大学院講義「先端医科学研究概論」

臨床ビッグデータ解析のための公開 Web ツールの紹介
臨床ビッグデータ解析のための一般公開されているウェブツールの紹介

佐倉 絵里 前園 絵里
バイオインフォマティクス 助教

2023-10-25

私のプロフィール

プロとしての経験

バイオインフォマティクス 助教2023年4月～現在
横浜市立大学先端医療研究センター

バイオインフォマティクス 研究者2020-2023 年 3 月
Craif Inc.、日本、分析チーム

さくら 前園
絵理博士

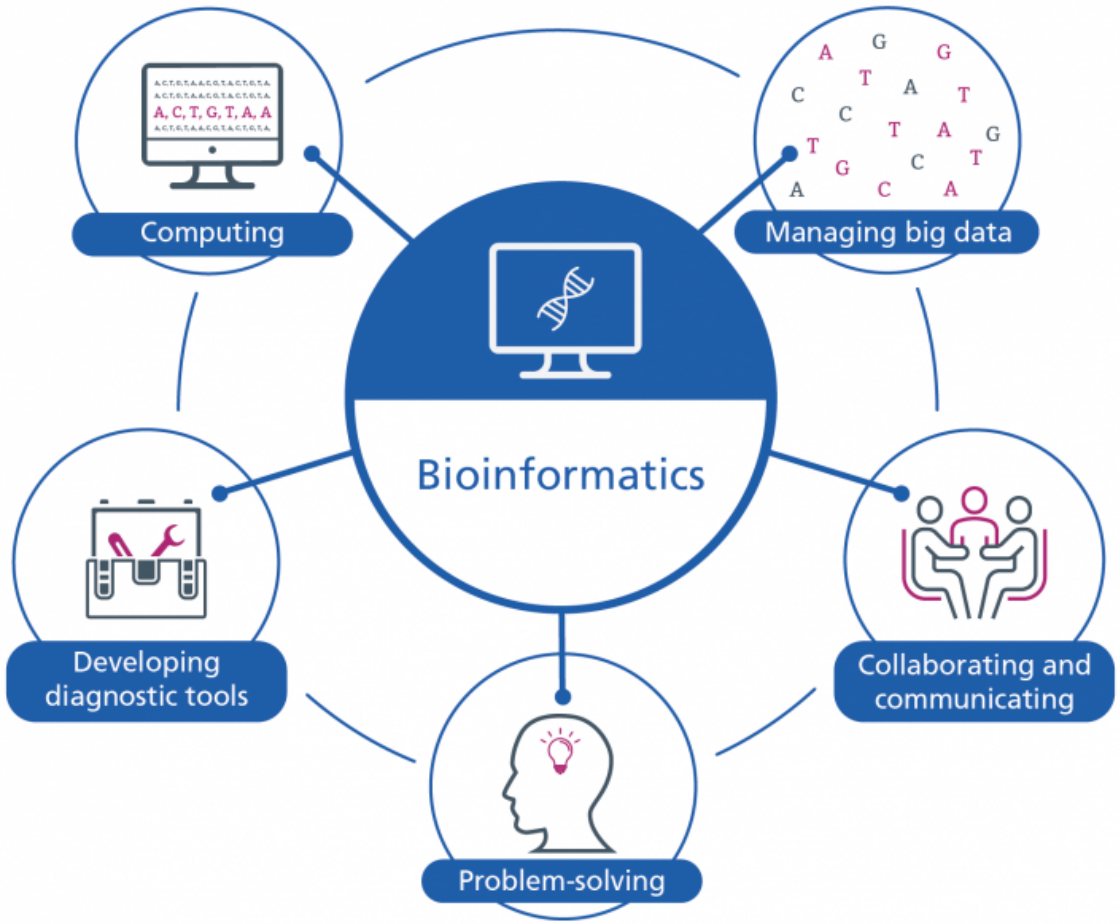
教育

筑波大学 2015-2020
博士課程修了 ヒューマンバイオロジー
大学院統合・グローバル専攻博士課程修了

スキル

研究設計・統計解析の計画と実施・臨床データ管理・R/Shiny/Bioconductor・Python/Streamlit/Django・JavaScript(Google apps script)・Gitを利用したバージョン管理・機械学習

バイオインフォマティクスとは何ですか?



私のプロフィール

プロとしての経験

バイオインフォマティクス助教 2023年4月～現在
横浜市立大学 先端医療研究センター

バイオインフォマティクス研究者 2020-2023 年 3 月
Craif Inc.、日本、分析チーム

さくら 前園
絵理博士

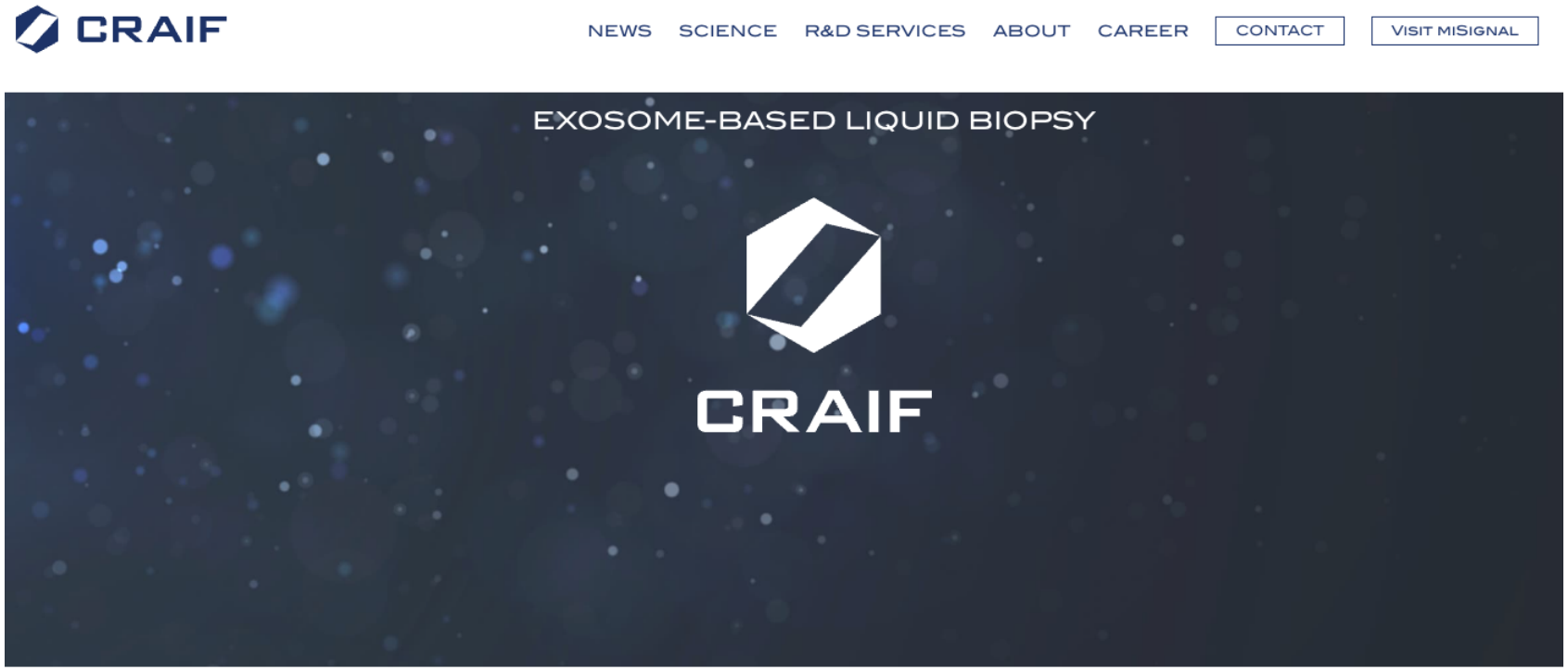
教育

筑波大学 2015-2020
博士課程修了 ヒューマンバイオロジー
大学院統合・グローバル専攻博士課程修了

スキル

研究設計・統計解析の計画と実施・臨床データ管理・R/Shiny/Bioconductor・Python/Streamlit/Django・JavaScript(Google apps script)・Gitを利用したバージョン管理・機械学習

過去の研究| 株式会社Craif バイオインフォマティクス研究員



人々が天寿を全うする社会の実現

Craifは、尿を利用したあらゆる疾患の早期発見と治療最適化を目指しています。
エクソソーム回収技術のパイオニアとして、精確で痛みのない疾患の早期発見方法を確立し
誰もが生涯にわたって健康でいられる社会を実現します。

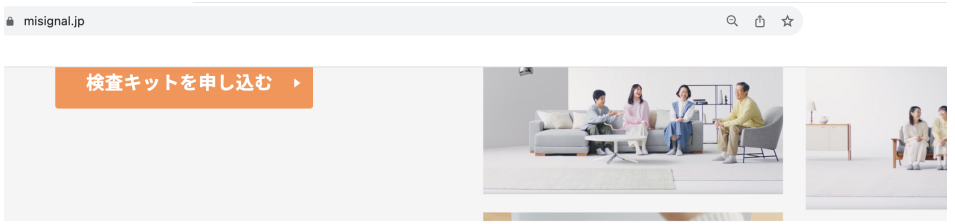
過去の研究| がんリスク評価キット: Craif miSignal®

7つのがんの種類

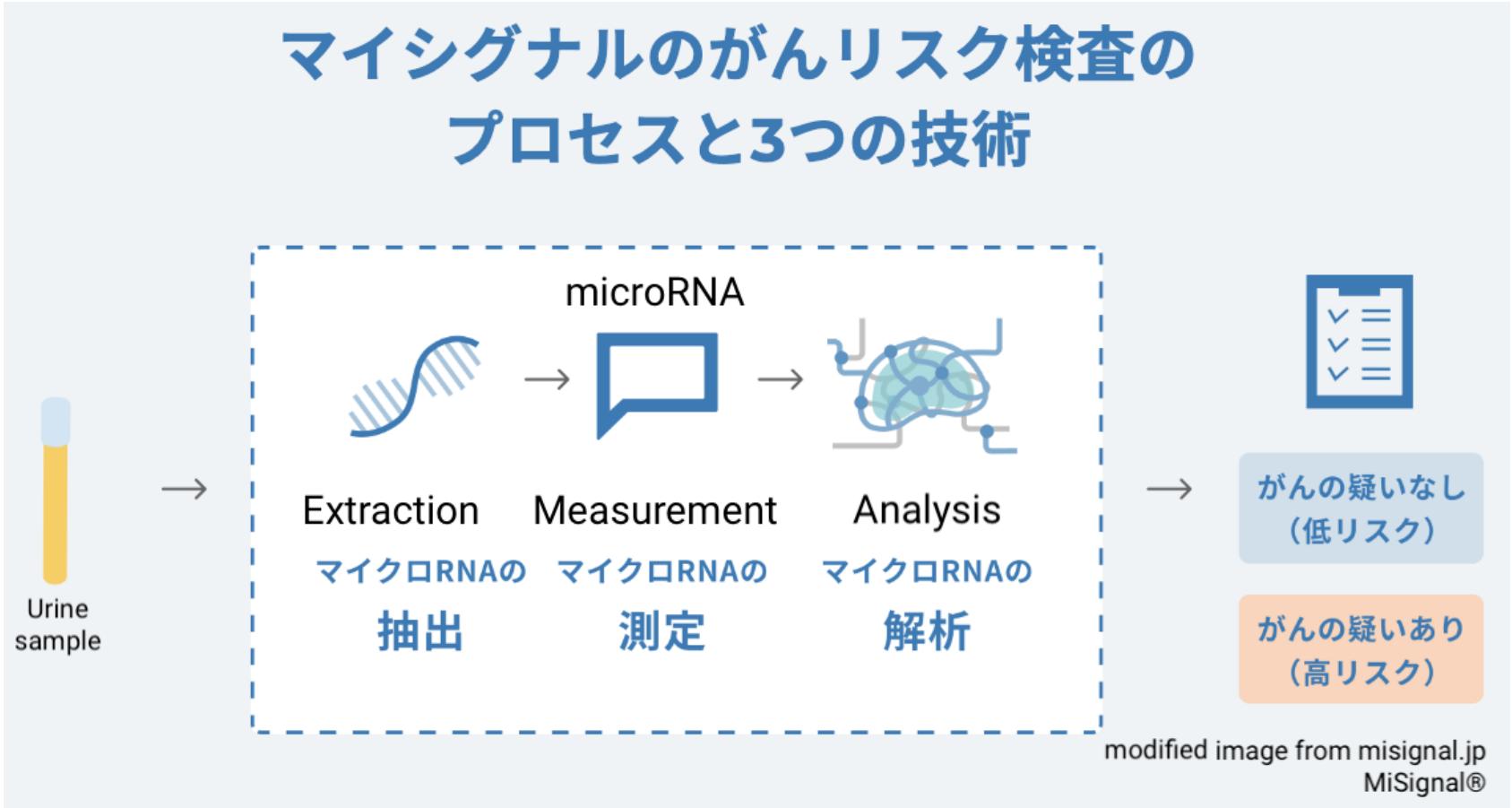
- 食道
- 肺
- 胸
- 胃
- 膵臓
- 結腸直腸
- 卵巣

責任

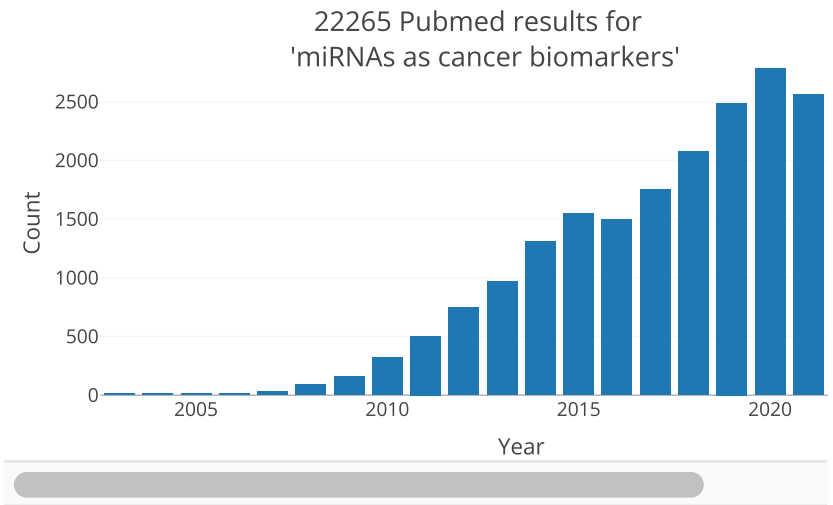
- 臨床データ管理
- miRNA発現データの前処理
- バイオマーカー発見分析
- 予測アルゴリズム開発



過去の研究| MiSignal® の仕組み

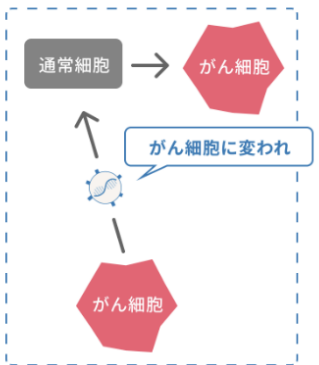


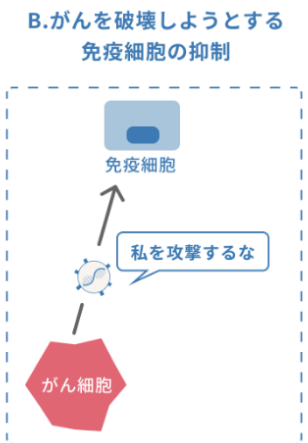
過去の研究| がんバイオマーカーとしてのマイクロRNA

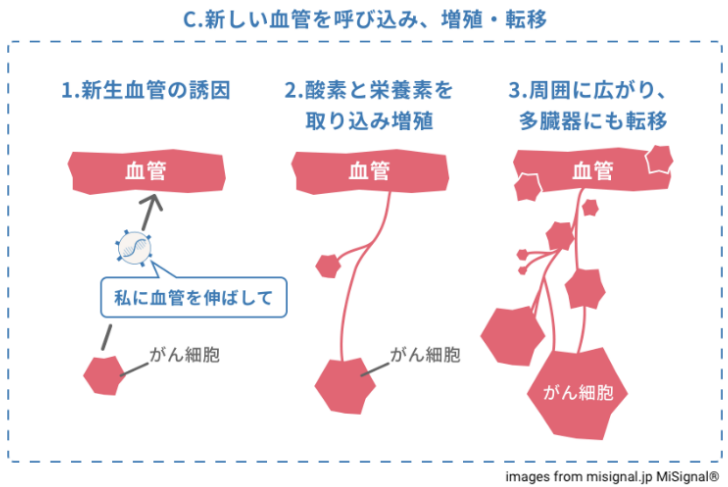


どうやって？

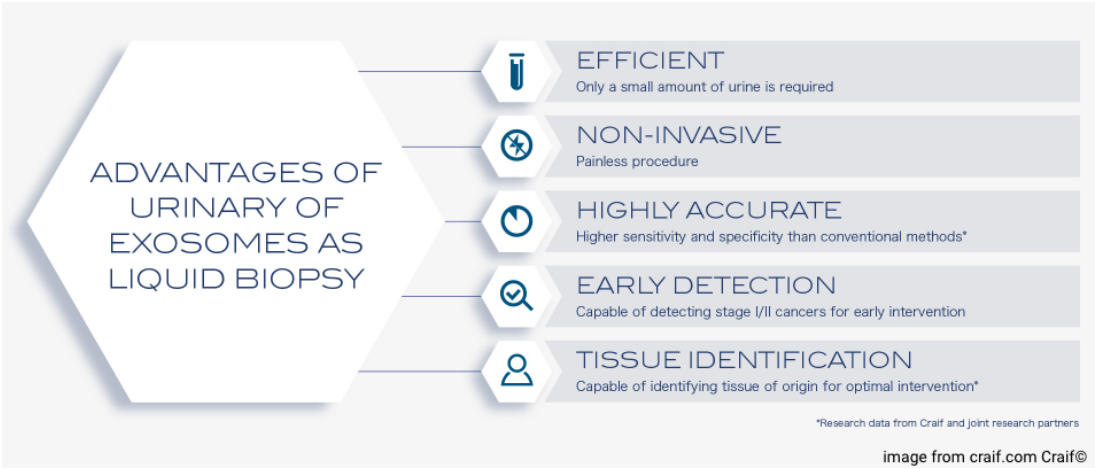
- **遺伝子発現の調節**への関与
- 診断、分類、予後、治療反応など、**がんのさまざまな側面との関連**性
- その**多用途性と非侵襲的検出により、**腫瘍学における貴重なツールとなっています。





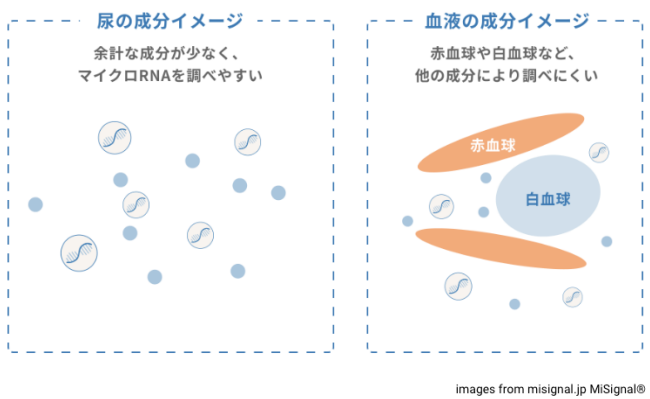


過去の研究| なぜ尿なのか？

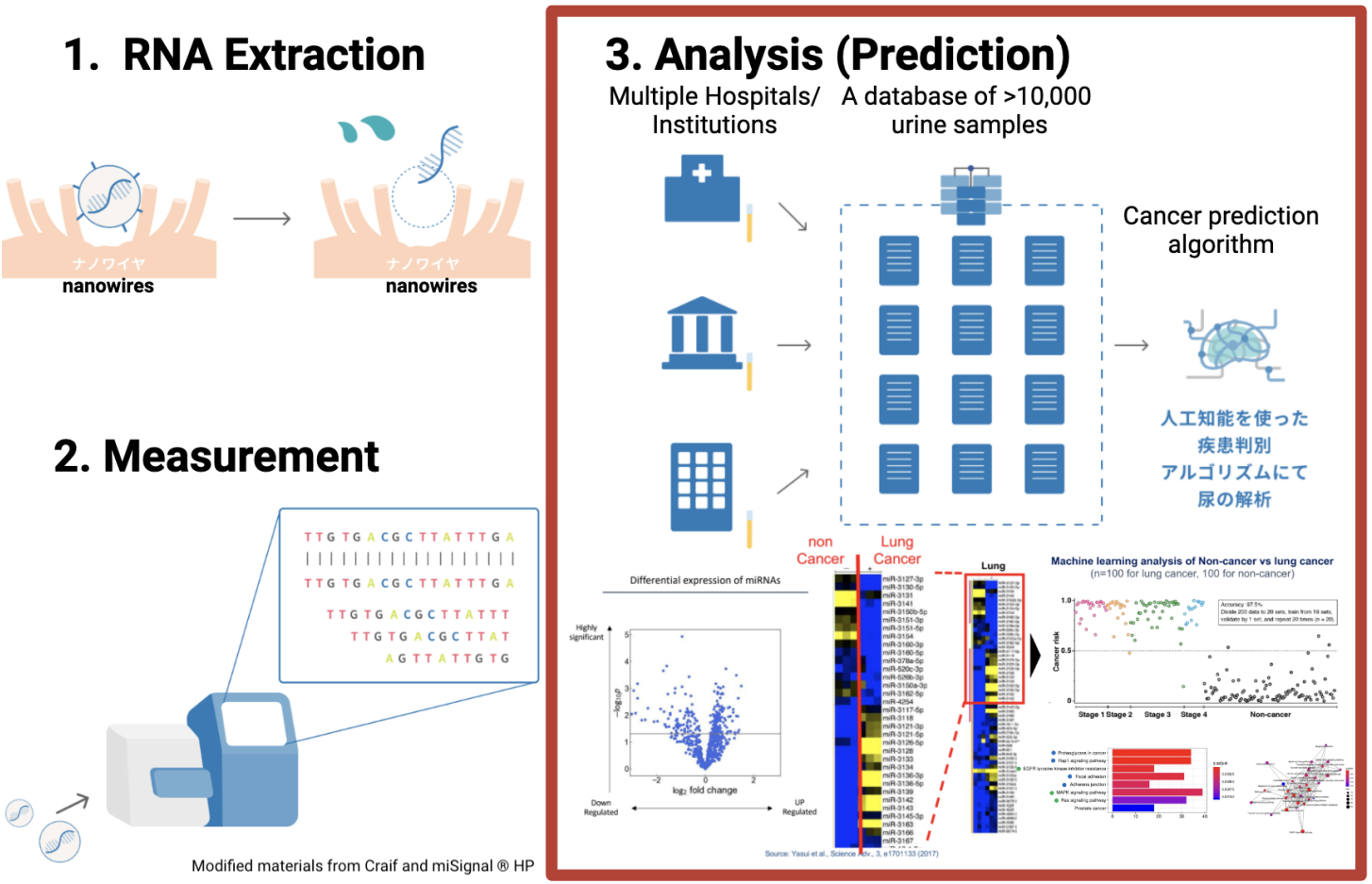


なぜ血ではないのでしょうか？

- 注射針が必要（非 侵襲的）
- より低い miRNA が検出されました (~ 600)。タンパク質などの不純物は検出を妨げる可能性があります (vs. >1300 ; 腎臓による濾過により汚染物質はほとんどありません)



過去の研究| miSignal® の開発におけるバイオインフォマティクスの役割



Modified materials from Craif and miSignal ® HP

私のプロフィール

プロとしての経験

バイオインフォマティクス 助教 2023年4月〜現在
横浜市立大学 先端医療研究センター

バイオインフォマティクス研究者 2020-2023 年 3 月
Craif Inc.、日本、分析チーム

教育

筑波大学 2015-2020
博士課程修了 ヒューマンバイオロジー
大学院統合・グローバル専攻博士課程修了

スキル

研究設計・統計解析の計画と実施・臨床データ管理・R/Shiny/Bioconductor・Python/Streamlit/Django・JavaScript(Google apps script)・Gitを利用したバージョン管理・機械学習

筑波大学 University of Tsukuba　CRAIF

さくら 前園
絵理博士

Bioinformatics Laboratory

バイオインフォマティクス研究室

https://www-user.yokohama-cu.ac.jp/~bioinfo/

YCU Bioinformatics Laboratory

Home Members Research News Education Recruitment

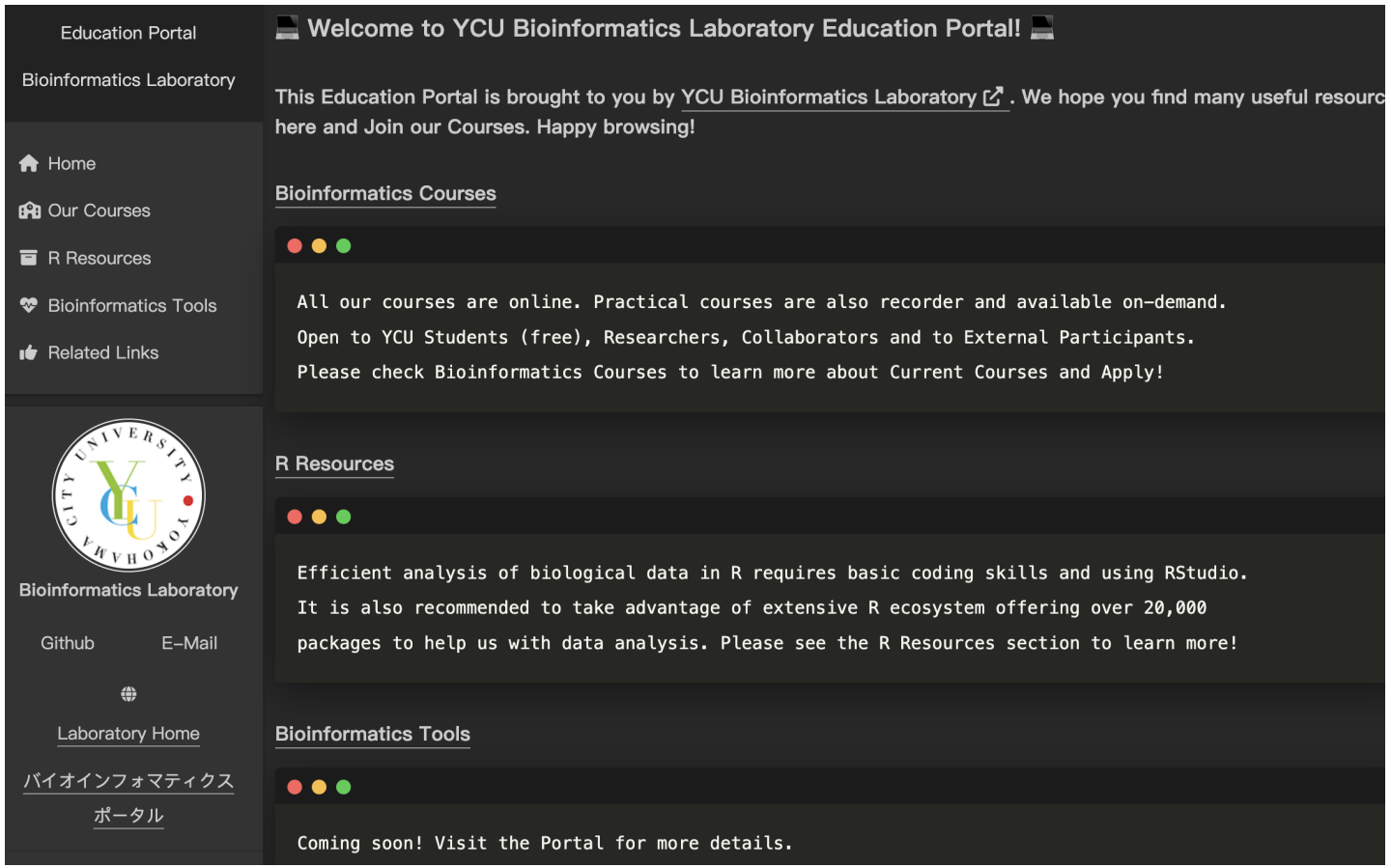
Yokohama City University (YCU) is an important and innovative Research & Educational Institution in Japan. **A**dvanced **M**edical **R**esearch **C**enter (AMRC/SENTANKEN) was established in 2006 at YCU Fukuura Medical Campus to further promote medically–relevant research through various industry and academic collaborations. **Bioinformatics Laboratory** was formed in 2009 as an integral part of AMRC.

Our Mission

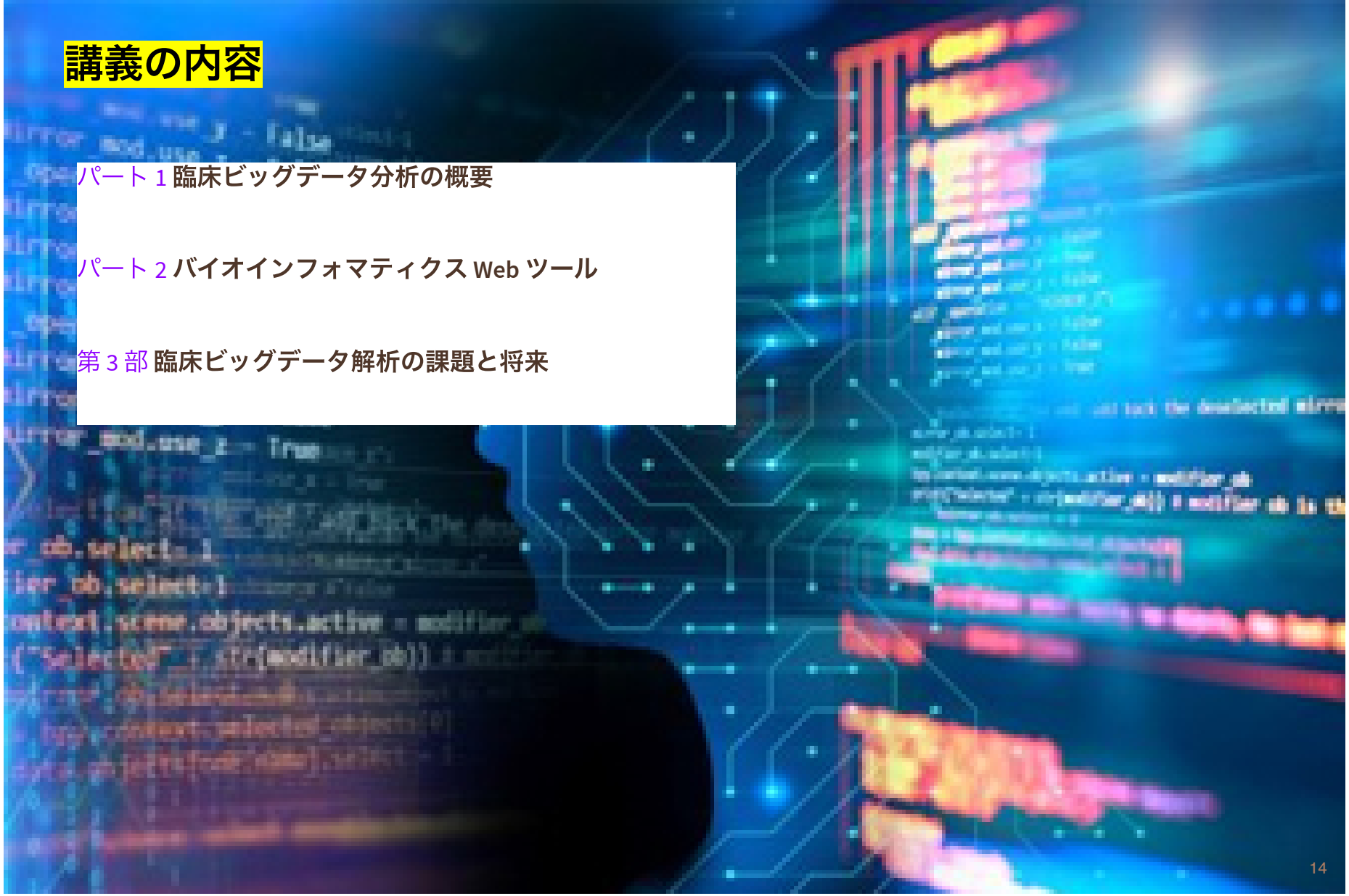
- 🧬 biologically–relevant research
- 📚 bioinformatics education
- 🤝 collaborations & internationalization

バイオインフォマティクス教育ポータル

https://edu.med.yokohama-cu.ac.jp/



講義の内容



パート 1 臨床ビッグデータ分析の概要

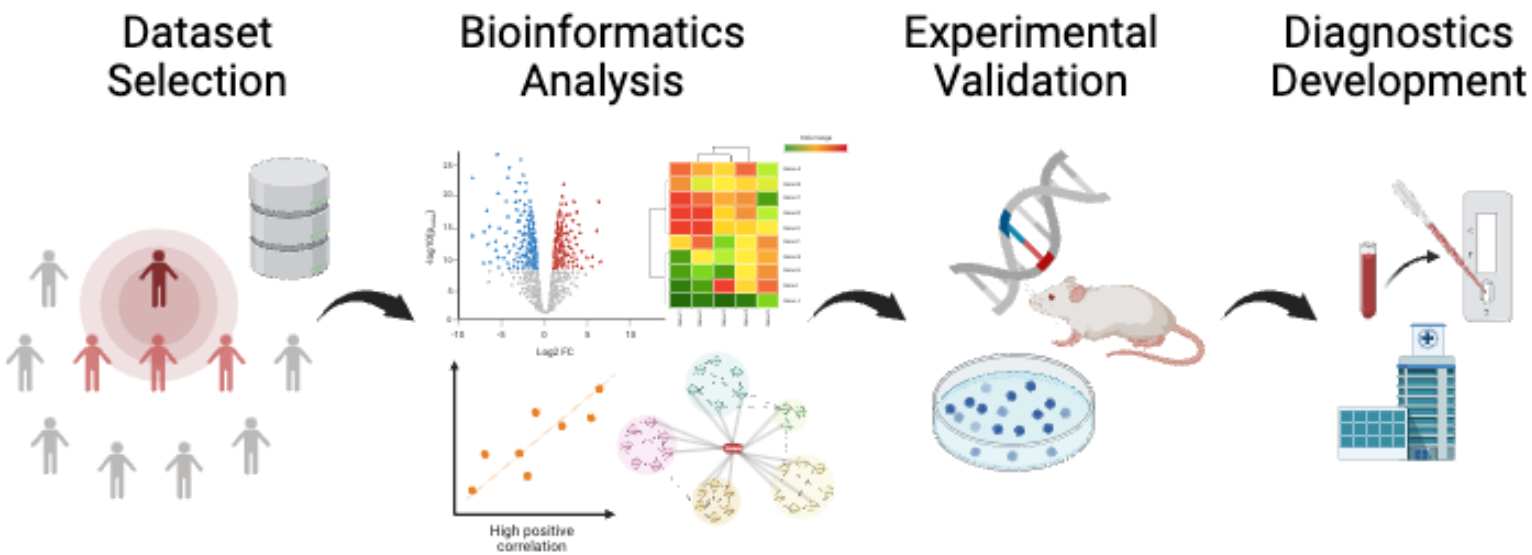
パート 2 バイオインフォマティクス Web ツール

第 3 部 臨床ビッグデータ解析の課題と将来

パート 1: 臨床ビッグデータ分析の概要

What is Clinical Big Data Analysis?

the process of extracting valuable insights from vast and diverse datasets related to healthcare and medicine

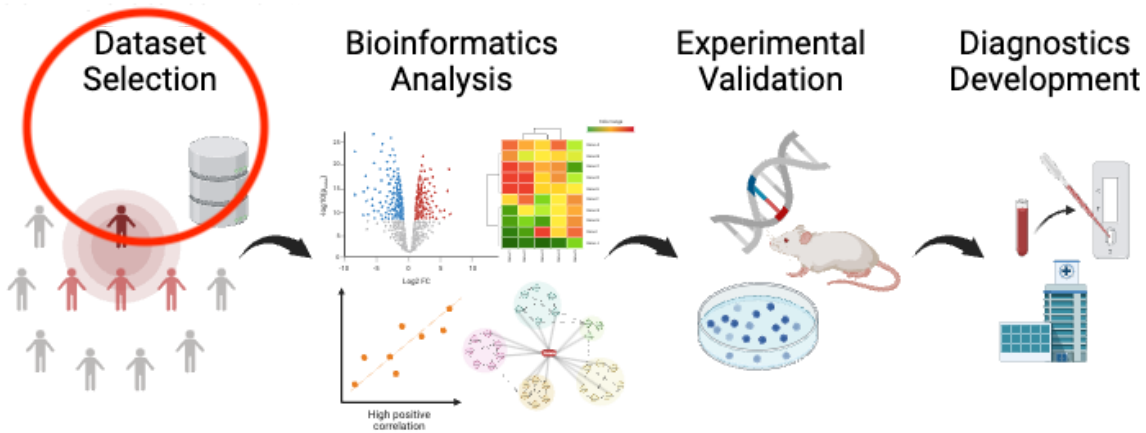


Importance of Clinical Big Data Analysis

The **complete sequencing of the human genome** has helped to unlock the genetic contribution for many diseases
Its applications include the following:

- **drug discovery**: Drug target identification and drug candidate screening can be accelerated, and safer/more effective drugs can be developed based on **molecular modelling and simulation**
- **gene therapy**: Identifying the **best gene target site** for each individual by taking their genetic profile into consideration can reduce the risk of unintended side effects
- **personalized medicine**: A patient's **genetic profile** can assist the doctor to predict susceptibility to certain diseases, provide proper medication, and with the proper dose to reduce side-effects
- **preventive medicine**: Genomics, proteomics, and metabolomics data are analyzed for possible **disease biomarkers** to develop screening tests that identify the disease at an early stage

It always starts with...



Types of data involved

- **Patient Data**: Demographic information, lifestyle factors, and health-related behaviors
- **Electronic Health Records (EHRs)**: Comprehensive patient records, including medical history, treatment, and lab results
- **Genomics Data**: Information about an individual's genetic makeup, including DNA sequences and variations

Available Big Data (databases) and how to access them

Clinical

- store and manage data related to patients' medical history, diagnoses, treatments, and outcomes
  - Surveillance, Epidemiology, and End Results (SEER)
  - National Health and Nutrition Examination Survey (NHANES)
  - The Cancer Proteome Atlas (TCPA)
  - The Cancer Genome Atlas (TCGA)
  - Medical Information Mart for Intensive Care (MIMIC)

Genomic

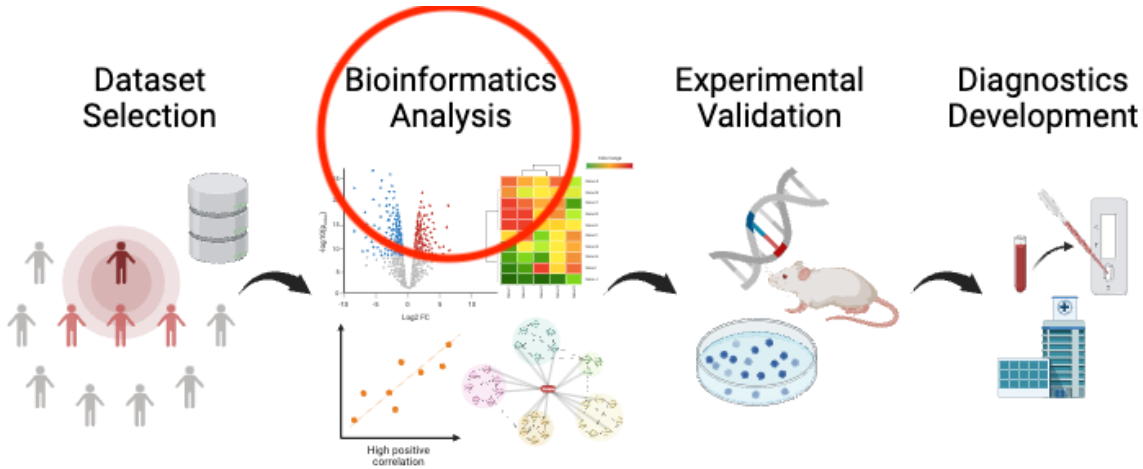
- store and manage genetic information, including DNA sequences, gene annotations, and variations in DNA
  - Database of Genomic Structural Variation (dbVar) and Database of Genotypes and Phenotypes(dbGaP)
  - Gene, GenBank, and RefSeq
  - Gene Expression Omnibus (GEO) & GEO Datasets and Genome Data Viewer (GDV)
  - International Genome Sample Resource (IGSR)
  - Cancer Single-cell Expression Map (CancerSCEM)

How to access?

Databases provide instructions but it is usually via the following:

⭐**Direct download from the website** ⭐**FTP server** ⭐**API (Shell/Python/R)**

After selecting data, analysis can be done…



- via R, Python, and other programming languages
- via **Publicly-available web tools**

Part 2:Bioinformatics Web Tools

Publicly Available Web Tools

- **online** applications **accessible to a wide audience, often free/low-cost**
- aid in processing, managing, and interpreting large clinical/genomic datasets

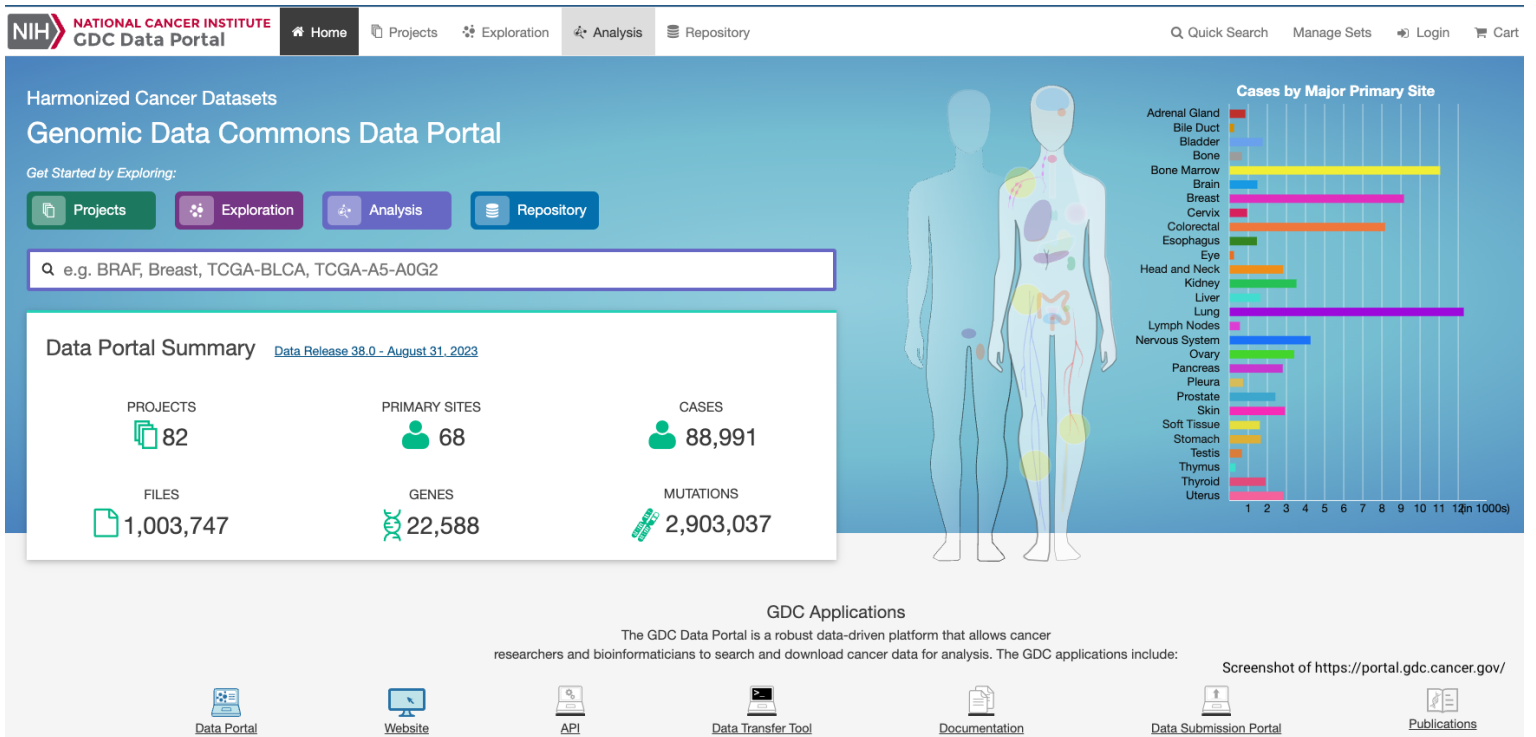
> **Main purpose**
>
> Public Web tools **turn data into interpretable results democratizing advanced data analysis without complex installations, programming skills, or high expenses**

Advantages

- **Accessibility**: easy access to advanced data analysis capabilities for healthcare professionals and researchers
- **Cost-Effectiveness**: often free/affordable, reducing financial barriers
- **Community Support**: users benefit from a **collaborative community**, sharing knowledge and solutions, enhancing the tools' utility and troubleshooting capabilities

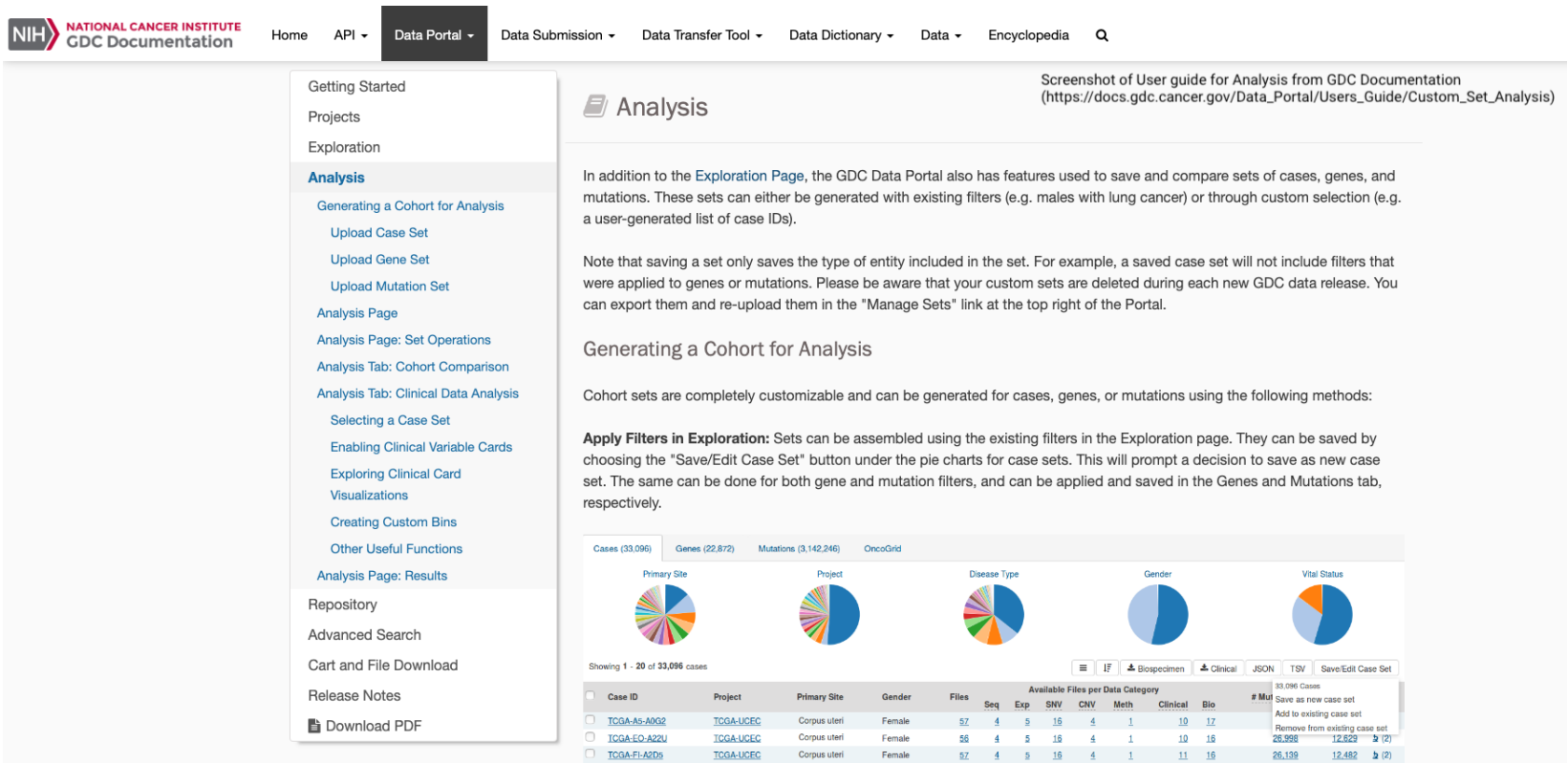
Examples of Publicly available web tools

- GDC Data Portal
- cBioPortal
- UCSC Xena
- TANRIC
- miEAA
- RCoV19
- CancerSCEM analyze modules
- Integrative Genomics Viewer (IGV)
- RNAseq analysis on the Web



How to Access and Use Web Tools

- Most explain how to access tools on their websites (web links, registration, etc.)
- A step-by-step guide (documentation) on how to use one or more of these tools for data analysis is usually provided

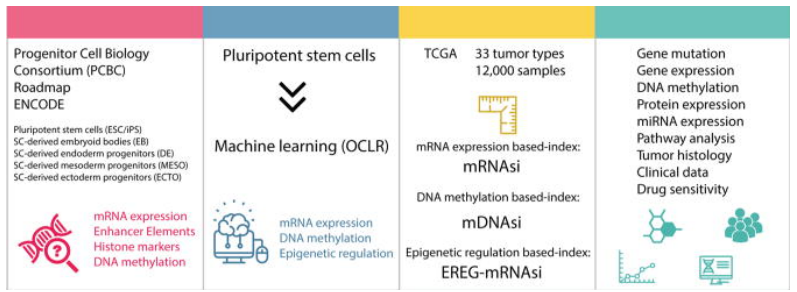


Case Studies – Publishing using Web tools

⭐ Case Study 1: GDC Data Portal

TCGA enabled the researchers to analyze cancer stemness in ~12,000 samples of 33 tumor types

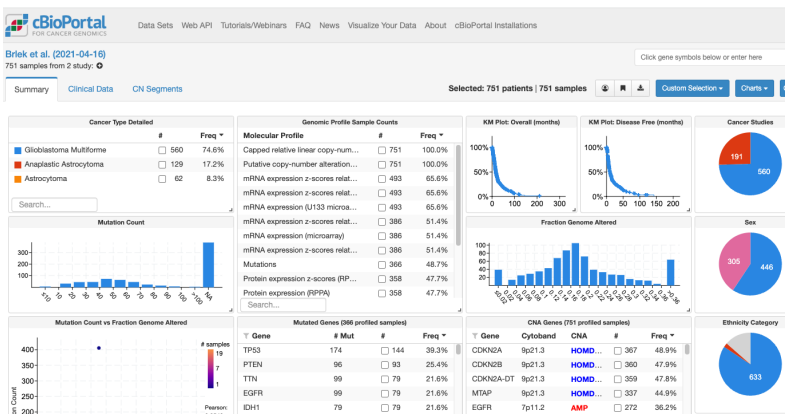
Publication: Fujimoto K., Ito K., Saito Y., et al. **Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation**. *Cell Reports*, 23(11), 3306-3320.e10, 2018. https://doi.org/10.1016/j.cell.2018.03.034



⭐ Case Study 2: cBioPortal

The researchers investigated AKT1, AKT2, AKT3, CHUK, GSK3β, EGFR, PTEN, and PIK3AP1 as participants of EGFR-PI3K-AKT-mTOR signaling using data from cBioPortal

Publication: Brlek, P.; Kafka, A.; Bukovac, A.; Pećina-Šlaus, N. **Integrative cBioPortal Analysis Revealed Molecular Mechanisms That Regulate EGFR-PI3K-AKT-mTOR Pathway in Diffuse Gliomas of the Brain**. *Cancers* 2021, 13, 3247. https://doi.org/10.3390/cancers13133247



Tips & Best Practices and Pitfalls to Avoid

DOs

Know Your Data – Understand the format and quality of your data

Take your time with Data prep – Clean and preprocess data as needed

Select Appropriate tools – Choose the right tool for your analysis

Read Documentation – Study tool guides and understand their limitations

Pay attention to Parameters – Set tool parameters carefully

Record Parameters – Keep records for reproducibility

Validate results – Verify results with independent data or experiments

Secure Data – Comply with data privacy regulations

Seek Help – Collaborate or ask for assistance if needed

DON'Ts

Misinterpret your Data – Be cautious in result interpretation

Take Data Quality for granted – Assess and preprocess data to ensure quality

Use all Data when unnecessary – Analyze relevant subsets for efficiency

Depend on one Tool – Use multiple tools for comprehensive analysis

Ignore Updates – Use the latest tool versions

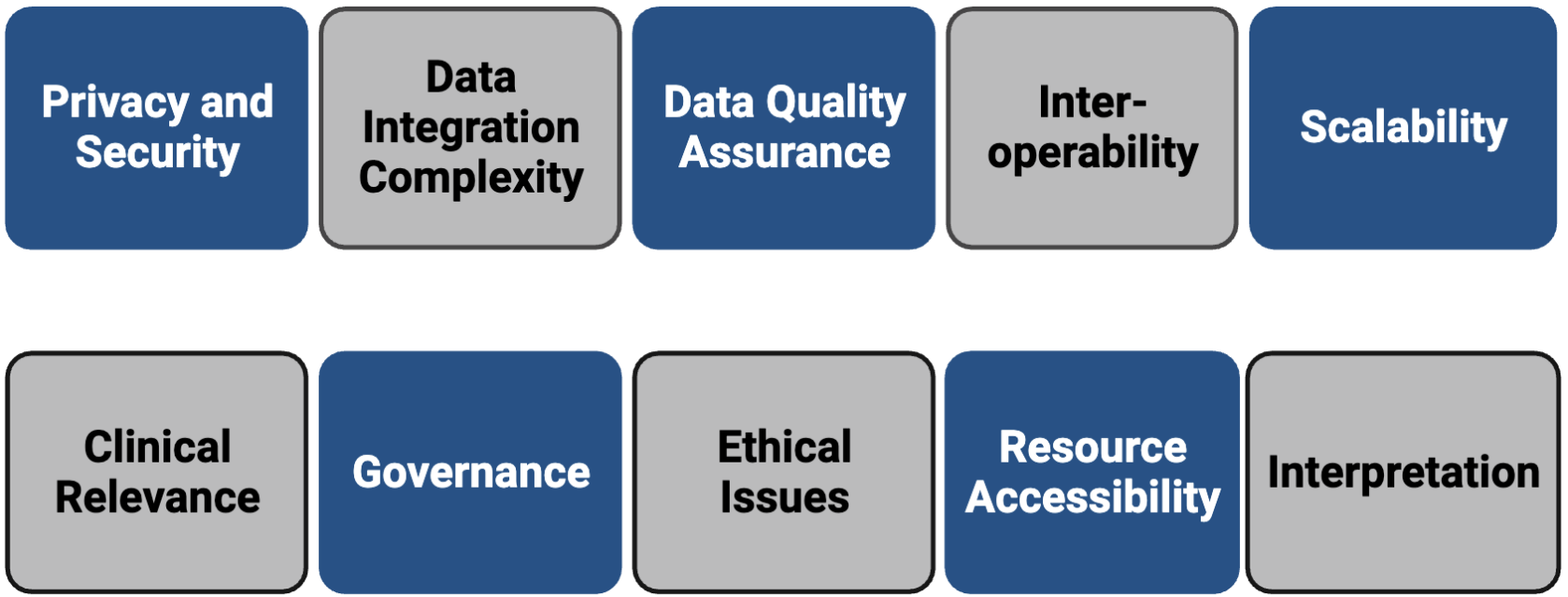
Forgo Resource Check – Check hardware for computational capacity

Forget Publication Quality – Follow best practices for reporting

Neglect Ethical Considerations – Respect ethical guidelines and permissions

Part 3: Challenges and Future of Clinical Big Data Analysis

Key Challenges in Clinical Big Data Analysis with existing web tools



Future of Clinical Big Data Analysis

Collaboration and Integration

Interdisciplinary Collaboration

Teams of healthcare providers, data scientists, and researchers working together to drive innovation

Integration into Routine Healthcare

Seamless incorporation of data analysis into everyday healthcare practices for data-driven decision-making and personalized care

Global Data Sharing

Enhanced collaboration and sharing of data among healthcare institutions and researchers to deepen disease understanding and improve treatments

Current research | Project: Clinical data analyzer

迅速な臨床ビッグデータクリーニングと解析のための統合的ノーコードウェブアプリ

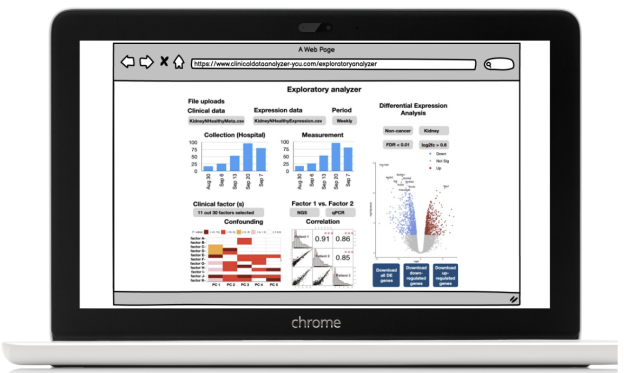
Problem

- The **quality of the input data is critical** to the final results and their interpretation
- **HOWEVER**, in Healthcare and Medicine, there are many examples of rich but **unorganized, incomplete**, and **inconsistent data**

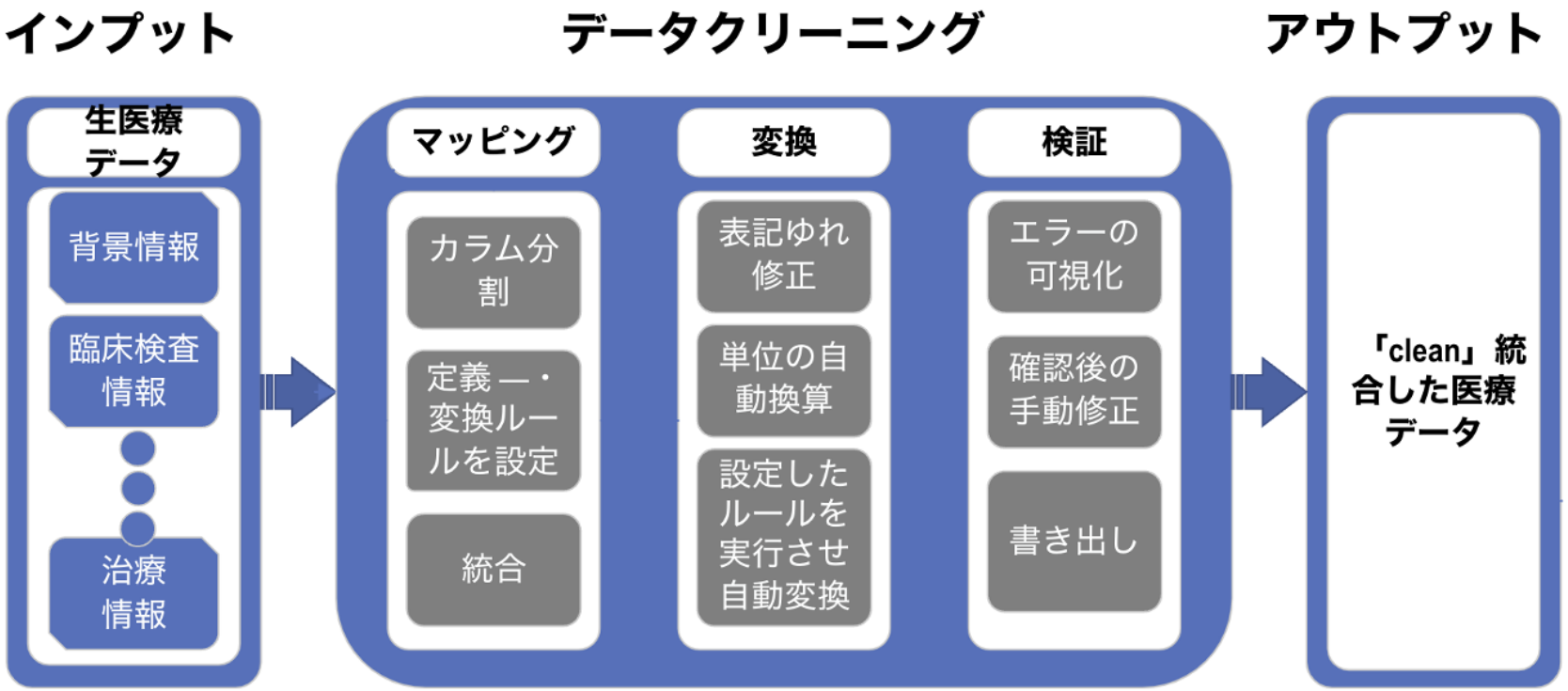
Solution: Clinical Data Analyzer

Integrated no-code web app development for rapid clinical big data cleaning and analysis (**Collaborators wanted!**)

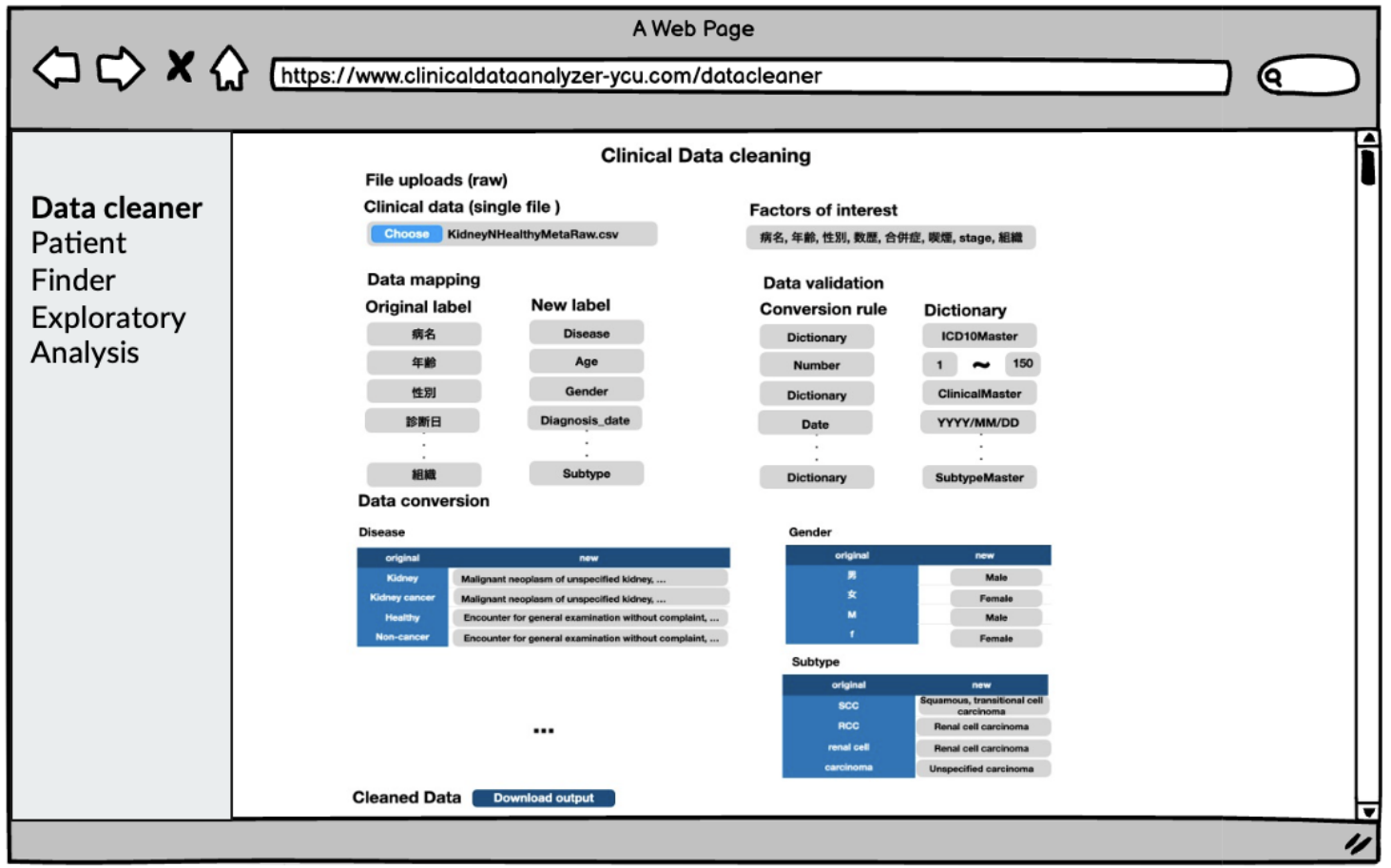
- a free web application that allows medical practitioners to quickly and easily construct initial hypotheses from data with so-called no-code
- consists of **three** major tools:
    1. Data cleaner
    2. Patient Finder
    3. Exploratory Analyzer



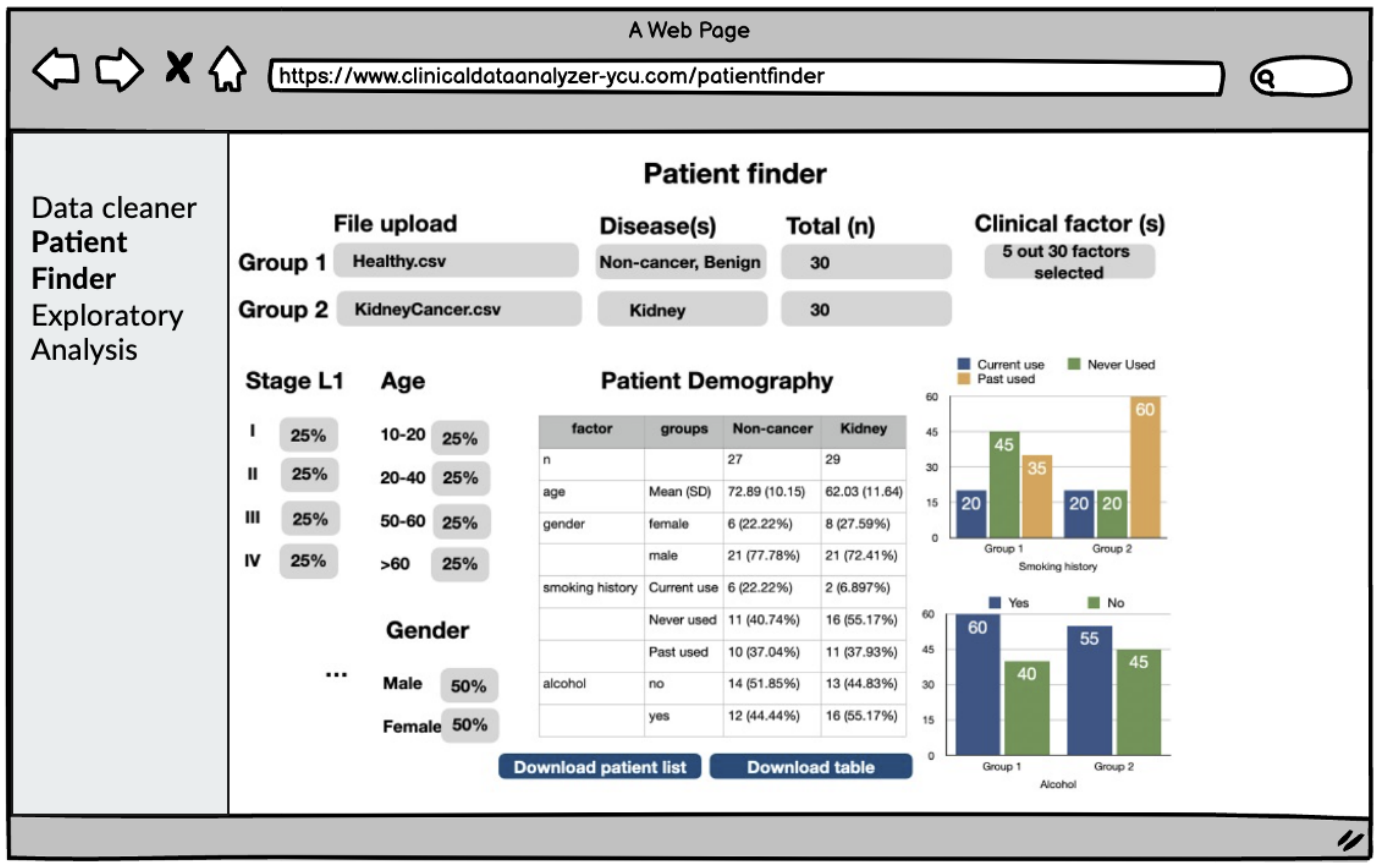
Current research | 1. Data cleaner: a semi-automated preprocessing data cleaning tool (1)



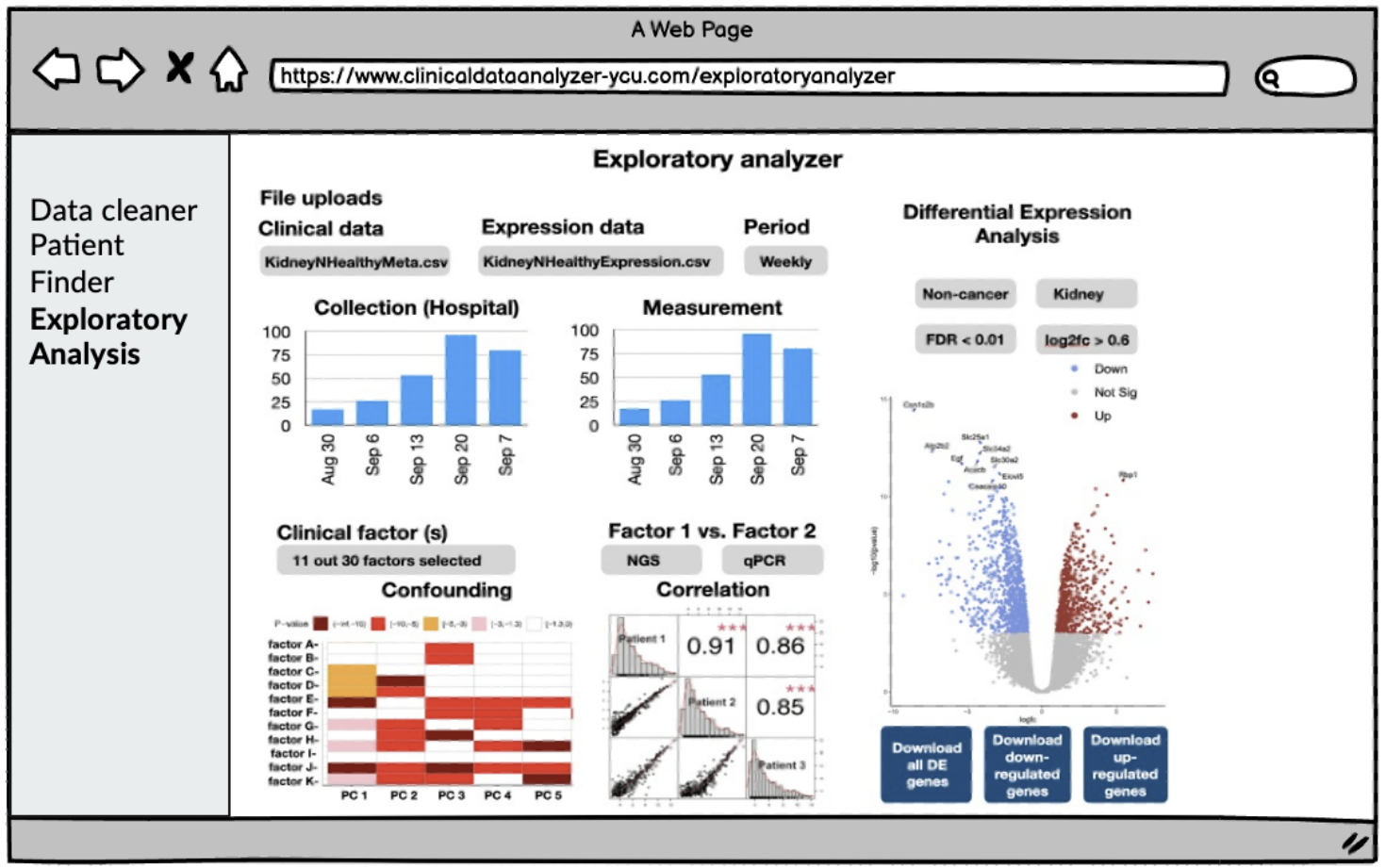
Current research | 1. Data cleaner: a semi-automated preprocessing data cleaning tool (2)



Current research | 2. Patient Finder: a patient cohort selector and visualizer tool



Current research | 3. Exploratory Analyzer: an interactive data analysis tool

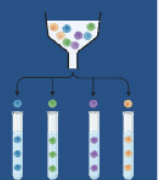


Future research | Projects/Research Interest

Creation of databases and user-friendly web resources for biological and medical data analysis and exploration

Computational assessment of gene signatures across different cancer types

Identification of biomarkers for the diagnosis and treatment of diseases, including cancer



Employing a multi-dimensional genomic approach to enhance our comprehension of stage 0 cancers

Interested in becoming a collaborator?

Contact me: sakura.maezono[at]yokohama-cu.ac.jp
Contact us: bioinfo[at]yokohama-cu.ac.jp

Summary

1. Clinical Big Data Analysis has facilitated the **extraction of valuable insights** from the continuously expanding healthcare and medical data

2. バイオインフォマティクス Web ツールは、アクセシビリティ、費用対効果、コミュニティ サポートの存在を通じて**高度なデータ分析を民主化します。**これらのツールを効果的に使用するには、正しいデータ入力、ドキュメントの読み取り、データ検証の知識が必要です。

3. **協力し、Web ツールを使用し、世界中でデータを共有することで、**病気についての理解を深め、現場のデータ品質、拡張性、ガバナンスなどの課題に対処できます。

持ち帰りメッセージ

バイオインフォマティクス Web ツールを活用して、増え続ける臨床ビッグデータに飛び込むことができます。

コラボレーションとデータの共有を通じて、**あなたは**病気に関する私たちの集合的な知識に積極的に貢献し、医療のより明るい未来を育むことができます。

質問とディスカッション