلى بلوكشين إذا أراد ذلك.

يحتوي طلب البيانات الشخصية من المستخدمين على بعض الميزات المميزة:

- لا يستوفي جميع المستخدمين معايير البحث المحدد. للتحقق مما إذا كانت البيانات مناسبة ، يلزم الوصول إلى المعلومات الشخصية.
- لا ينبغي نقل المعلومات الشخصية خارج حاسوب المستخدم بشكل صريح (بشكل مباشر) أو ضمنيًا ( warrant canary ")
- في حالة استيفاء المستخدم للمعايير ، سيحصل على عرض لمشاركة بياناته الشخصية للحصول على مكافأة.

خلاصة1

بما أن نقل المعلومات غير مسموح به ، يجب إجراء الفحص في بيئة منعزلة. الشفرة القابلة للتنفيذ في هذه البيئة لها وصول كامل إلى المعلومات الخاصة ولكن لا يمكنها التفاعل مع أجزاء أخرى من النظام.

خلاصة 2

لا يُسمح بإرسال نتيجة تنفيذ هذه الدالة المعزولة كإجابة على العملاء المحتملين لأنها تهدد الخصوصية أيضًا.

خلاصة 3

من المعقول اختيار مجموعة البيانات الشخصية المقيدة ، والتي تعتبر ضرورية للتوصل إلى نتيجة في خطوة منفصلة للتحقق. سيرى المستخدم على البيانات التي يستند إليها القرار على شاشة عرض سوق التداول.

خلاصة 4

يتم سرد البيانات التي يتم نقلها بشكل صريح على شاشة عرض سوق التداول. يمكن فقط نقل البيانات التي تم طلبها لإجراء التحقق بشكلٍ ضمني ، وستكون أيضًا قائمة هذه البيانات متاحة للمستخدم.

طلب البيانات الشخصية:

- **في الخطوة الأولى** ، يتم طلب الوصول إلى بيانات المستخدم الشخصية.

فقط معرفات البيانات التي تم سردها في هذا الطلب ستكون متاحة لفحص الوظيفة.

- في الخطوة الثانية ، يتم تنفيذ مدونة وظيفة التحقق داخل بيئة منعزلة ، وتمثل نتيجة تنفيذها العرض المصمم فيما يتعلق بتبادل البيانات أو إلغاء العرض. في الحالة الأخيرة لا يتم نقل أي بيانات في أي مكان.

في الحالة الأولى ، يتم إعلام المستخدم عند صياغة عرض التداول ، ثم يقوم هو بفحص العرض ، وقائمة البيانات المستخدمة أثناء التدقيق وقائمة البيانات التي سيتم نقلها إذا وافق المستخدم على العرض.

إذا رفض المستخدم العرض ، فلن يتم نقل أي بيانات.

إذا وافق المستخدم على العرض ، فسيتم نقل البيانات المعروضة صراحةً إلى المستخدم فقط.

# فهرس


*[1]* NHGRI. *Talking Glossary of Genetic Terms. Word «Genome»*.
URL: https://www.genome.gov/glossary/index.cfm?id=90.

[2] Venter J.C., Smith H.O., Adams M.D. "The Sequence of the Human Genome". In: *Clinical Chemistry* 61.9 (2015), pp. 1207–1208.
URL: http://clinchem.aaccjnls.org/content/61/9/1207.long.

[3] Adams M.D. Venter J.C. Smith H.O. "The Sequence of the Human Genome". In: *Science* 291.5507 (2001), pp. 1304–1351. ISSN: 0036-8075.
DOI: 10.1126/science.1058040.
URL: http://science.sciencemag.org/content/291/5507/1304.

[4] Wetterstrand KA. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*.
URL: www.genome.gov/sequencingcostsdata.

[5] Melissa Gymrek et al. "Identifying Personal Genomes by Surname Inference". In: *Science* 339.6117 (2013), pp. 321–324. ISSN: 0036-8075.
DOI: 10.1126/science.1229566.
URL: http://science.sciencemag.org/content/339/6117/321.

[6] H. Christina Fan et al. "Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood". In: *Proceedings of the National Academy of Sciences* 105.42 (2008), pp. 16266–16271.
DOI: 10.1073/pnas.0808319105.
URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2562413/.

[7] Sboner, Andrea and Mu, Xinmeng Jasmine and Greenbaum, Dov and Auerbach, Raymond K. and Gerstein, Mark B. "The real cost of sequencing: higher than you think!" In: *Genome Biology* 12.8 (Aug. 2011), p. 125. ISSN: 1474-760X.
DOI: 10.1186/gb-2011-12-8-125.
URL: https://doi.org/10.1186/gb-2011-12-8-125.

[8] Rachel R. J. Kalf et al. "Variations in predicted risks in personal genome testing for common complex diseases". In: *Genet Med* 16.1 (Jan. 2014), pp. 85–91. ISSN: 1098-3600.
DOI: 10.1038/gim.2013.80.
URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3883880/.

[9] Karen Norrgard. “Calculation of Complex Disease Risk”. In: *Nature* (2008). URL: https://www.nature.com/scitable/topicpage/calculation-of-complex-disease-risk-756.

[10] Kathryn Nave. “How Craig Venter is fighting ageing with genome sequencing”. In: *WIRED UK* (2016).
URL: http://www.wired.co.uk/article/craig-venter-human-longevity-genome-diseases-ageing.

[11] Amalio Telenti et al. “Deep sequencing of 10,000 human genomes”. In: *Proceedings of the National Academy of Sciences* 113.42 (2016), pp. 11901–11906.
DOI: 10.1073/pnas.1613365113.
eprint: http://www.pnas.org/content/113/42/11901.full.pdf.
URL: http://www.pnas.org/content/113/42/11901.abstract.

[12] Luke Timmerman. “Google Translate Star Leaves Venter’s Human Longevity For Illumina-Backed Grail”. In: *Forbes* (2016).
URL: https://www.forbes.com/sites/luketimmerman/2016/09/27/google-translate-star-leaves-venters-human-longevity-for-illumina-backed-grail.

[13] João Sá Sousa et al. “Efficient and secure outsourcing of genomic data storage”. In: *BMC Medical Genomics* 10.2 (July 2017), p. 46. ISSN: 1755-8794.
DOI: 10.1186/s12920-017-0275-0.
URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5547444/.

[14] Zachary D. Stephens et al. “Big Data: Astronomical or Genomical?” In: *PLOS Biology* 13.7 (July 2015), pp. 1–11.
DOI: 10.1371/journal.pbio.1002195.
URL: https://doi.org/10.1371/journal.pbio.1002195.

*[15]* Erika Check Hayden. “Privacy protections: The genome hacker”. In: *Nature* (2013).
URL: http://www.nature.com/news/privacy-protections-the-genome-hacker-1.12940.

[16] Laure-Anne Pessina. “Reconstructing a face from DNA: an EPFL alumnus takes the stage at the 2016 TED Conference”. In: *School of Engineering (Federal Institute of Technology Lausanne)* (2016).
URL: http://sti.epfl.ch/page-129921-en.html.

[17] Melanie Swan. “Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen”. In: *Journal of Personalized Medicine* 2.3 (2012), pp. 93–118. ISSN: 2075-4426.
DOI: 10.3390/jpm2030093.
URL: http://www.mdpi.com/2075-4426/2/3/93.

[18] Patrick Lin. “Blockchain: The Missing Link Between Genomics and Privacy?” In: *Forbes* (2017).
URL: https://www.forbes.com/sites/patricklin/2017/05/08/blockchain-the-missing-link-between-genomics-and-privacy.

[19] Justin Zimmerman. “DNA Block Chain Project Boosts Research, Preserves Patient Anonymity”. In: *CoinDesk* (2014).
URL: https://www.coindesk.com/israels-dna-bits-moves-beyond-currency-with-genes-blockchain/.

[20] D.S. Brama. “Method, System and Program Product for Transferring Genetic and Health Data”. US Patent App. 14/218,865. July 2015.
URL: https://www.google.com/patents/US20150205929.

[21] Melanie Swan. *Blockchain: Blueprint for a New Economy*.
URL: https://www.goodreads.com/book/show/24714901-blockchain.

[22] Matthew Herper. “Surprise! With $60 Million Genentech Deal, 23andMe Has A Business Plan”. In: *Forbes* (2015).
URL: https://www.forbes.com/sites/matthewherper/2015/01/06/surprise-with-60-million-genentech-deal-23andme-has-a-business-plan.

[23] “23andMe, Pfizer to Launch Inflammatory Bowel Disease Genetics Study”. In: *GenomeWeb* (2014).
URL: https://www.genomeweb.com/clinical-genomics/23andme-pfizer-launch-inflammatory-bowel-disease-genetics-study.

[24] “Genetic Wild West: 23andMe Raw Data Contains 75 Alzheimer’s Mutations”. In: *Alzforum* (2017).
URL: http://www.alzforum.org/news/community-news/genetic-wild-west-23andme-raw-data-contains-75-alzheimers-mutations.