VCU Board of Visitors

ACADEMIC, HEALTH AFFAIRS AND RESEARCH COMMITTEE

Thursday, April 23, 2026
2:00 pm.[1]
The Honorable Benjamin Lambert, III Board Room
1213 Clay Street
Richmond, VA

AGENDA

1. **CALL TO ORDER** — **Dr. V. Dale Jones**, *Chair*

2. **ACTION ITEMS:** — **Dr. V. Dale Jones**, *Chair*
   5 minutes (3:00 - 3:05 p.m.)
   a. Feb. 27, 2026 meeting minutes
   b. Proposal to establish a Bachelor of Science in Public Health (BSPH) in Public Health, School of Public Health
   c. Proposal to establish a Master of Science (MS) in Pharmaceutical Engineering, College of Engineering and School of Pharmacy
   d. Proposal to close the Bachelor of Science (BSEd) in Health and Physical Education, School of Education

3. **CLOSED SESSION** — **Dr. V. Dale Jones**, *Chair*
   10 minutes (3:05 - 3:15 p.m.)
   Freedom of Information Act Section 2.2-3711.A.2-
   Selection of the Board of Visitors Award Recipient
   which requires disclosure of scholastic records

4. **RETURN TO OPEN SESSION** — **Dr. V. Dale Jones**, *Chair*
   1 minute (3:15 - 3:16 p.m.)

5. **EXECUTIVE REPORT:**
   Leveraging AI for VCU's Academic Enterprise
   55 minutes (3:15 - 4:11 p.m.)

   Dr. Arturo Saavedra, *Interim Executive Vice President and Provost*

   Dr. Marlon Levy, *Senior Vice President for Health Sciences and CEO VCU Health*

   Dr. Srirama Rao, *Vice President for Research and Innovation*

[1] The start time for the Board of Visitors meeting is approximate only. The meeting may begin either before or after the listed approximate start time as Board members are ready to proceed.

6. **OTHER BUSINESS** — **Dr. V. Dale Jones**, *Chair*

7. **ADJOURNMENT** — **Dr. V. Dale Jones**, *Chair*

In accordance with the Board's operating procedures and in compliance with the Virginia Freedom of Information Act, there will be no opportunity for public comment at this meeting.

CONFIDENTIAL SUMMARY
ACADEMIC, HEALTH AFFAIRS AND RESEARCH COMMITTEE OF
THE VCU BOARD OF VISITORS
April 23, 2026

OPEN SESSION

ACTION ITEMS:

- Feb. 27, 2026 meeting minutes
- Proposal to establish a Bachelor of Science in Public Health (BSPH) in Public Health, School of Public Health
- Proposal to establish a Master of Science (MS) in Pharmaceutical Engineering, College of Engineering and School of Pharmacy
- Proposal to close the Bachelor of Science (BSEd) in Health and Physical Education, School of Education

FOR INFORMATION/COMMITTEE REVIEW:

1) Items that may be action items at upcoming committee meeting:
   - Draft Policy for Post-Tenure Review

2) Items that the board needs to be aware of, but will not require action:
   - None

EXECUTIVE REPORT:

Joint Report from the Provost and Senior Vice President for Academic Affairs; the Senior Vice President for Health Sciences and CEO VCU Health; and the Associate Vice President for Research and Innovation

Interim Executive Vice President and Provost **Arturo Saavedra, M.D., Ph.D.**; Senior Vice President for Health Sciences & CEO of VCU Health **Marlon Levy M.D., MBA**, and Vice President for Research and Innovation **P. Srirama Rao, Ph.D.** will present on **Leveraging AI across the VCU enterprise.**

VCU's growth in size and impact has also increased the complexity in which the organization must function. Leveraging AI to increase functionality, productivity and efficiencies has become a crucial exercise across the academic enterprise. The session for the Academic, Health Affairs and REsearch Committee will provide a brief overview of three efforts across campus focused on implementing AI technologies. The panel will include:

Milos Manic, Ph.D.
Professor, College of Engineering
Director, VCU Convergence_AI

Preetam Ghosh, Ph.D.
Professor, College of Engineering
Faculty director, High Performance Research Computing Core

Sandeep Kothiwale, Ph.D.
Enterprise data scientist
VCU Health System Authority

Proposed New Degree Program Brief

New Degree Program: Bachelor of Science in Public Health (BSPH) in Public Health

Overview

Virginia Commonwealth University (VCU) seeks approval to establish a Bachelor of Science in Public (B.S.P.H) in Public Health. The program will be administered by the Dean's Office within the School of Public Health.

The purpose of the BSPH in Public Health degree is to develop scholars with competence in all foundational domains of a CEPH-accredited bachelor of public health degree program. Students will develop expertise in the basic processes, approaches, and interventions used to identify and address major health-related needs and concerns of populations. Students will also gain knowledge in statistics, the sciences, the history and foundations of public health and disease, U.S. and global health systems, health communication, and the legal, ethical, and regulatory dimensions of health. Core coursework also builds skills in public health data collection and use, as well as the design and implementation of population-level health interventions. By integrating foundational theory with applied practice, the curriculum prepares graduates for entry-level roles related to health and well-being in community and governmental agencies, health promotion and disease prevention, healthcare settings, health policy, research, and related industries.

Delivery Format

The proposed program will be offered in a face-to-face delivery format.

Target Implementation Date

Fall 2027

Demand and Workforce Development

Data from the Bureau of Labor Statistics (BLS) show a projected 13.5% increase in employment demand over ten (10) years for aligned occupational titles, while the Virginia Employment Commission (VEC) projects a 14.9% increase for the same set of occupations. Similarly, the Virginia Office of Education Economics (VOEE) reports an average five-year workforce growth projection of 16.7% for these aligned occupational titles, and Virginia GO Region 4 projects a 13.6% increase over the same period. The included Lightcast report indicates a job-posting intensity of approximately 2:1—meaning that for every two postings, there is one unique job posting. This is comparable to the regional average posting intensity of 3:1 across all occupations, suggesting that employers are exerting a typical level of effort to hire for these roles.

External Competition

The same or related degree programs in Public Health or Community Health exist at the following State Council of Higher Education for Virginia (SCHEV) institutions: George Mason University, Old Dominion University, Radford University, Virginia State and Polytechnic University, and College of William & Mary.

Target Population

The target audience for the BSPH in Public Health degree program is for those who interested in addressing community health problems, promoting health, planning, implementing, and evaluating healthcare programs and interventions.

Impact on Existing Programs

The proposed degree program will not replace or eliminate any existing degree programs at VCU.

Impact on Faculty

The proposed program will require 1.9 FTE of faculty instructional effort in the initiation year, rising to 14.6 FTE by the target year 2031-2032. The School of Public Health has a total of 54 faculty members. Twelve (12) faculty members will teach core and required courses for the proposed BSPH in Public Health degree program. No adjunct faculty will be utilized to initiate and sustain the proposed degree program.

Funding

The School of Public Health anticipates needing to hire two (2) additional faculty members with expertise in social and behavioral sciences and epidemiology to teach in the BSPH program as well as two (2) additional staff members to assist with advising/student services. The Dean of the School of Public Health has committed resources for a total of two (2) new faculty members who will teach in the proposed degree program beginning Fall of 2029.

Return on Investment

The Foundation for Research on Economic Opportunity contains comparative Return on Investment Data for Bachelor of Science in the field of Public/Community Health. Comparative data from public institutions in the state of Virginia is provided below.

Alignment with the VCU Mission

The proposed BSPH in Public Health degree program aligns with the mission of Virginia Commonwealth University. The proposed degree program will prepare students to design and implement public health interventions that promote health "equity" in diverse populations. By integrating knowledge from multiple disciplines, including epidemiology, health policy, community engagement, and health promotion, students will address "real-world" public health challenges. Graduates will improve population "health" and well-being in communities, governmental agencies, healthcare systems, and other sectors by implementing health and disease prevention strategies.

Next Steps

- **February 19, 2026** – University Undergraduate Curriculum Committee (UUCC) -**APPROVED**
- **February 26, 2026**- University Committee on Academic Affairs (UC-AA) **APPROVED**
- **March 5, 2026** - University Council (UC) **APPROVED**
- **April 13, 2026** - President's Cabinet (PC)-**APPROVED**
- **April 23-24, 2026** – AHAC/Board of Visitors (BOV)-*Pending Internal Approval*

Proposed New Degree Program Brief
New Degree Program: Master of Science (M.S.) in Pharmaceutical Engineering

Overview
Virginia Commonwealth University (VCU) seeks approval to establish a Master of Science (M.S.) degree program in Pharmaceutical Engineering to be offered at the Monroe Park Campus in Richmond, Virginia. The proposed program will be administered by the College of Engineering and the School of Pharmacy.

The purpose of the proposed M.S. in Pharmaceutical Engineering degree program is to prepare students to participate in areas of health sciences that form the core of therapeutic drug discovery, development and manufacture. Due to the leadership of VCU in pharmaceutical research and education, and the flexible options proposed herein, there is a unique opportunity to establish a M.S. program that will support local pharmaceutical industry needs as well as establish a national and international student population.

Delivery Format
The proposed program will be offered in face-to-face and fully-online delivery formats.

Target Implementation Date
Fall 2027

Demand and Workforce Development
The Bureau of Labor Statistics (BLS) does not track data for Pharmaceutical Engineering occupational titles. Pharmaceutical engineers are listed in the "Engineering, All Other" category–a list of various engineering titles. For this category, the BLS shows that the need for this group is expected to grow 2.1% between 2024 and 2034 (10-year outlook). The Virginia Employment Commission (VEC) suggests an expected 4.1% increase in demand between 2022 and 2032. Similarly, the included Virginia Office of Education Economics (VOEE) report's five-year (5-year) workforce projection suggests a 3.3% growth within VA.

External Competition
No standalone degree program in Pharmaceutical Engineering exists at public institutions in the Commonwealth of Virginia. Virginia Tech offers an M.S/M.Eng. in Chemical Engineering and the University of Virginia offers an M.S./M.E. in Chemical Engineering.

Target Population
The target audience for the proposed degree program includes students who are interested in pharmaceutical engineering work within pharmaceutical laboratories and manufacturing facilities.

Impact on Existing Programs
The proposed degree program will not replace or eliminate any existing degree programs at VCU. There are two (2) similar degree programs at VCU: 1) M.S. in Pharmaceutical Sciences and 2) M.S. in Engineering with a concentration in Chemical and Life Science. There are significant differences between the proposed and existing programs; both the dean of the College

of Engineering and the dean of the School of Pharmacy have expressed their support for this program.

Impact on Faculty

The proposed program will require 1.5 FTE faculty instructional effort in the initiation year, rising to 3.0 FTE by the target year 2029-30. The Department of Pharmaceutics has a total of 14 faculty members. The Department of Chemical and Life Science Engineering has a total of 14 faculty members. Faculty members from both departments will teach the core and required courses for the proposed program.

Funding

No additional funding will be required to initiate or maintain the proposed program. All required courses are already present in the Ph.D. in Pharmaceutical Engineering program. The program will be supported by enrollment services, library support, and career services already in place as well as faculty support services.

Return on Investment

None of the fields of study in the Foundation for Research on Equal Opportunity (FREOPP) database align directly with the proposed degree program. The field that is most closely, but not completely, aligned is "Master's Degree in Pharmacy, Pharmaceutical Sciences, and Administration." FREOPP does not have relevant data for public institutions in Virginia. Return on investment data for graduates in these fields from non-VA institutions has been included as a PDF at the end of this document.

Alignment with the VCU Mission

The proposed degree program aligns with VCU's mission and will train VCU students to participate in areas of "health sciences" that form the core of therapeutic drug discovery, development and manufacture. Graduates will participate in advanced coursework and cutting-edge research under the leadership of VCU faculty and work with PhD program students and post-doctoral scholars to become experts who will contribute to global, national and state efforts in health sciences.

Next Steps

- **February 24, 2026** – University Graduate Council (UGC) **APPROVED**
- **February 26, 2026** - University Committee on Academic Affairs (UC-AA) **APPROVED**
- **March 5, 2026-** University Council **APPROVED**
- **April 13, 2026** - President's Cabinet (PC)- **APPROVED**
- **April 23-24, 2026** – AHAC/Board of Visitors (BOV)-*Pending Internal Approval*

Proposed Program Discontinuance Brief
Program Discontinuance: Discontinue the Bachelor of Science in Education (B.S.Ed.) in Health and Physical Education degree program at Virginia Commonwealth University (VCU).

Overview
Virginia Commonwealth University (VCU) seeks approval to discontinue the Bachelor of Science in Education (B.S.Ed.) in Health and Physical Education (CIP code: 13.1314) degree program. The degree program is located in the School of Education, Department of Teaching and Learning.

Delivery Format
The B.S.Ed. in Health and Physical Education is offered in the face-to-face delivery format.

Target Implementation Date
The program is scheduled for discontinuation in fall 2027.

Rationale
Enrollment trends and demand for the B.S.Ed. in Health and Physical Education have been insufficient to sustain the program. Student enrollment in the degree program has averaged 13 students per year since inception-far below expectations. The School of Education no longer has the full-time faculty required to teach core and required courses. Additionally, student enrollment and graduation data indicate that the degree program would not meet the productivity and viability standards of the State Council of Higher Education for Virginia (SCHEV) when reviewed during the next cycle.

Impact on Students
A teach-out plan is in place for the seven (7) students currently enrolled in the B.S.Ed. in Health and Physical Education degree program. All students are expected to graduate in spring 2030. The last term that students will be able to complete the B.S.Ed. in Health and Physical Education degree program is spring 2031.

Institutional records show that two (2) students have "stopped out" since spring 2025, and those students have been considered. There is a three (3) semester period in which students may return and complete the degree program. All core and required courses will continue to be offered as such courses. The two (2) students will be notified in writing about the discontinuation of the degree program. Faculty advisors will also work with the students to transfer to another degree program, if a degree program is appropriate.

Impact on Faculty and Staff
There is no full-time faculty teaching in the program. The program coordinator for the program is also coordinating the early childhood and elementary programs and will continue to do so after this program's discontinuance.

Funding
No funding is needed or requested for the discontinuance of the degree program.

Next Steps

March 19, 2026 - University Undergraduate Curriculum Committee (UUCC) **APPROVED**
April 2, 2026 – University Committee on Academic Affairs (UC-AA) **APPROVED**
April 9, 2026 – University Council **APPROVED**
April 13, 2026 – President's Cabinet **APPROVED**
April 23-24, 2026 – AHAC/Board of Visitors *Pending Institutional Approval*

Leveraging AI across the VCU enterprise

VCU Board of Visitors

Academic, Health Affairs and Research Committee

Presenters

Milos Manic, Ph.D.

Professor, College of Engineering
Director, VCU Convergence_AI

Preetam Ghosh, Ph.D.

Professor, College of Engineering
Faculty director, High Perfomance Research Computing Core

Sandeep Kothiwale, Ph.D.

Enterprise Data Scientist
VCU Health System Authority

VCU Convergence_AI

VCU Convergence represents what’s possible when higher education breaks down barriers and operates without walls.

Milos Manic, Ph.D.

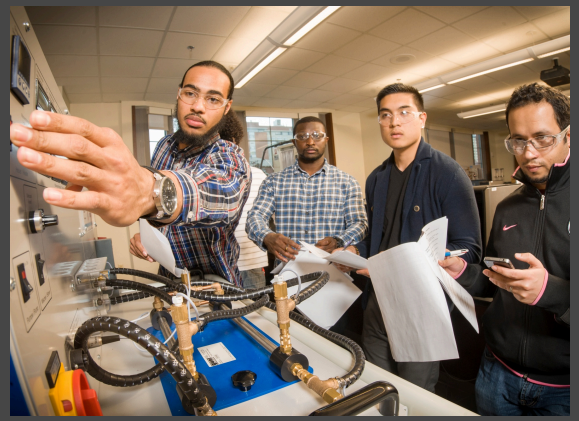
In the last eight months...

Transdisciplinary research

- Seed grants
- DoE Genesis Mission

Interdisciplinary Curricula

- Minors
- Micocredentials (Ethics and Philosophy of AI)
- Applied AI graduate certificate (pending SCHEV approval)



Transformative, experiential learning

- Vertically Integrated Projects (VIPs)
- Axiom Studios

Events and workshops

- Enterprise AI Workshop
- VCU Convergence AI Summit: AI for the Public Good

AI in healthcare research: From multi-omics to clinical data

Preetam Ghosh, Ph.D.

A Multi-Scale Healthcare Matrix

Scale I: Cellular & Molecular

Decoding single-cell spatial multi-omics and predicting synergistic drug targets.

AI Core

Scale II: Clinical Interventions

Transforming massive, heterogeneous Electronic Health Records (EHR) into AI-ready decision support networks.

Scale III: Population Dynamics

Deploying physics-informed machine learning to optimize pandemic management and epidemiological forecasting.

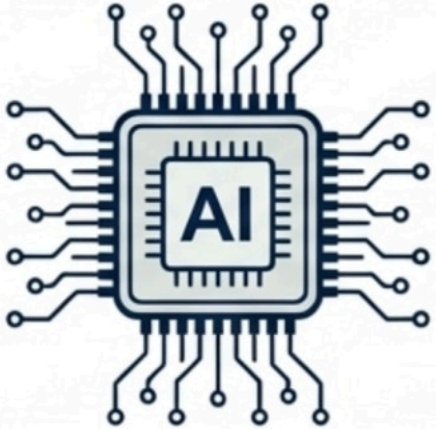
The Translational Pipeline: From Data to Discovery

Pillar I: Decoding Biology

Single-Cell Spatial Multiomics

Understanding cellular heterogeneity, rare transcripts, and gene regulatory networks.

The AI Bridge



Machine Learning & Deep Neural Networks translating biological insights into actionable targets.

Pillar II: Engineering Interventions

AI-Based Drug Design

Predicting drug-target affinity, synergy, and driving drug repurposing.

The Unified AI-Driven Healthcare Ecosystem

Synthesis Statement:

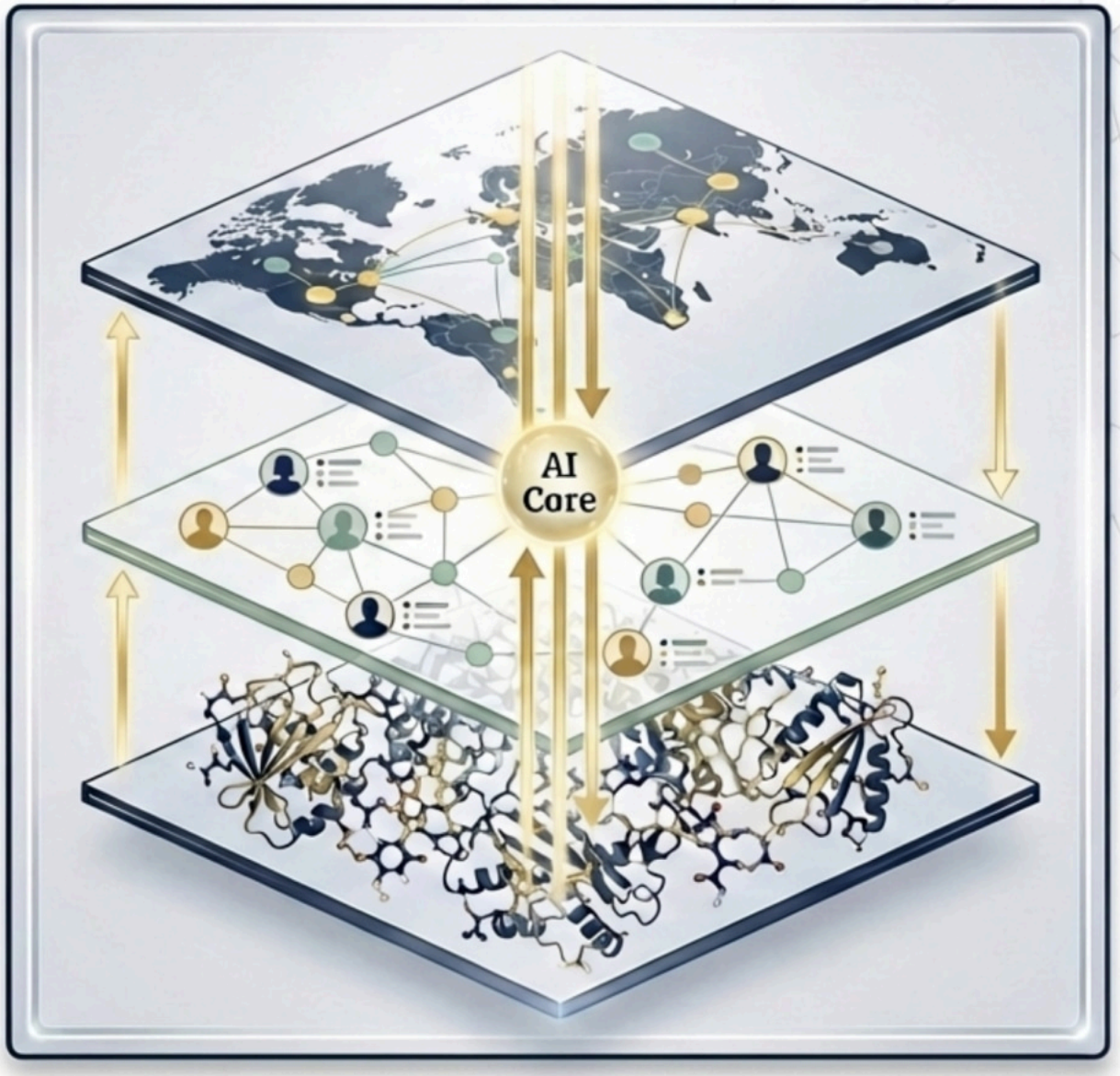
We do not view biology, clinical care, and epidemiology as separate disciplines. They are interconnected data layers that can be handled by modern-AI.

The Feedback Loop:

Population-level spread dynamics (Scale III) inform localized clinical patient similarity graphs (Scale II), which in turn map down to target specific molecular drug synergies (Scale I).

The Advantage:

A holistic, multi-modal systems biology framework capable of modeling human health across the entire continuum.



Expanding the Matrix: Future Frontiers

Telemedicine & Remote Monitoring

Robot-Assisted Surgery Integration

Intelligent Healthcare Chatbots

The translation of multi-omics and big data into tangible health technologies.

Mental Health Support Analytics

Rehabilitation & Physical Therapy Tracking

Non-Pharmaceutical Interventions (Wearables)

AI in healthcare delivery at VCU Health

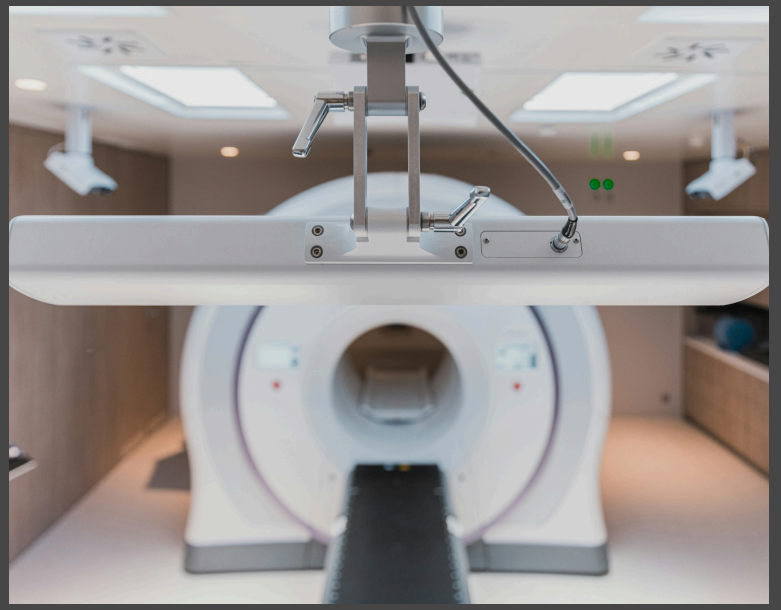
AI is embedded across clinical care, operations, and enterprise workflows, supported by a growing data and governance foundation.

Sandeep Kothiwale, Ph.D.

AI embedded in clinical workflows

EHR (EPIC)

- Predictive and generative AI embedded directly in clinical workflows
- Decision support and documentation assistance at point of care



Imaging (radiology)

- AI for time-critical detection of acute findings
- Workflow prioritization integrated with PACS

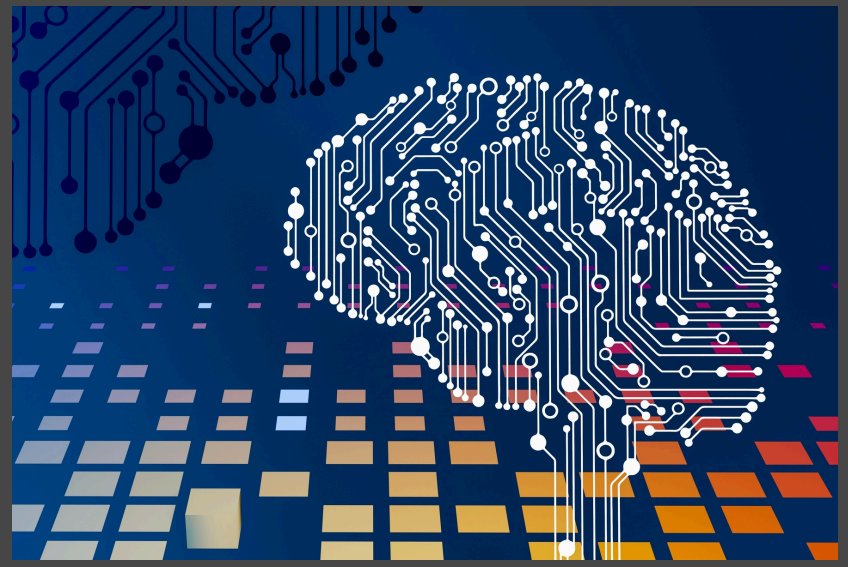
Enterprise clinical tools

- Clinical-facing AI assistance across delivery
- Chart-level insights and task support outside the EHR

AI embedded in the EHR

Capabilities

- Drafting and summarization assistance with human-in-the-loop
- Predictive risk and utilization models
- Workflow-embedded clinical insights and communication



Outcomes

- Enables consistent, scalable decision support
- Reduces documentation burden and variation in practice

Radiology AI and enterprise clinical tools

Radiology AI

- **AI for detection of time-critical findings:** stroke, pulmonary embolism, etc...
- **Workflow prioritization integrated directly with PACS**

Enterprise clinical tools

- **Clinician-facing AI assistance across specialties**
- **Chart-level insights and tasks supported beyond the EHR**

Enterprise data foundation and custom AI

Governed data

Clinical and operational data with quality, provenance, and permitted-use controls

Enterprise data foundation

Unified, secure data foundation supporting analytics and AI

AI products

Predictive models, NLP and agent-based workflows developed in-house

Deployed solutions

Clinical decision support and operational workflow assistance

Monitoring and oversight

Performance monitoring, safety review, and governance alignment

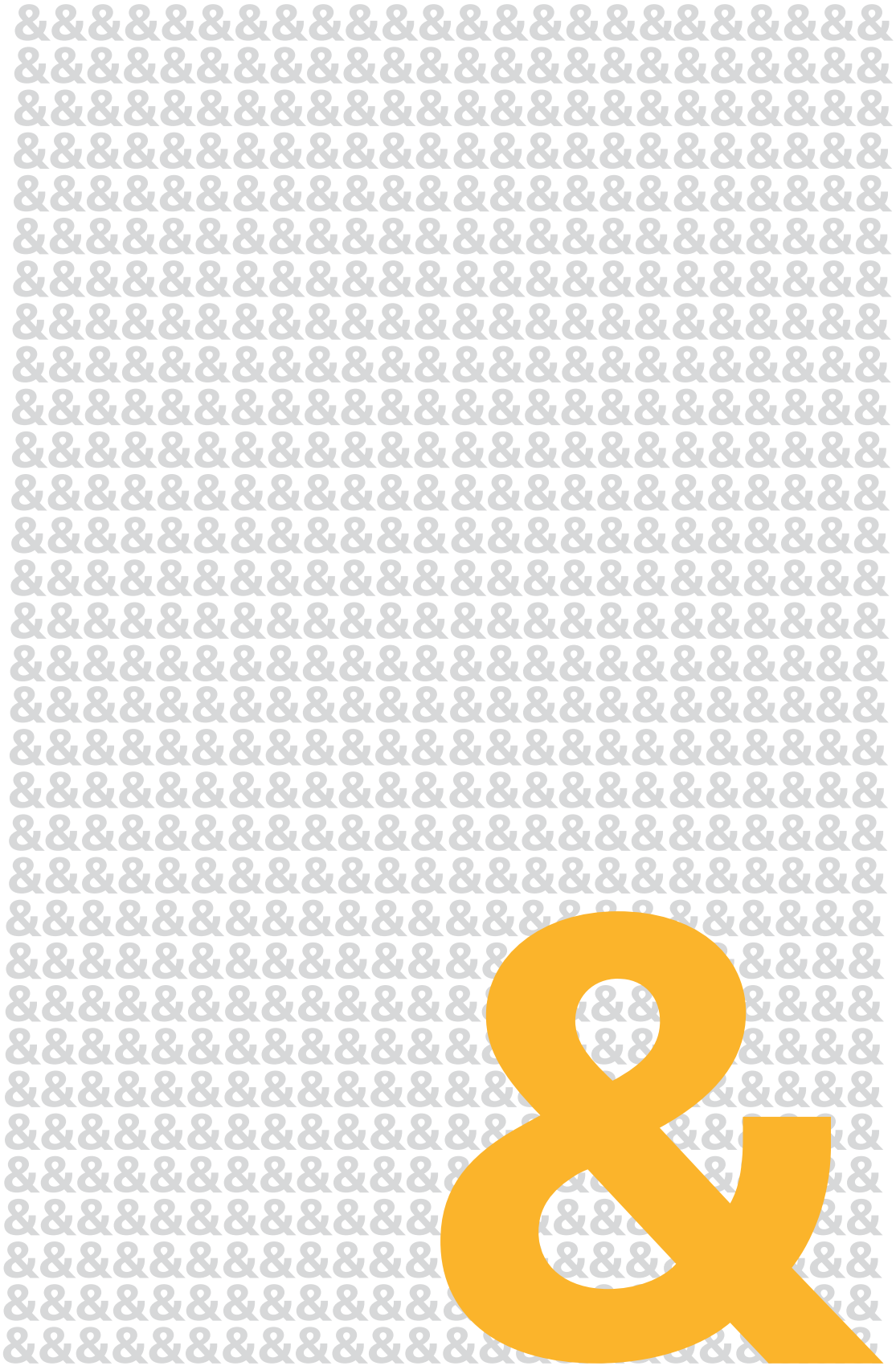
Leveraging AI across the VCU enterprise

VCU Board of Visitors

Academic, Health Affairs and Research Committee

VCU CONVERGENCE_AI PRE-READS

Minor in AI Studies in Humanities and Sciences



VCU College of Humanities and Sciences

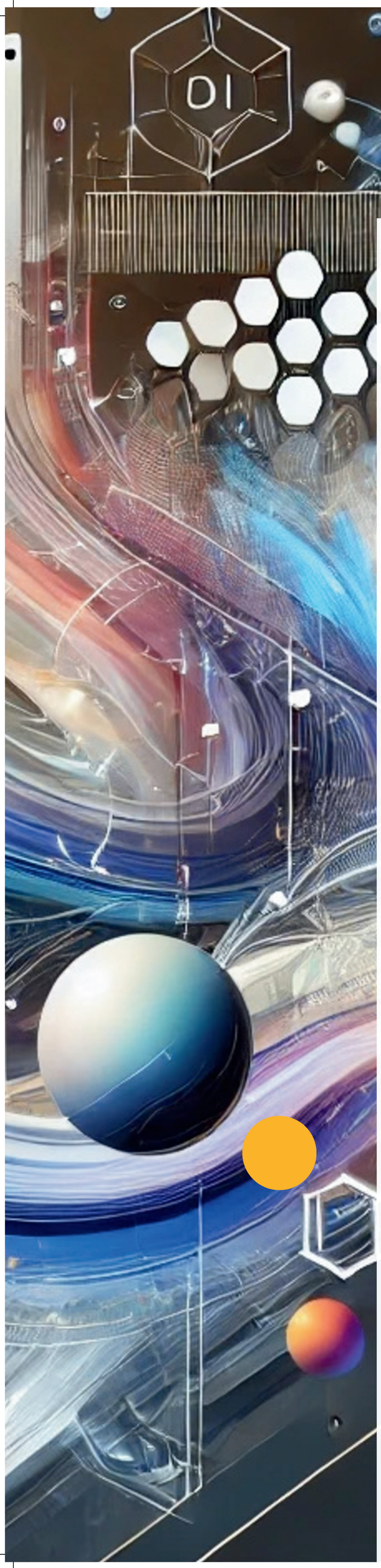
Minor
in AI Studies in Humanties and Sciences

ANTH/FRLG 336	Language, Society, and AI: Sociolinguistic Perspectives
ENGL 350	Digital Rhetoric
ENVS 225	Personal Knowledge Management: Building Your AI-Supported Second Brain
FRLG 392	AI in Intercultural Communication
GSWS 420	Algorithmic Inequalities: Gender, Race and Power in Big Tech
HIST 395	Computer Cultures and AI
HIST 396	Climate Change and AI: A History
HPEX 305	Artificial Intelligence for Health Promotion and Disease Prevention
HUMS 204	AI Literacy
HUMS 392	Topics in AI
LFSC 501	AI in Life Sciences
MASC 311	AI in Mass Media
MASC 312	Creativity With AI
MASC 313	Strategic Communication With AI
MASC 314	Storytelling With AI
MATH 170	The Language of Artificial Intelligence
MATH 310	Linear Algebra
MATH 370	Mathematical Foundations for Artificial Intelligence
NEXT 275	AI, Texts and Cultures
NEXT 375	Critical AI Studies
PHIL 202	Ethics of Artificial Intelligence
PHIL 332	Philosophy of Artificial Intelligence
POLI 393	AI Governance and Ethics
PHYS 220	Introduction to Complexity Science in the AI Era
SOCY 301	AI, Automation and Work
STAT 355	Data Science for AI
STAT 421	Statistical Computing for Machine Learning and Artificial Intelligence

Teachers and guest speakers wanted!

Interested in guest lecturing or teaching an AI-focused course in the future? **Come teaching with us.** We are always looking for knowledgeable, forward-thinking collaborators. Share your information with us, and we will follow up about opportunities to get involved.

VCU College of Humanities and Sciences



Here is where creators, educators, and innovators shape what’s next — together.

The Media & AI Initiative is hands-on, industry-connected, and ethically grounded. We empower smart storytellers to create real impact.

We’re builders

We partner with industry to design solutions, lead workshops, and equip organizations navigating AI transformation. We author thought leadership and conduct original research that advances the field. We build AI classes across undergraduate, graduate, and professional programs, design curriculum, and train faculty.

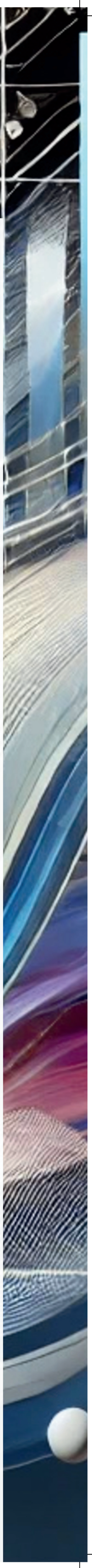
We don’t just study the future of media and AI. We build it.

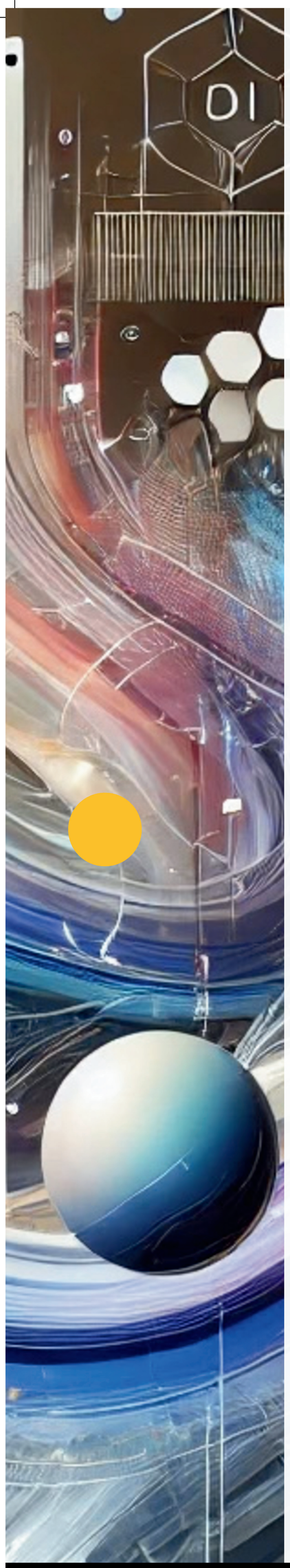
We’re different

We are backed by a top R1 research university — deeply rooted in the humanities and sciences with access to business, political science, education, health care, and more.

Media+AI

mediaandai.org





Media+AI

We are

Builders.
Collaborators.
Educators.
Innovators.
Leaders.
Practitioners.
Researchers.

You can

Teach as adjunct faculty.
Guest lecture.
Design and lead workshops.
Co-create research studies.
Write and publish.
Pursue grants.
Prototype tools.
Launch pilot projects.
Mentor students.
Sponsor challenges.
Bridge industry and classroom.
Advise on ethics.
Test emerging platforms.
Host labs.
Shape curriculum.
Build cross-sector partnerships.

You should Join us.

Sign up.
Get connected.
Shape what's next.

mediaandai.@vcu.edu

mediaandai.org

AI IN HEALTHCARE RESEARCH PRE-READS

AI in Healthcare: From Multi-Omics to Clinical Data

Preetam Ghosh, Professor of Computer Science,
Faculty Director: High Performance Research Computing Core

VCU College of Engineering | Department of Computer Science

VCU College of Engineering

A Multi-Scale Healthcare Matrix

Scale I: Cellular & Molecular

Decoding single-cell spatial multi-omics and predicting synergistic drug targets.

AI Core

Scale II: Clinical Interventions

Transforming massive, heterogeneous Electronic Health Records (EHR) into AI-ready decision support networks.

Scale III: Population Dynamics

Deploying physics-informed machine learning to optimize pandemic management and epidemiological forecasting.

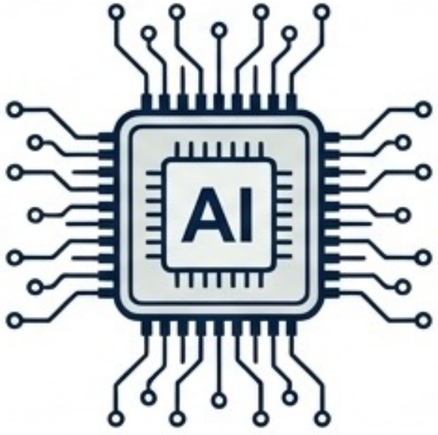
The Translational Pipeline: From Data to Discovery

Pillar I: Decoding Biology

Single-Cell Spatial Multiomics

Understanding cellular heterogeneity, rare transcripts, and gene regulatory networks.

The AI Bridge



Machine Learning & Deep Neural Networks translating biological insights into actionable targets.

Pillar II: Engineering Interventions

AI-Based Drug Design

Predicting drug-target affinity, synergy, and driving drug repurposing.

Scale I: Overcoming Single-Cell Transcriptomics Bottlenecks

High Technical Dropouts: Low RNA capture obscures rare transcripts.

Solution: MOGClass

Stabilizes sparse single-cell data by aligning clusters with bulk RNA-seq co-expression patterns using Graph Neural Networks.

Annotation Ambiguity: Manual marker annotation is subjective and imprecise.

Solution: CORTADO

Applies hill-climbing optimization to maximize differential expression and enforce sparsity, surpassing baseline methods.

Computational Expense: Processing massive datasets yields highly variable networks.

Solution: COFFEE & CHAI

Utilizes a 'Wisdom of Crowds' consensus clustering approach and Borda count voting to map single composite Gene Regulatory Networks.

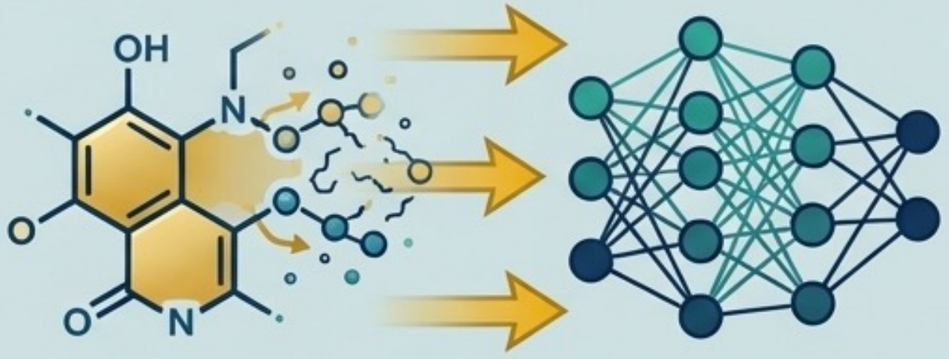
Scale I: The Drug Discovery Paradigm Shift

The Legacy Approach

Restricted by physical trial-and-error screening. Expensive and highly prone to clinical failure.

The AI Alternative

The GramSeq-DTA architecture. A quantitative deep learning framework fusing gene expression information for analyzing Drug-Target Binding (DTB) and Drug-Target Affinity (DTA).



The Scale

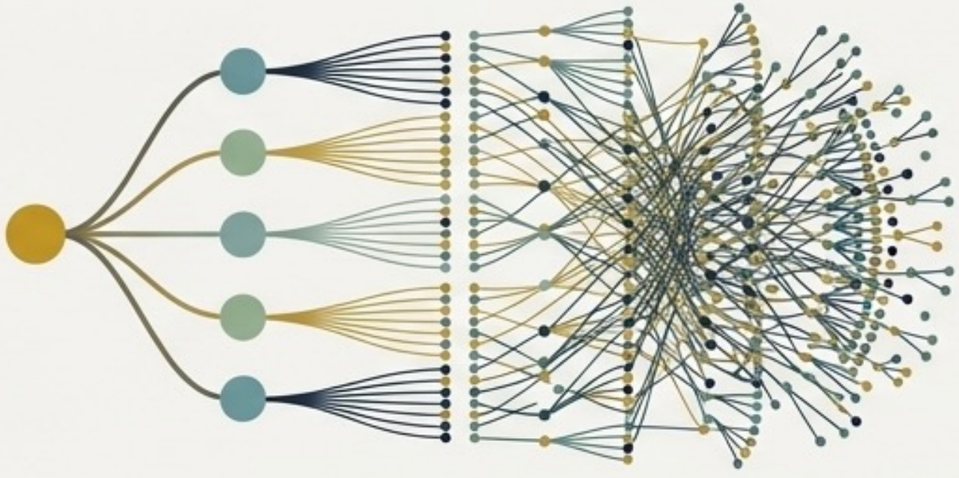
AI models evaluate libraries of ~8 million compounds against human proteome targets in days, drastically reducing clinical trial failure risks.

Scale I: The Next Frontier in Drug Synergy Prediction

The Combinatorial Explosion

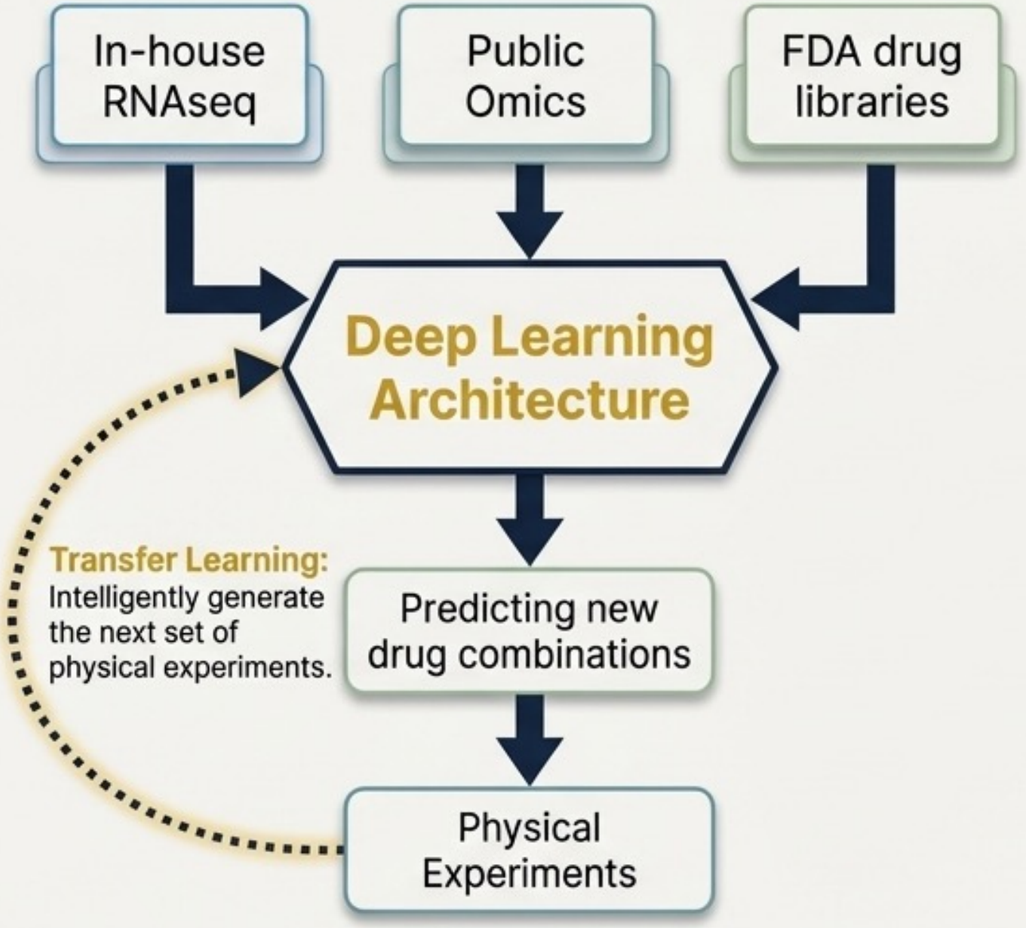
Monotherapies often fail against complex diseases. We require combination therapies, but experimental validation across diverse cell lines is statistically impossible to achieve manually.

1 + 1 = 3 (Synergistic)
1 + 1 = 2 (Additive)
1 + 1 = 0.5 (Antagonistic)



The Prediction Architecture

Data-to-Discovery Loop



Scale II: Engineering AI-Ready Clinical Data Pipelines

Care Delivery Data

CPRS
Patient Info

TPS
Dosimetry

TMS
Treatment Data

HINGE System

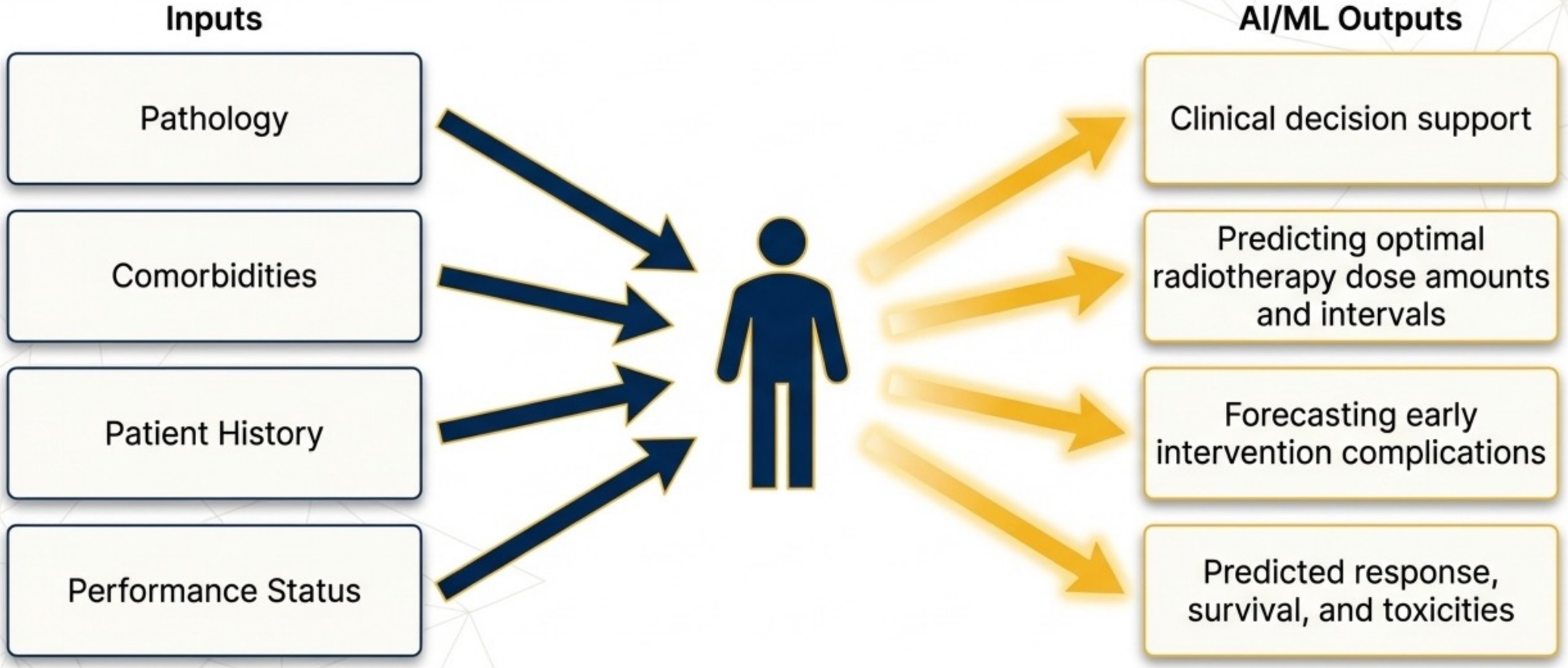
Automated extraction, transformation, loading (ETL), and Natural Language Processing (NLP) mapping clinical data to standard ontologies.

RO-CDW Core Data

An integrated RDF Graph-Database structuring demographics, survival details, DVH curves, and lab values, making massive US-VHA datasets AI-ready for continuous quality improvement.

Scale II: Precision Radiotherapy & Outcomes Modeling

How best to treat the individual patient across 40 US VHA centers nationwide?



Scale II: Mapping Temporal Patient Similarity Graphs

The Framework

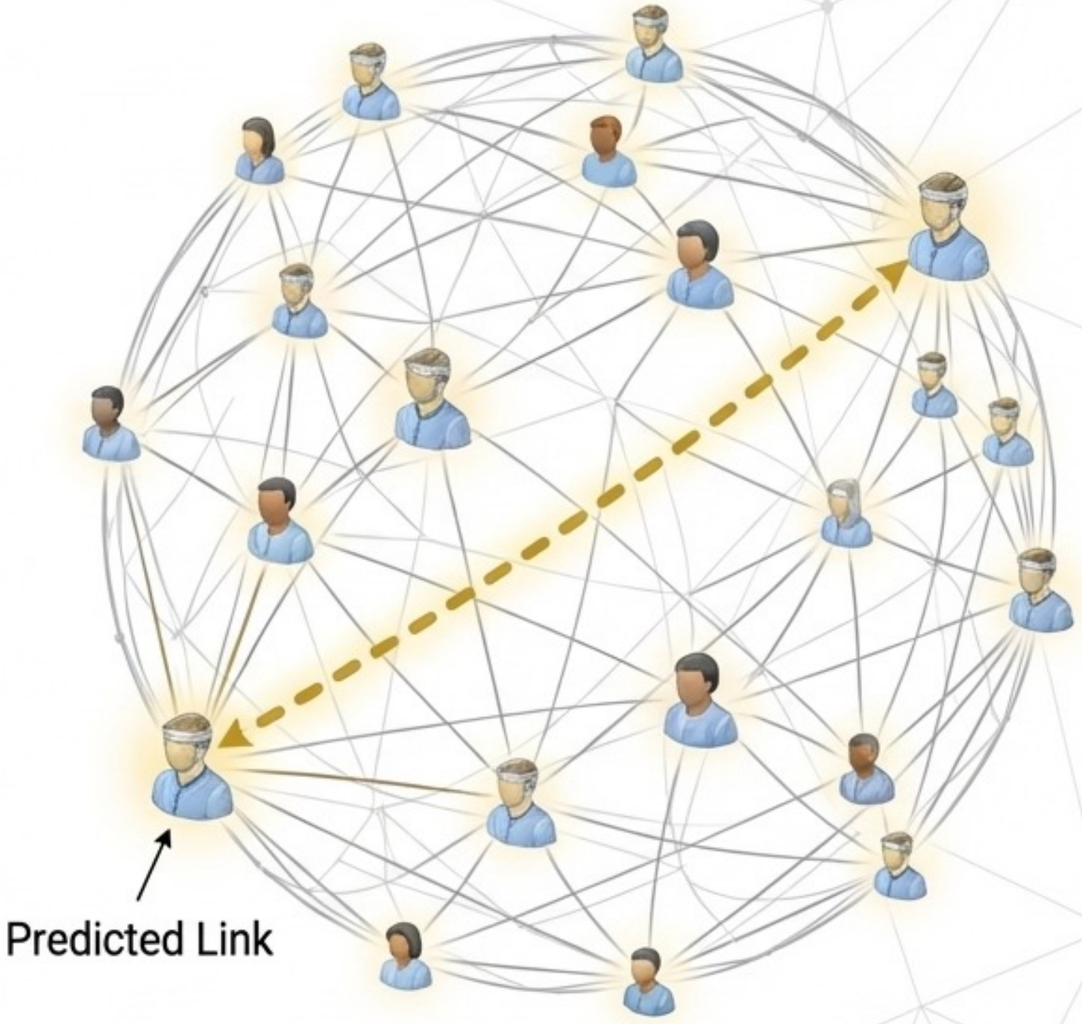
Modeling heterogeneous clinical data (EHR) using Heterogeneous Graph Neural Networks.

The Mechanism

Given a Graph G = (V, E), the system uses the Training Graph (G_T) to predict future disease progression (E_P) based on the assortativity of complex networks.

The Result

Lightly scaled models that create patient-specific similarity networks, allowing practitioners to track and predict disease subtyping and survival rates over time.



Scale III: Population Dynamics & Pandemic Management

An online unified pandemic management architecture requiring multiple modalities of models and data.

Vaccine Allocation

- Demographic data
- Health info

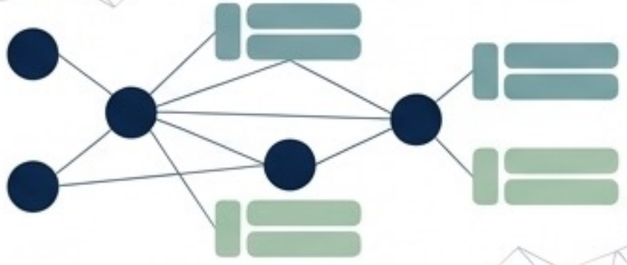
Lockdown Health

- Infection rates
- Economy

Mobility Scheduling

- Human traffic patterns
- Urban planning

Central Knowledge Graph



Predictive Outcome Indicators

- Infected
- Recovery Rates
- Health
- Economic Impact

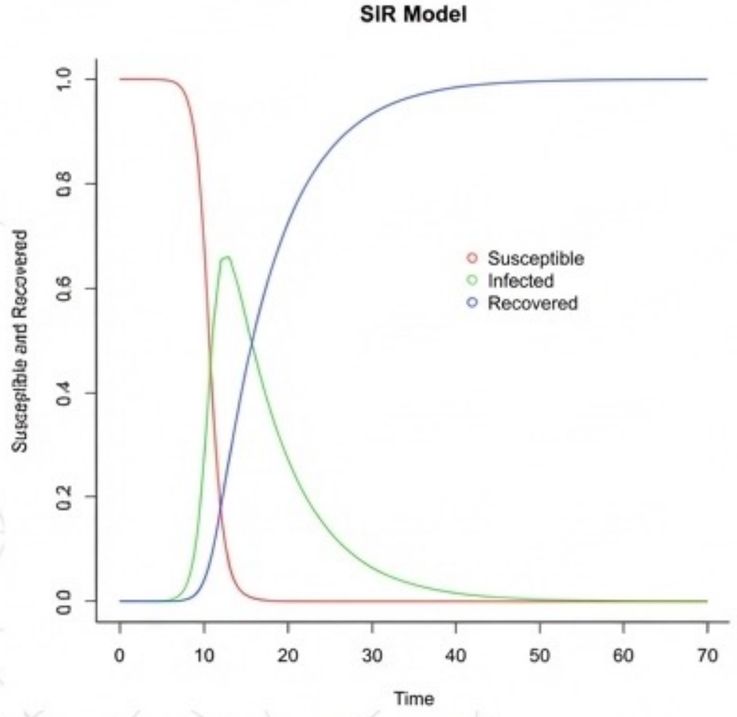
Keeping the "Man in the Loop" for final decision-making.

Scale III: The PIML Panacea vs. Standard SIR Models

Traditional Compartmental SIR

Relies on static governing differential equations (dS/dt, dI/dt, dR/dt).

Fails to adapt to long-term population dynamics, healthcare shifts, and geographic variances.



$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

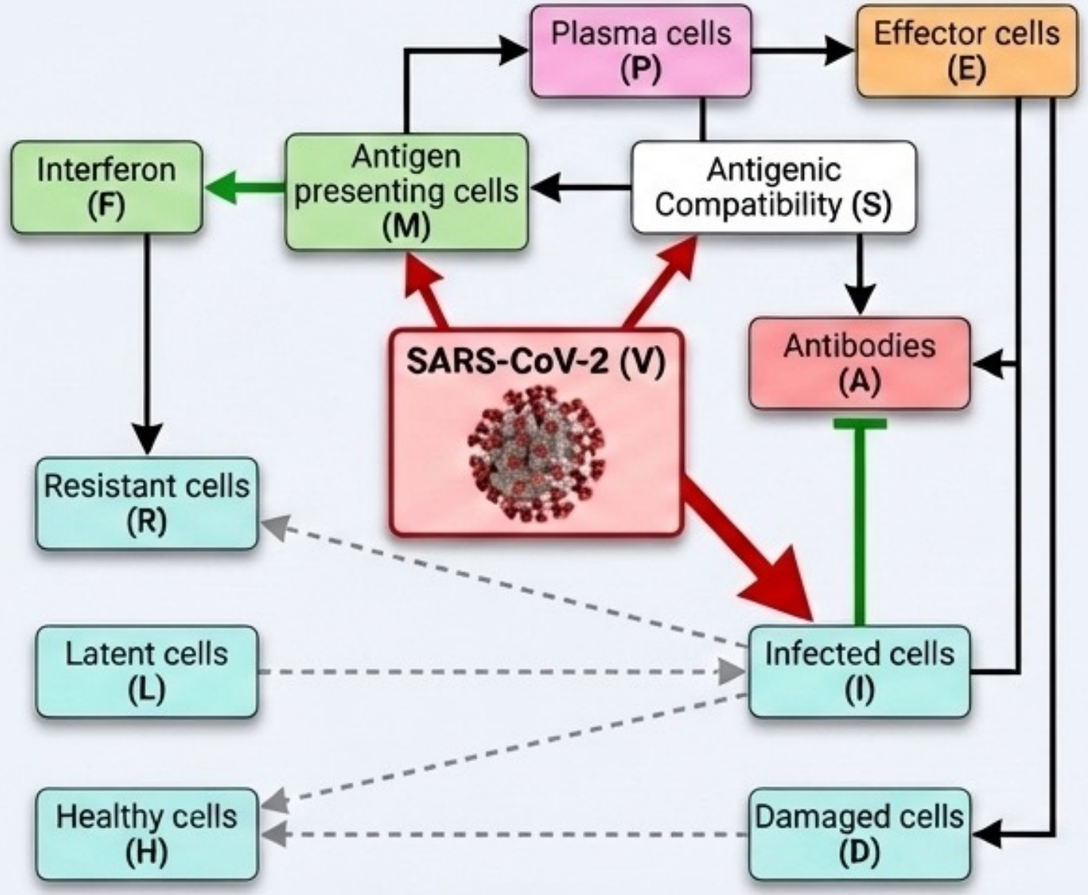
Physics-Informed Machine Learning (PIML)

- **Automated Parameter Estimation**: Uses SIRPINN to automatically estimate and tune transmission (β) and recovery (γ) rates across regions.
- **Adaptive Scaling & Clustering**: Employs K-Means and Agglomerative clustering to group geographic areas with similar epidemic patterns in real-time.
- **Cross-Region Dynamics**: Handles nonlinearities in disease spread data far better than static equations.

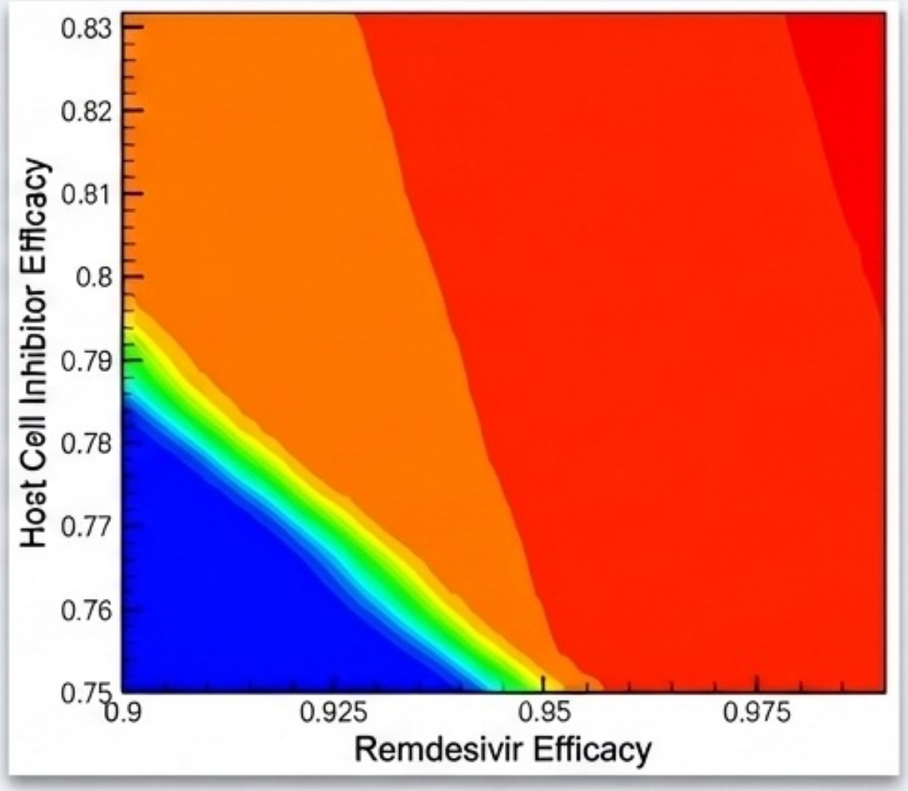
Scale III: Systems Biology of Pathogen Spread

Biological Network Modeling: Utilizing spatio-temporal, stochastic, and ODE models to map the exact pathways of infectious diseases.

The Interaction Matrix



Therapeutic Simulation



Generating precise stability analyses and drug combination predictions to inform large-scale epidemiological interventions.

The Unified AI-Driven Healthcare Ecosystem

Synthesis Statement:

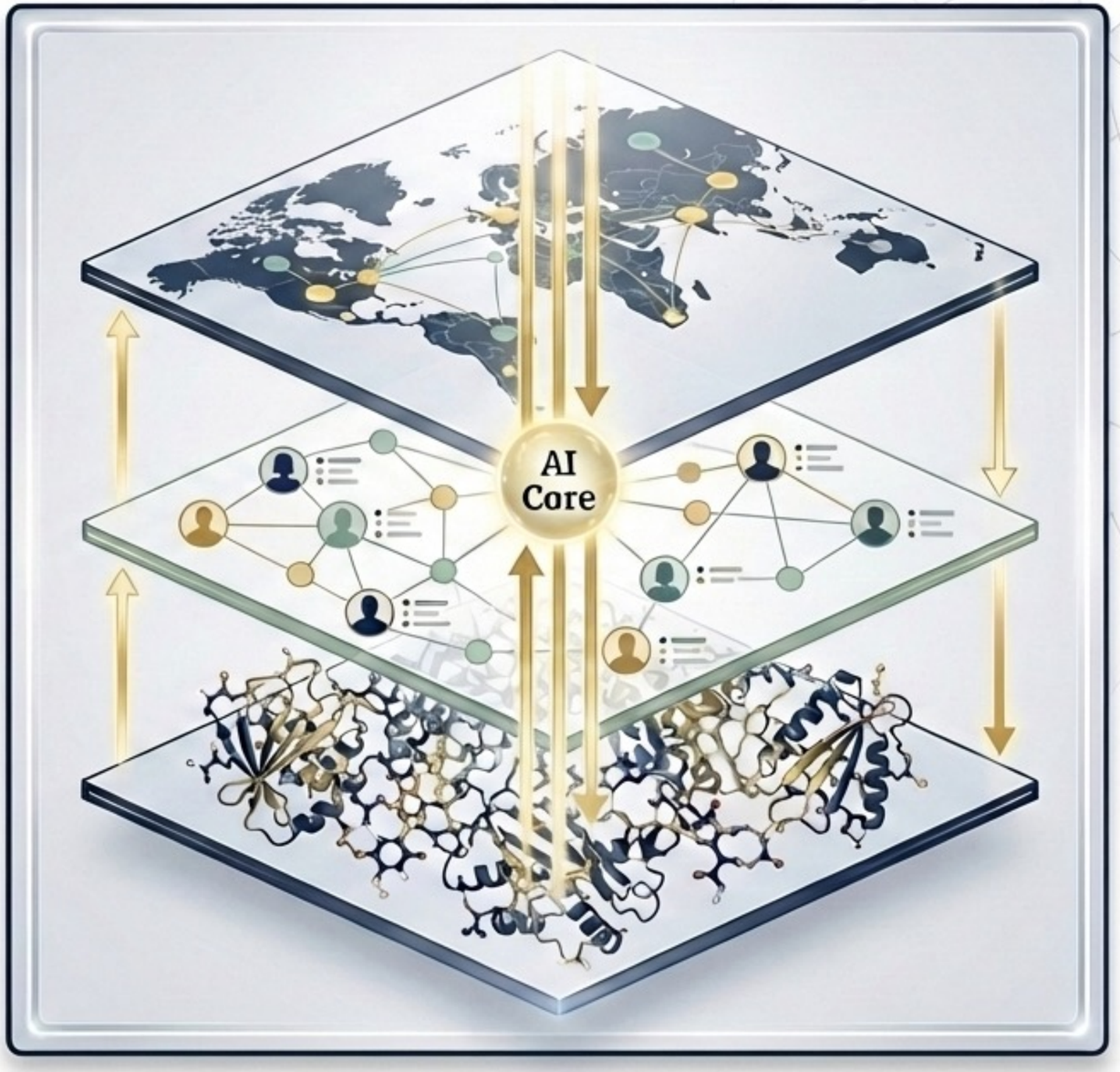
We do not view biology, clinical care, and epidemiology as separate disciplines. They are interconnected data layers that can be handled by modern-AI.

The Feedback Loop:

Population-level spread dynamics (Scale III) inform localized clinical patient similarity graphs (Scale II), which in turn map down to target specific molecular drug synergies (Scale I).

The Advantage:

A holistic, multi-modal systems biology framework capable of modeling human health across the entire continuum.



Expanding the Matrix: Future Frontiers

Telemedicine & Remote Monitoring

Robot-Assisted Surgery Integration

Intelligent Healthcare Chatbots

The translation of multi-omics and big data into tangible health technologies.

Mental Health Support Analytics

Rehabilitation & Physical Therapy Tracking

Non-Pharmaceutical Interventions (Wearables)

nature reviews genetics

https://doi.org/10.1038/s41576-023-00580-2

Review article

Check for updates

Methods and applications for single-cell and spatial multi-omics

Katy Vandereyken[1,2,3,4], Alejandro Sifrim[1,2,3,4], Bernard Thienpont[1,2,3,4] & Thierry Voet[1,2,3]

Abstract

The joint analysis of the genome, epigenome, transcriptome, proteome and/or metabolome from single cells is transforming our understanding of cell biology in health and disease. In less than a decade, the field has seen tremendous technological revolutions that enable crucial new insights into the interplay between intracellular and intercellular molecular mechanisms that govern development, physiology and pathogenesis. In this Review, we highlight advances in the fast-developing field of single-cell and spatial multi-omics technologies (also known as multimodal omics approaches), and the computational strategies needed to integrate information across these molecular layers. We demonstrate their impact on fundamental cell biology and translational research, discuss current challenges and provide an outlook to the future.

Sections

[1]KU Leuven Institute for Single Cell Omics (LISCO), University of Leuven, KU Leuven, Leuven, Belgium. [2]Department of Human Genetics, University of Leuven, KU Leuven, Leuven, Belgium. [3]Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD, USA. [4]These authors contributed equally: Katy Vandereyken; Alejandro Sifrim; Bernard Thienpont. e-mail: thierry.voet@kuleuven.be

Introduction

Humans and many other Eukaryota are composed of billions of cells, belonging to vastly heterogeneous cell types and functional cell states determined by both cell-intrinsic and cell-extrinsic factors. Intrinsically, there is a complex interactive molecular hierarchy of the different 'omics' layers within a cell: from genome and epigenome to transcriptome, proteome and metabolome, and back. Extrinsically, the functional state of a cell can be modulated by its neighbouring cells through direct physical interaction (such as receptor–ligand interactions), through signalling molecules secreted by one cell that can act through receptors on remote cells (such as morphogen signalling pathways), or by other microenvironmental factors (such as chemical compound gradients)[1]. Consequently, investigating how multicellular organisms develop from a totipotent single cell and subsequently function, age and develop disease, necessitates single-cell and spatial multi-omics approaches (also known as multimodal omics approaches).

Robust technologies for unimodal (mono-omics) measurements of individual cells, such as single-cell RNA sequencing (scRNA-seq) methods[2], have already evolved to revolutionize the discovery and understanding of cell types as well as their different functional cell states, cell plasticity upon exposure to external stimuli and drugs, and cell differentiation or reprogramming trajectories[3]. The power of these technologies is underscored by the instigation of the Human Cell Atlas (HCA)[1,4] and other consortium-based resources afterwards, which are primarily aimed at creating cellular reference maps of organisms, including the position, function and characteristics of every cell type. However, to develop fundamental understanding of the molecular hierarchy from genome to phenome in individual cells, multi-omics methodologies at single-cell and spatial resolution are necessary. They enable investigation of the intermolecular dynamics between gene regulation on the epigenome level and gene expression on the transcriptome and/or proteome levels unambiguously in the same single cells across development, ageing and disease. Additionally, these technologies enable investigating the impact of acquired genetic variation in the genome of single cells on their own function and phenotypic features, as well as (surrounding) tissue function and more, as exemplified throughout this Review. In recent years, this field has advanced phenomenally and is rapidly maturing both technologically and computationally, enabling broad applications to understand cell biology (Fig. 1).

In this Review, we discuss the fundamental technological and computational principles, state of the art and applicative value of modern single-cell and spatial multi-omics. We focus on methods that provide a comprehensive 'omics' view of at least one molecular analyte, and for single-cell multi-omics are based on next-generation sequencing (NGS), whereas for spatial multi-omics we highlight both NGS-based and imaging-based methods. Although many algorithms have been tailored to analyse the individual molecular layers, here we primarily focus on computational techniques for the integration of information across the different data modalities to maximally leverage the potential of these multi-omics technologies. We end with an outlook to the future.

Single-cell multi-omics methods

Multi-ome measurements from single cells are enabled by different methodological approaches, which can be categorized according to whether the distinct molecular analytes are uncoupled before, during or after sequencing library preparation. As described in Fig. 2 and further below for specific methods, each of these different principles comes with specific advantages and limitations.

Single-cell genomics-plus-transcriptomics

Soon after the establishment of single-cell DNA sequencing (scDNA-seq) and single-cell RNA sequencing (scRNA-seq) protocols[2,5], methods for genome-plus-transcriptome sequencing of individual cells were developed (Fig. 1) that rely on one of the four basic principles for multi-ome analysis (Fig. 2).

A first set of methods applies physical separation of DNA and RNA before sequencing library preparation (Fig. 2a). In G&T-seq[6,7] (Supplementary Fig. 1a), oligo-dT bead-mediated precipitation of polyadenylated (poly(A)) RNA molecules enables their physical separation from the remaining molecules in the cell's lysate, including the nuclear and mitochondrial DNA, either manually or robotically. Captured transcripts are then on-bead primed for full-length cDNA amplification through reverse transcription (RT), template switching and PCR using a Smart-seq2-like reaction[8], allowing for both short-read and long-read sequencing[7]. Long-read sequencing is preferred for transcript isoform detection. The genomic DNA (gDNA) in the collected supernatant is subjected to whole-genome amplification (WGA) using a method of choice: multiple displacement amplification (MDA), PCR or displacement preamplification followed by PCR (DA-PCR). Single-cell transcriptogenomics applies a similar principle as G&T-seq and is compatible with targeted DNA exome sequencing following MDA[9]. The freedom of choice for downstream processing of the separated poly(A) RNA and gDNA presents a major advantage. Indeed, WGA is not error-free, and different WGA methods present different biases, making some more suitable for the detection of specific classes of genetic variants, as reviewed previously[10,11]. Similarly, different scRNA-seq methodologies present different performances in sensitivity, transcript coverage and throughput[12].

As alternatives to oligo-dT bead-based separation of poly(A) RNA and gDNA, multiple methods partition cytosolic RNA from nuclear DNA using a two-step cell lysis (Fig. 2b). In SIDR-seq[13] (Supplementary Fig. 1b), single cells are first subjected to hypotonic lysis, enabling the nucleus to be isolated from the supernatant containing RNA using antibody-conjugated magnetic microbeads. Then, the nucleus is lysed and subjected to MDA-based scDNA-seq, and the supernatant RNA is subjected to Smart-seq2-based scRNA-seq. In DNTR-seq[14] (Supplementary Fig. 1c), the nuclei of single cells are precipitated by centrifugation after cell membrane lysis. While the supernatant cytosol is isolated for Smart-seq2-like scRNA-seq, the nuclear DNA is subjected to direct tagmentation. The latter enables direct PCR-based library preparation for scDNA-seq, thereby circumventing the classic approach of WGA before scDNA-seq library preparation and in part the resulting artefacts associated with it[14,15]. However, these nuclear–cytosolic partitioning methods are less amenable to comprehensive characterization of mitochondrial DNA and nuclear RNA, and they are confined to the use of intact cells as input. Furthermore, plate-based assays such as G&T-seq, transcriptogenomics, SIDR-seq and DNTR-seq are inherently low throughput and, despite (partial) automation on liquid-handling robotics, still require up to a few days' time[6]. To circumvent this, integrated on-chip microfluidics approaches have been devised that enable separation and parallel processing of cytosolic RNA and nuclear DNA[16,17]. Such systems largely avoid the introduction of operator bias, can miniaturize the reactions and hence reduce reagent costs, and have the capacity to increase throughput, although this remains to be shown.

Another set of methods relies on a preamplification-and-split approach for uncoupling DNA and RNA molecular analytes (Fig. 2c). In DR-seq[18] (Supplementary Fig. 1d), poly(A) RNA-derived first-strand cDNA is subjected to quasilinear amplification together with the gDNA

in a single tube. This reaction is then split, with both aliquots containing preamplified cDNA and gDNA. In one, the preamplified gDNA is further PCR-amplified and converted to a sequencing library, with the caveat that contaminating cDNA will be co-amplified and sequenced into reads indistinguishable from the gDNA sequences. In the sister reaction, the cDNA is further amplified using in vitro transcription (IVT), followed by RT and PCR. Because only the cDNA is tagged with T7 promoter sequence for IVT, gDNA will not be co-amplified, resulting in a 3′-end scRNA-seq library. In the more recently developed TARGET-seq[19], a method for targeted mutation detection and parallel transcriptome characterization of the same single cells, poly(A) cDNA is co-amplified with a mutation of interest targeted by cDNA primers and gDNA primers in a single reaction. This pot is then split to prepare the genotyping and transcriptome sequencing libraries separately. Although DR-seq and TARGET-seq minimize the risk of losing (deoxy)ribonucleic acids in comparison to above-described methods relying on physical separation of DNA and RNA, they can suffer from RNA-derived reads contaminating the gDNA analysis, and are also plate-based and low throughput.

Instead, scONE-seq[20] (Supplementary Fig. 1e) follows the principle of seq-split (Fig. 2d) and differentially barcodes gDNA and RNA through, respectively, tagmentation with a 6-nucleotide DNA-barcode-containing adaptor and reverse transcription with a 6-nucleotide RNA-barcode-containing RT-primer. Differentially labelled gDNA and cDNA is then co-amplified and converted to a sequencing library in a single-tube reaction. Following NGS, gDNA-derived and RNA-derived reads are distinguished by their barcode sequence. This methodology hampers sequencing RNA-seq and DNA-seq libraries separately to optimal depths and is plate-based and low throughput, but indicates that relatively simple one-tube reactions are possible for multi-omics measurements of single cells.

Recently, a highly scalable plate-based technology was established based on the principle of single-cell combinatorial indexing (Fig. 2e) with a three-level indexing scheme and combined with linear IVT-based amplification of the cells' genome and transcriptome (Supplementary Fig. 1f). This sci-L3-RNA/DNA co-assay enables analysis of at least tens of thousands of single nuclei, with the possibility to increase the throughput to more than 1 million cells profiled per experiment[21]. Proof of concept of sci-L3-RNA/DNA was presented on mixtures of male mouse and female human cell lines, proving that the single-cell transcriptomes were organized into the two expected cell clusters and that the matching single-cell genomes were of the correct sex, although high-sensitivity and high-resolution profiling per cell was not demonstrated.

The study of both the genome and the transcriptome of the same cell enables one to unambiguously investigate the impact of acquired DNA mutations, such as DNA copy number aberrations, on gene expression in the same cell. This has important applications for understanding intratumoural heterogeneity, enabling the investigation of the development of different phenotypic cancer cell states among the different genetic subclones that arise, or even within a single genetic subclone. For instance, DNTR-seq identified minor subclones having genetic copy number alterations with associated transcriptional perturbations in paediatric acute lymphoblastic leukaemia[14]. Additionally, transcriptional signatures of the WNT pathway activation learned from scRNA-seq could be explained by mutations detected in the scDNA-seq data from the same cells[17]. Furthermore, using tumour model systems exposed to treatment, or direct longitudinal sampling of patient tumour specimens before and during treatment, and analysing them by single-cell genome-plus-transcriptome sequencing will allow investigation of which genetic subclones are more fit to tolerate the drug selection. Additionally, it will allow the study of how cells within these genetic subclones putatively apply cell plasticity to change their gene expression repertoire and accommodate different phenotypic cancer cell states able to withstand drug treatment and, eventually, acquire resistance[22]. In turn, these approaches might enable the identification of potential cancer cell vulnerabilities, such as druggable molecular players involved in the acquisition of drug tolerance.

Beyond the field of oncology, these multi-omics methods are important for understanding the pathogenesis of other disorders in which somatic genetic variation plays a putative role, including in neurological disorders such as Alzheimer disease, Parkinson disease and others[23–25]. With the recent discovery that normal tissues are also subjected to an extraordinary amount of mutation, technologies enabling the analysis of the genome and transcriptome of the same single cells will be important to study the role or impact of acquired mutations on phenotypic and functional cellular states, and how these in turn impact development[26], cellular competition between normal and genetically aberrant cells[27], tissue homeostasis, normal phenotypic variation and ageing[28]. The power of multi-omics methods in this respect is illustrated by the detection of different transcriptional responses to acquired DNA copy number aberrations[6,7,14,18] and other forms of mutations[9]. For instance, it was shown using DNTR-seq that structural DNA imbalances lead to both linear and nonlinear transcriptional dosage effects, whereby several genes important for cancer cell growth, such as *MYC* and *TCF7L2*, demonstrated strong dosage compensation and were shown to be mostly unaffected by copy number alterations[14].

Single-cell genome-plus-transcriptome sequencing is also a valuable tool to study the efficacy and safety of genome editing in germline therapy. CRISPR–Cas9 genome editing has potential as a therapeutic tool for the correction of disease-causing mutations. Genome editing of human embryos or germ cells provides the means for introducing heritable genetic alterations, which may reduce the burden of genetic disease in specific familial situations[29]. Its use is currently a hot topic of international debate around ethics, safety and efficacy. Single-cell genome-plus-transcriptome sequencing approaches will be pivotal to assess on-target and off-target genome edits plus cell phenotypic consequences, as only a few cells are available for analysis from the treated and subsequently in vitro-cultured human embryo. Recently, *OCT4* (also known as *POU5F1*) CRISPR–Cas9-targeted and control human preimplantation embryos were investigated using single-cell G&T-seq[30,31] as well as single-cell or low-input mono-omics DNA-seq. Regions of loss of heterozygosity in genome-edited cells that spanned beyond the *OCT4* on-target locus, as well as segmental loss and gain of the *OCT4*-containing chromosome 6, were detected in the genomic data, which collectively resulted in unintended genome edits being identified in ~16% of the human embryo cells analysed. The transcriptome data suggested that the loss of heterozygosity does not lead to the misexpression of other genes adjacent to the *OCT4* locus.

Furthermore, genome-plus-transcriptome sequencing approaches allow genetic variation detected in DNA sequences to be confirmed in the RNA sequences of the same cell, increasing the reliability of the genotyping call[7,9,13]. This principle has been shown for the detection of single nucleotide variants contained within expressed genes, forms of structural variation resulting in the expression of fusion genes, and copy number variants resulting in gene expression dosage effects. The genomic and mitochondrial DNA variants detected by genome-plus-transcriptome sequencing approaches, considering potential imperfections[15], can furthermore be leveraged

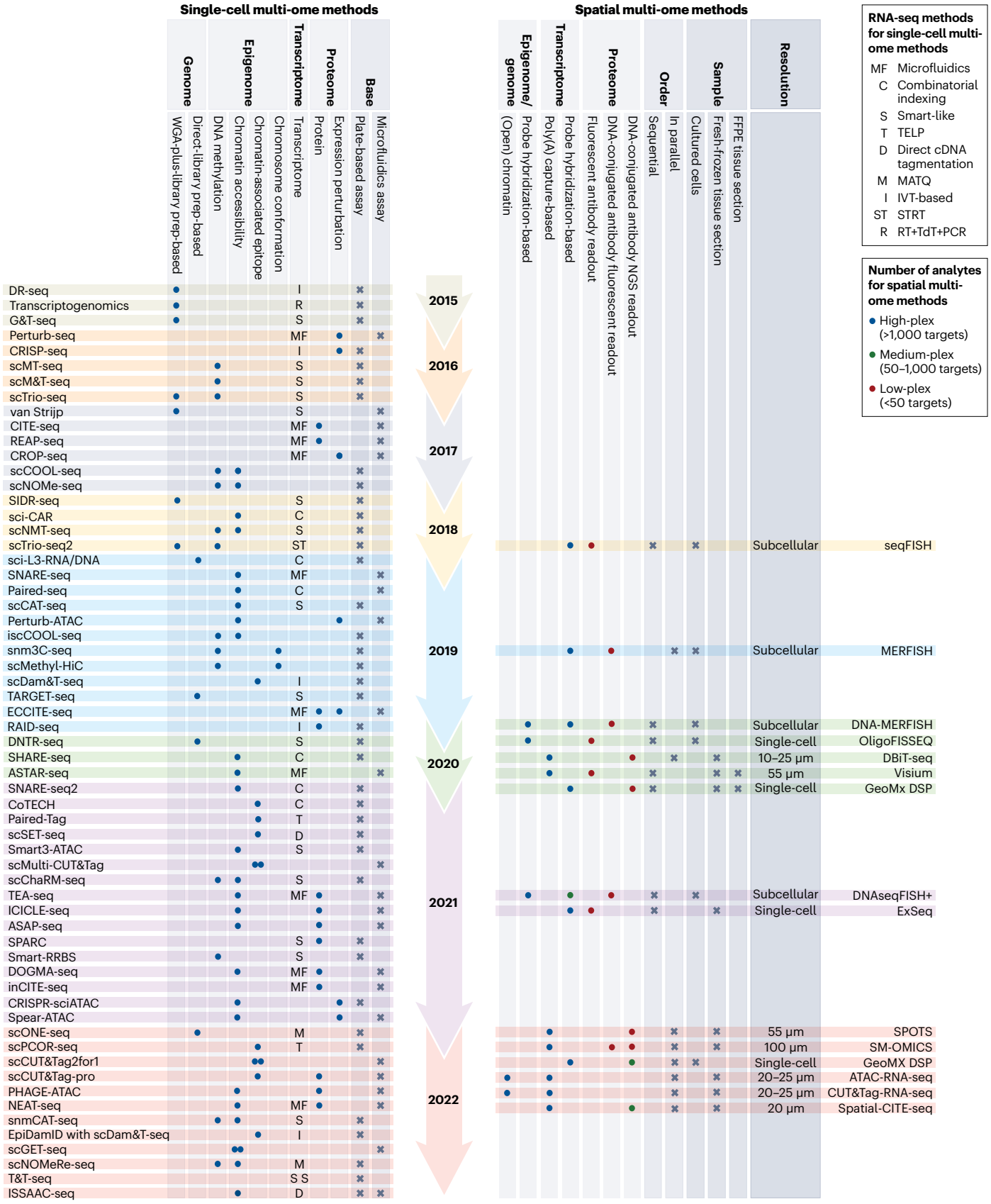

Single-cell multi-ome methods
Genome
Epigenome
Transcriptome
Proteome
Base
WGA-plus-library prep-based
Direct-library prep-based
DNA methylation
Chromatin accessibility
Chromatin-associated epitope
Chromosome conformation
Transcriptome
Protein
Expression perturbation
Plate-based assay
Microfluidics assay
DR-seq
Transcriptogenomics
G&T-seq
Perturb-seq
CRISP-seq
scMT-seq
scM&T-seq
scTrio-seq
van Strijp
CITE-seq
REAP-seq
CROP-seq
scCOOL-seq
scNOMe-seq
SIDR-seq
sci-CAR
scNMT-seq
scTrio-seq2
sci-L3-RNA/DNA
SNARE-seq
Paired-seq
scCAT-seq
Perturb-ATAC
iscCOOL-seq
snm3C-seq
scMethyl-HiC
scDam&T-seq
TARGET-seq
ECCITE-seq
RAID-seq
DNTR-seq
SHARE-seq
ASTAR-seq
SNARE-seq2
CoTECH
Paired-Tag
scSET-seq
Smart3-ATAC
scMulti-CUT&Tag
scChaRM-seq
TEA-seq
ICICLE-seq
ASAP-seq
SPARC
Smart-RRBS
DOGMA-seq
inCITE-seq
CRISPR-sciATAC
Spear-ATAC
scONE-seq
scPCOR-seq
scCUT&Tag2for1
scCUT&Tag-pro
PHAGE-ATAC
NEAT-seq
snmCAT-seq
EpiDamID with scDam&T-seq
scGET-seq
scNOMeRe-seq
T&T-seq
ISSAAC-seq
2015
2016
2017
2018
2019
2020
2021
2022
Spatial multi-ome methods
Epigenome/genome
Transcriptome
Proteome
Order
Sample
Resolution
(Open) chromatin
Probe hybridization-based
Poly(A) capture-based
Probe hybridization-based
Fluorescent antibody readout
DNA-conjugated antibody fluorescent readout
DNA-conjugated antibody NGS readout
Sequential
In parallel
Cultured cells
Fresh-frozen tissue section
FFPE tissue section
Subcellular
seqFISH
Subcellular
MERFISH
Subcellular
DNA-MERFISH
Single-cell
OligoFISSEQ
10–25 µm
DBiT-seq
55 µm
Visium
Single-cell
GeoMx DSP
Subcellular
DNAseqFISH+
Single-cell
ExSeq
55 µm
SPOTS
100 µm
SM-OMICS
Single-cell
GeoMX DSP
20–25 µm
ATAC-RNA-seq
20–25 µm
CUT&Tag-RNA-seq
20 µm
Spatial-CITE-seq
RNA-seq methods for single-cell multi-ome methods
MF Microfluidics
C Combinatorial indexing
S Smart-like
T TELP
D Direct cDNA tagmentation
M MATQ
I IVT-based
ST STRT
R RT+TdT+PCR
Number of analytes for spatial multi-ome methods
High-plex (>1,000 targets)
Medium-plex (50–1,000 targets)
Low-plex (<50 targets)

Fig. 1 | Timeline of single-cell and spatial multimodal methods. In addition to the year of publication, other key features of the methods are indicated. For single-cell multi-omics assays, this includes the nature of molecular analytes they analyse as well as the method used for cell barcoding. For spatial multi-omics assays, this includes the resolution, sample type, order and number of analytes that can be profiled simultaneously. FFPE, formalin-fixed paraffin-embedded; IVT, in vitro transcription; MATQ, multiple annealing and dC-tailing-based quantitative single-cell RNA-seq; STRT: single-cell tagged reverse transcription sequencing; RT, reverse transcription; TdT, terminal deoxynucleotidyl transferase; TELP, tailing extension ligation and PCR.

for the construction of a genetic lineage tree of the cells, which can be annotated with cell type and functional phenotypic states of cells using the RNA sequencing data of the same cells. As an alternative to analysing naturally occurring somatic mutations, high-throughput methods relying on CRISPR-scarring are available in model systems[32], which through recent improvements may enable accurate lineage recording as well as the capturing of ancestral transcriptional states[33,34]. In combination with scRNA-seq readouts, these promise to revolutionize our understanding of cellular differentiation trajectories.

Single-cell epigenomics-plus-transcriptomics

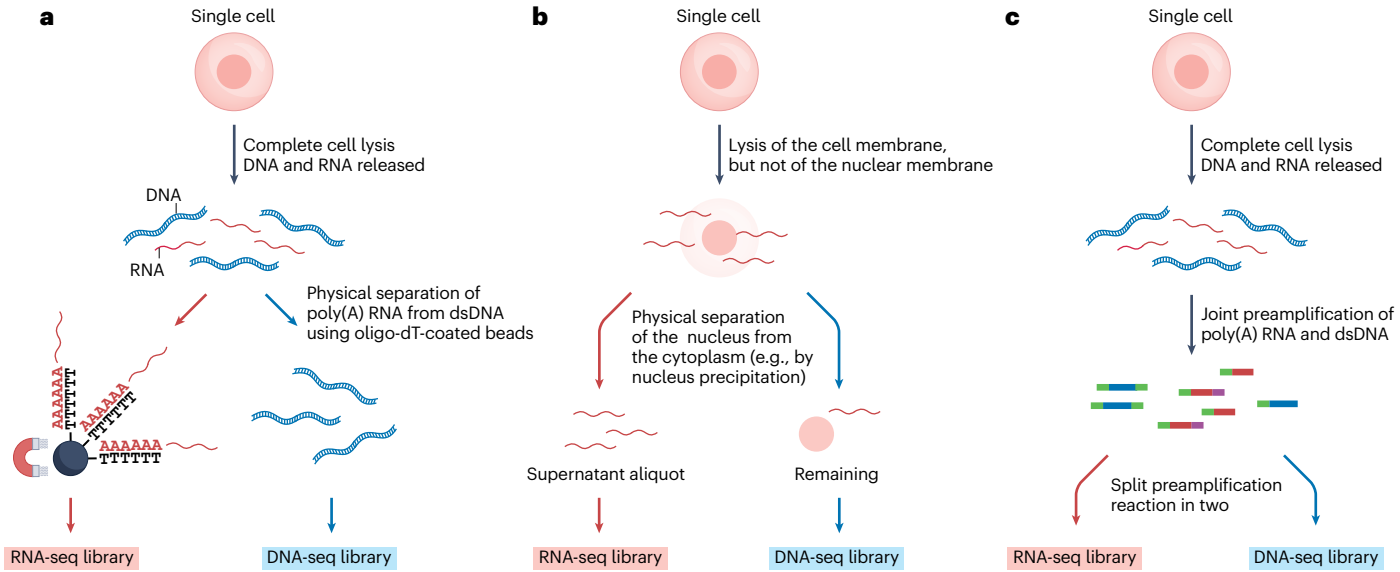
Advances in the past few years have also moved the mark for methods that can profile a cell's epigenome and transcriptome in parallel (Fig. 1). As with methods analysing the genome and transcriptome, their original designs relied on the physical separation of RNA from DNA or the nucleus, and the subsequent profiling of these separated fractions. More recently, methods for the differential marking of RNA and epigenetic information encoded in DNA have been described, relying on the separation of reads originating from both through restriction digestion, PCR or molecular barcoding. These later methods do not require upfront separation and are therefore more readily parallelized in higher throughput, via preamplification and split, seq-split or combinatorial indexing principles (Figs. 2 and 3).

Most known layers of epigenetic information, including DNA methylation, chromatin accessibility, histone modifications, and binding of transcription factors (TFs) and chromatin remodelling complexes, can be recovered from single cells in parallel to the transcriptome. Such methods abound but range in sensitivity, specificity and ease of use (Fig. 1). The presence of epigenetic modifications can be read either directly from the DNA sequence, as is the case for DNA methylation, or indirectly by first encoding them in the sequence through DNA methylation and/or tagmentation. We discuss the latter approach first, as this strategy is most widely used.

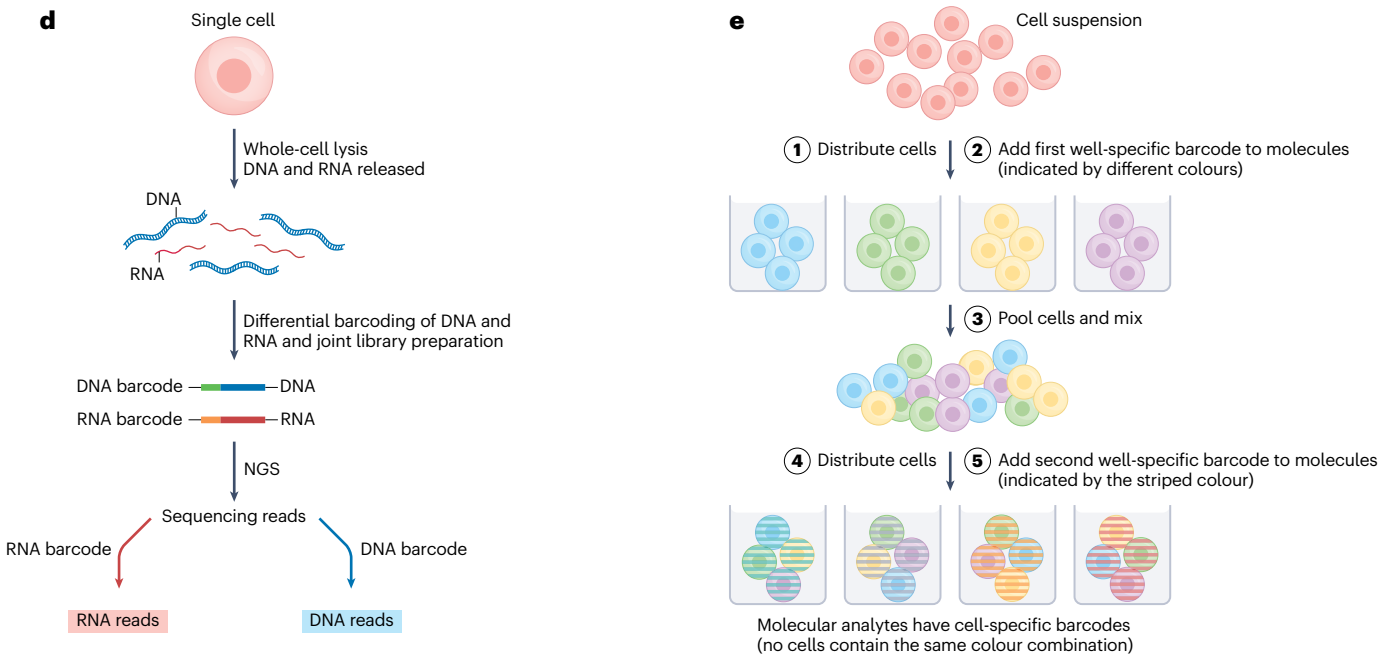
Tagmentation-based methods come in a wide range of types. Most common are methods that jointly profile the transcriptome and chromatin accessibility (Fig. 3). Here, accessible DNA is recovered as transposon-insertion-flanked regions using an assay for transposase-accessible chromatin (ATAC). Examples include scCAT-seq[35] and Smart3-ATAC[36], which are plate-based; sci-CAR-seq[37], SHARE-seq[38], SNARE-seq2[39] and Paired-seq[40] (Fig. 3a,b), which rely on combinatorial indexing; SNARE-seq[39,41], ASTAR-seq[42] and the commercially available 10x Genomics Multiome technology, which rely on microfluidics for cell barcoding (Fig. 3c); and ISSAAC-seq, which is amenable to both plate-based and microfluidics-based cell barcoding[43,44]. scGET-seq represents an unusual type of such methods, profiling both accessible and inaccessible chromatin but not the transcriptome[45]. Important features to be considered in tagmentation-based method selection include the ease of use of commercially available methods, the higher throughput of combinatorial indexing-based methods, the typical lower cost (but difficulty of establishment) of non-commercial, laboratory-developed methods, and the coverage, sensitivity and specificity obtained across cellular modalities. A systematic benchmark of these methods is unfortunately currently lacking. In most methods, transcriptome and accessible chromatin libraries are prepared in a common reaction and separated after indexing using magnetic beads, restriction enzymes or specific PCR primers (Fig. 3). A key advantage of the joint profiling of transcriptome and chromatin accessibility is that the link between gene expression and TF binding is more readily evaluable. In hair follicle cells, for example, SHARE-seq analyses of differentiation trajectories revealed TFs becoming expressed, with their activity being revealed in ATAC profiles later in pseudotime as binding sites became accessible, before expression of the associated target genes[38]. As such, these multi-omic analyses in dynamic systems enable TF activity to be readily assigned to target genes. The ease of use of some of these methods has spearheaded their application in biomedicine and other domains.

However, the aforementioned methods quantify open chromatin without addressing the causes of accessibility changes, such as shifts in histone post-translational modifications (PTMs) or TF binding. To tackle this limitation, other methods have been developed in which tagmentation is not randomly targeting accessible regions but directed towards specific histone PTMs or TFs. Conjugating the transposase to specific antibodies enables joint profiling of transcriptome and epitopes. Cell and modality barcoding can also occur through DNA–RNA separation, preamplification and split, seq-split or combinatorial indexing principles (Fig. 2), with transcriptome and DNA libraries being separated by restriction enzymes, PCR or beads. Examples of such methods include scPCOR-seq[46], coTECH[47], Paired-Tag (Fig. 3b)[48] and scSET-seq[49]. scCUT&Tag2for1[50] and scMulti-CUT&Tag[51] are distinct variations on these methods, involving two epitopes being targeted using different antibody-conjugated transposases. The distribution of both epitopes in a single cell can be learned from their different genomic distributions (for example, broad or narrow peaks) in scCUT&Tag2for1[50], or from epitope-specific barcoding enabled by transposases loaded with different oligonucleotides in scMulti-CUT&Tag[51]. A potent illustration of these methods was the profiling of mouse brain cells for both transcripts and different histone modifications using Paired-Tag, which identified distinct categories of genes each regulated by different epigenetic mechanisms[48].

Epigenetic information can also be read through DNA methylation profiling, relying either on 5-methylcytosine (5mC) or N^6-methyladenine (6mA) (Fig. 4). 5mC is a prevalent endogenous epigenetic modification of DNA, found almost exclusively in a CpG context, that represses ectopic and heterochronic gene transcription initiation. In the most basic approach, endogenous 5mC DNA methylation is quantified in single cells by bisulfite sequencing. Key issues precluding more widespread adoption of single-cell 5mC analyses are the costs associated with library preparation and whole-genome sequencing, as well as the technical complexity. Indeed, bisulfite treatment of DNA causes its denaturation and fragmentation, and comes with a need for purification, leading to DNA losses. Library preparation can occur either through random-priming-based methods, such as post-bisulfite adapter tagging (PBAT), which is costly but yields genome-wide profiles covering 5–50% of the genome, or through reduced-representation bisulfite sequencing (RRBS), which is more cost-effective but only



Principle 1: physical separation		Principle 2: preamplification and split
G&T-seq; transcriptogenomics; scChaRM-seq; scNMT-seq	SIDR-seq; DNTR-seq; scCAT-seq	DR-seq; SNARE-seq; DOGMA-seq



Principle 3: seq-split	Principle 4: combinatorial indexing
scONE-seq; scDam&T	sci-L3 DNA RNA co-assay; sci-CAR-seq; SNARE-seq2

covers 1–3% of the genome. Recent evolutions in less destructive DNA methylome analytics and target capture may serve to develop more attractive workarounds[52,53]. DNA methylation profiled on its own through single-cell DNA sequencing enables concomitant chromosomal copy number profiling. More often, DNA methylomes are profiled together with a cell's transcriptome, chromatin structure and/or chromatin accessibility.

Methods that produce transcriptome profiles alongside targeted DNA methylome profiles – as in scMT-seq, scTrio-seq and Smart-RRBS – or genome-wide DNA methylome profiles – as in scM&T-seq and scTrio-seq2 – have been established[54–58]. These typically involve physical separation of DNA and RNA (Fig. 2a,b).

Endogenous CpG DNA methylation is often assessed in conjunction with chromatin accessibility. For this, a GpC methyltransferase is added

Fig. 2 | The four general principles for multi-ome measurements from single cells. All principles are visualized with RNA and DNA as example analytes. Principle 1 is based on physical separation of the distinct molecular analytes (parts **a**,**b**). **a**, Following complete lysis of the isolated cell or nucleus, poly(A) RNA hybridizes to oligo-dT-coated paramagnetic beads, and following magnetic pulldown, the supernatant that contains the genomic DNA is transferred to a new reaction vessel[7,9]. Alternatively, biotinylated nucleotides are incorporated into RNA-derived cDNA, allowing their capture with streptavidin-coated paramagnetic beads (not shown)[42]. Advantages include flexibility in downstream processing of DNA and/or RNA and compatibility with intact cells and nuclei from fresh and frozen tissue. Disadvantages include potential loss of RNA and/or DNA molecules during physical separation. **b**, In an alternative approach, lysis conditions that rupture the cell but not the nuclear envelope allow separation of nuclear from cytoplasmic molecular analytes, either by precipitating the nucleus with centrifugation or magnetic pulldown followed by aspiration of the cytosolic supernatant[13,14] or by microfluidic-controlled nucleus-from-cytoplasm separation[16,17]. Advantages include flexibility in downstream processing of DNA and RNA and availability of non-poly(A) RNA. Disadvantages include loss of nuclear RNA plus some cytoplasmic RNA during nuclear–cytoplasmic separation, loss of mitochondrial analytes, need for intact single cells, incompatibility with frozen tissue and likely incompatibility with mitotic cells (in which the nuclear envelope disaggregates). **c**, In principle 2, termed preamplification and split, distinct analytes are differentially tagged and jointly preamplified, followed by splitting the preamplification reaction for analyte-specific sequencing library preparations[18]. Advantages include minimal risk of analyte loss and compatibility with intact cells and nuclei from fresh and frozen tissue. Disadvantages include limited flexibility, as the preamplification protocol needs to be suitable for all analytes of interest, and risk of cross-contaminating molecular analytes. **d**, Principle 3, termed seq-split, involves analyte-specific barcoding and sequencing library preparation in a single-pot reaction[20]. Multi-omic information is uncoupled computationally following sequencing. Advantages include minimal risk of analyte loss and compatibility with intact cells and nuclei from fresh and frozen tissue. Disadvantages include that libraries cannot be sequenced separately to optimal depth for each modality, and potential risk of cross-contaminating molecular analytes. **e**, In principle 4, termed combinatorial indexing, molecular analytes of single cells are tagged without isolating single cells[21]. Multiple cells or nuclei are deposited per well of a multi-well plate, whereby each cell or nucleus serves as a reaction container. Each receives an analyte-specific tag and a well-specific barcode. By pooling, mixing and randomly re-distributing the cells or nuclei in subsequent rounds of well-specific barcoding, molecular analytes uniquely barcoded per cell are obtained. Combination with concepts of principle 2 and/or 3 achieves single-cell or single-nucleus multi-omics. Advantages include that there is no need for isolating single cells, ability to achieve extremely high throughput, and compatibility with intact cells and nuclei from fresh and frozen tissue. Disadvantages include typically lower sensitivity, risk of analyte loss and limited flexibility in whole-genome and whole-transcriptome amplification protocols. dsDNA, double-stranded DNA; NGS, next generation sequencing; poly(A), polyadenylated.

to isolated nuclei where it can methylate open-chromatin-associated accessible DNA (Fig. 4a,b). Subsequent bisulfite sequencing can disclose the methylated GpC dinucleotides that mark these accessible regions, in addition to the endogenous methylation that in most cell types is nearly exclusive to CpG dinucleotides. GpC methylase-based methods show higher coverage per promoter than ATAC-based methods[59], albeit at substantially higher sequencing cost per cell, and make it easier to distinguish open from truly closed regions, which are not directly detected in the sparse scATAC-seq signals. A limitation relative to regular DNA methylome analyses is that at cytosines flanked on both sides by a guanine, methylation can be attributed to endogenous as well as exogenous processes. These are discarded from analyses. Relevant methods include scCOOL-seq[60], iscCOOL-seq[61] and scNOME-seq[62], which jointly profile accessibility and DNA methylation, and trimodal methods such as scNMT-seq[59], scNOMeRe-seq[63], scChaRM-seq[64] and snmCAT-seq[65], which in addition uncover gene expression profiles from the same cell (Fig. 4a,b). All are plate-based assays and based on DNA–RNA separation, apart from snmCAT-seq, which is seq-split based (Fig. 2d; Fig. 4b).

Noteworthy alternatives that do not profile transcriptomes are snm3C-seq[66] and scMethyl-HiC[67], as they profile the DNA methylome in parallel to higher-order chromatin structure (Fig. 4c). Here, a 3C or HiC-like single-cell method to capture nuclear organization is combined with bisulfite conversion, to enable joint profiling of DNA methylation and chromatin structure[66,67]. Methods that jointly also capture the transcriptome have yet to be described.

Despite being challenging techniques, the multi-omics nature of the resultant data enables a very in-depth analysis of cells, revealing hierarchies of changes to the epigenome and transcriptome during development, differentiation or pathogenesis. For example, stages of mouse gastrulation were profiled using scNMT-seq, which demonstrated that DNA methylation and chromatin accessibility patterns at ectodermal enhancers are pre-established in epiblast cells and stable during ectoderm differentiation, whereas mesodermal and endodermal enhancers are inactive in epiblast cells but actively remodelled following differentiation to mesoderm or endoderm[68]. Accessibility and DNA methylation changes seemed to be tightly coordinated. These studies illustrate that multi-omics profiling of single cells for transcriptomic and epigenetic layers is feasible and provides fundamental insights unattainable using mono-omics methods.

Apart from methylated cytosines, methylation of adenines to 6mA can also be profiled in DNA (Fig. 4d). In contrast to 5mC, 6mA is an ultra-rare base in mammalian DNA[69] and can thus provide nearly unambiguous DNA marking when artificially introduced. Typically, a prokaryotic DNA adenine methyltransferase (Dam) is used to methylate adenines in GATC context for DNA-sequence encoding of epigenetic information. Dam can either be expressed to mark accessible chromatin[70–72] or be tethered to endogenously expressed proteins using insertion mutagenesis to mark their binding sites in live cells. Two DNA–RNA separation-based methods applying adenine methylation to 6mA have been described: scDam&T-seq, with Dam tethered to chromatin-associated proteins to mark their nuclear location (DamID)[70], and scDam&T-seq with EpiDamID, in which Dam is tethered to protein domains or nanobodies that recognize PTMs[73] (Fig. 4d). As 6mA cannot be detected directly through short-read sequencing, quantification is relative, relying on Dam activity and a 6mA-specific restriction digest. In contrast to the other chromatin profiling methods described above, the resulting profiles typically reflect the aggregate of residence times of the proteins marked, leading to better signal-to-noise ratios but poorer temporal resolution. The reliance on endogenous tethering limits the scope of these methods to systems amenable to transgenesis. Interestingly, it extends their scope beyond only those proteins for which a specific antibody is available, and at least scDam&T-seq enables a method of analysing chromatin occupancy that is orthogonal to antibody-based methods. Additionally, these methods may suffer less from the bias towards accessible chromatin that characterizes transposase-based tagging approaches. Interestingly, these methods share several features with Fibre-seq-based approaches[74], in which chromatin patterns are read through 6mA

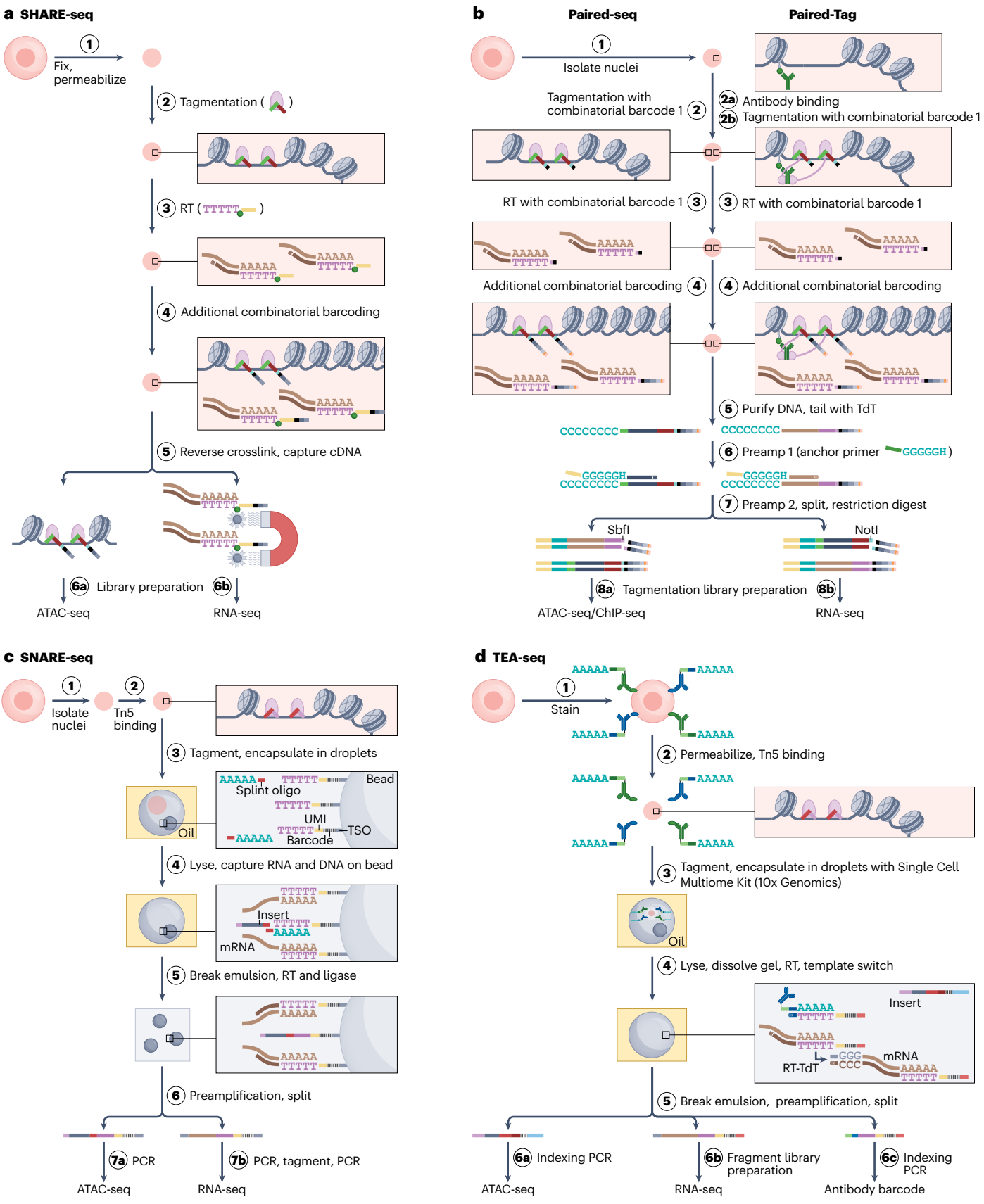

a SHARE-seq
1 Fix, permeabilize
2 Tagmentation ( )
3 RT (TTTTT)
4 Additional combinatorial barcoding
5 Reverse crosslink, capture cDNA
6a Library preparation
6b
ATAC-seq
RNA-seq
b
Paired-seq
Paired-Tag
1 Isolate nuclei
Tagmentation with combinatorial barcode 1
2
2a Antibody binding
2b Tagmentation with combinatorial barcode 1
RT with combinatorial barcode 1
3
3 RT with combinatorial barcode 1
Additional combinatorial barcoding
4
4 Additional combinatorial barcoding
5 Purify DNA, tail with TdT
6 Preamp 1 (anchor primer GGGGGH)
7 Preamp 2, split, restriction digest
SbfI
NotI
8a Tagmentation library preparation
8b
ATAC-seq/ChIP-seq
RNA-seq
c SNARE-seq
1 Isolate nuclei
2 Tn5 binding
3 Tagment, encapsulate in droplets
Splint oligo
Bead
UMI
TSO
Barcode
Oil
4 Lyse, capture RNA and DNA on bead
Insert
mRNA
5 Break emulsion, RT and ligase
6 Preamplification, split
7a PCR
7b PCR, tagment, PCR
ATAC-seq
RNA-seq
d TEA-seq
1 Stain
2 Permeabilize, Tn5 binding
3 Tagment, encapsulate in droplets with Single Cell Multiome Kit (10x Genomics)
Oil
4 Lyse, dissolve gel, RT, template switch
Insert
mRNA
RT-TdT
5 Break emulsion, preamplification, split
6a Indexing PCR
6b Fragment library preparation
6c Indexing PCR
ATAC-seq
RNA-seq
Antibody barcode

Fig. 3 | Selected tagmentation-based methods for single-cell multi-omic analyses. Summaries of experimental workflows highlighting how and in what order different modalities are probed and separated for analysis, while retaining single-cell information. Shown are SHARE-seq[38] (part **a**), Paired-seq2[40] (part **b**, left) and Paired-Tag[48] (part **b**, right), SNARE-seq[39] (part **c**) and TEA-seq[82] (part **d**). In all methods shown, an assay for transposase-accessible chromatin (ATAC) precedes reverse transcription (RT) of polyadenylated RNA. Cell barcoding can occur through successive rounds of combinatorial indexing (parts **a**,**b**) or by compartmentalizing cells and barcoded oligonucleotides in microdroplets (parts **c**,**d**). DNA fragments originating from mRNA and DNA can be separated by binding to paramagnetic beads (part **a**), differential restriction digestion (part **b**) or using specific PCR primers (parts **c**,**d**). Part **b** illustrates how similar workflows can either map accessible chromatin (left) and chromatin-associated proteins or their post-translational modifications (right). TEA-seq illustrates that barcoded oligonucleotides conjugated to antibodies can be detected using approaches similar to those developed for measuring gene expression. cDNA, complementary DNA; ChIP–seq, chromatin immunoprecipitation followed by sequencing; RNA-seq, RNA sequencing; TdT, terminal deoxynucleotidyl transferase; TSO, template-switching oligonucleotide; UMI, unique molecular identifier.

marking using single-molecule long-read sequencing. Combining both methods may thus facilitate high-throughput, haplotype-resolved and cell-type-specific characterization of chromatin structures in bulk.

Single-cell omics plus low-plex profiling of another analyte

Aside from profiling multiple comprehensive omics layers from the same cell, substantial progress was recently made in profiling single cells comprehensively for a single analyte and in a less comprehensive (low-plex) manner for another analyte. Most common among these methods are those quantifying the cell's transcriptome and/or accessible genome, alongside a limited set of cell surface or intranuclear proteins. These methods typically rely on antibodies tagged with a specific barcoded oligonucleotide, which can be captured and amplified alongside the transcriptome. Using mixes of dozens to hundreds of such barcoded antibodies that each recognize specific epitopes, subsequent barcode counting thus enables quantification of multiple proteins of interest in single cells alongside true omics analysis. Such methods enable the profiling of protein abundances in addition to gene expression or other modalities, but can also capture other information about proteins such as protein stability, PTMs and protein isoform expression. Omics layers shown to be amenable to this include the transcriptome (REAP-seq[75], CITE-seq[76], inCITE-seq[77], SPARC[78], ECCITE-seq[79] and RAID-seq[80]), open chromatin (ASAP-seq[81] and ICICLE-seq[82]) or both (DOGMA-seq[81] and TEA-seq[82]), as well as chromatin modifications (scCUT&Tag-pro[83]) (Fig. 3d). Epitopes available for profiling are mostly limited to the cell surface, although methods for intracellular (SPARC[78] and RAID-seq[80]) and intranuclear (inCITE-seq[77] and NEAT-seq[84]) epitopes have also been developed. Note that any plate-based method can theoretically also leverage antibody marking and cytometry to quantify a limited set of proteins per cell. A notable alternative approach (PHAGE-ATAC) was described recently in which barcoded phages that display a nanobody serve to bind cell surface epitopes[85]. The barcoded phage genomes can be quantified in conjunction with genome-wide chromatin accessibility profiles. Each of these methods is currently limited by the availability of specific antibodies or nanobodies, and although mass spectrometry-based methods for proteome-wide analysis of single cells have been developed[86], such analyses alongside other omics layers are currently lacking. A related method recently described is single-cell transcriptome and translatome sequencing (T&T-seq), in which cells are sorted into plates and lysates are distributed for total RNA-seq and for affinity purification of actively translating ribosomes. T&T-seq thus enables joint profiling of all transcripts and those transcripts that are being translated into proteins[87].

Finally, other molecular features have also been shown to accommodate barcoding, with guide RNAs as a key example. In cells expressing CRISPR-based gene editing, activation or inactivation systems, feature barcoding methods enable high-throughput profiling of the transcriptome or accessible chromatin in pools of single cells subjected to high-throughput genetic perturbation screens. Examples include CROP-seq, in which the guide RNA sequence is directly determined alongside the cell's transcriptome[88], and CRISP-seq[89] and Perturb-seq[90,91], in which each guide RNA has a unique barcode that is sequenced alongside the transcriptome. The latter two methods have lower specificity, as recombination can blur the link between barcode and guide RNA, and more recent implementations of Perturb-seq therefore apply direct guide-RNA sequencing. Methods to profile the impact of such perturbations on chromatin accessibility include CRISPR-sciATAC[92], Perturb-ATAC[91] and Spear-ATAC[93]. Recently, this approach was applied at scale when millions of single-cell transcriptomes were analysed, with each expressed gene inactivated in a subset of these cells. This strategy revealed the effect of inactivation of each expressed gene on the expression of all other genes, and thus represents a valuable resource for the in silico modelling of genetic perturbations[94]. Notably, a similar approach can be deployed to characterize gene interactions[95]. Here, cells are transfected with, on average, two different guide RNAs from a pool, generating pools of double-knockout cells. The combined impact of both perturbations can then be compared with the impact of single perturbations, providing direct quantification of a matrix of genetic interactions.

Spatial multi-omics methods

Methodologies for spatial multi-omics are developing rapidly to allow the study of different molecular analytes at up to subcellular resolution within their native tissue context (Fig. 1). Spatial multi-omics technologies were listed by *Nature* as one of the seven technologies to watch in 2022[96], with the basis for their development and ongoing innovations being a range of established spatial mono-omics methods (Box 1). Spatial multi-ome characterization of a sample, usually a fixed fresh-frozen or formalin-fixed paraffin-embedded (FFPE) tissue section, is often achieved by combining these spatial mono-omics methods. They can be applied separately on adjacent tissue sections, serially on the same tissue section if the quality of the different analytes can be maintained, or in parallel on the same tissue section if joint targeting and reading out of the different analytes is possible. The number of target analytes that can be analysed simultaneously at the different molecular levels varies between methodologies (Fig. 1). Often spatial omics measurements are also supplemented with histological stains, such as H&E (haematoxylin and eosin) staining, of the same or adjacent tissue sections, allowing integration with additional morphological annotations.

Spatial multi-omics via adjacent-section strategies

The application of spatial assays for different mono-omics layers of interest (Box 1) on adjacent or serial sections from the same tissue sample enables these techniques to be assayed in their most optimal

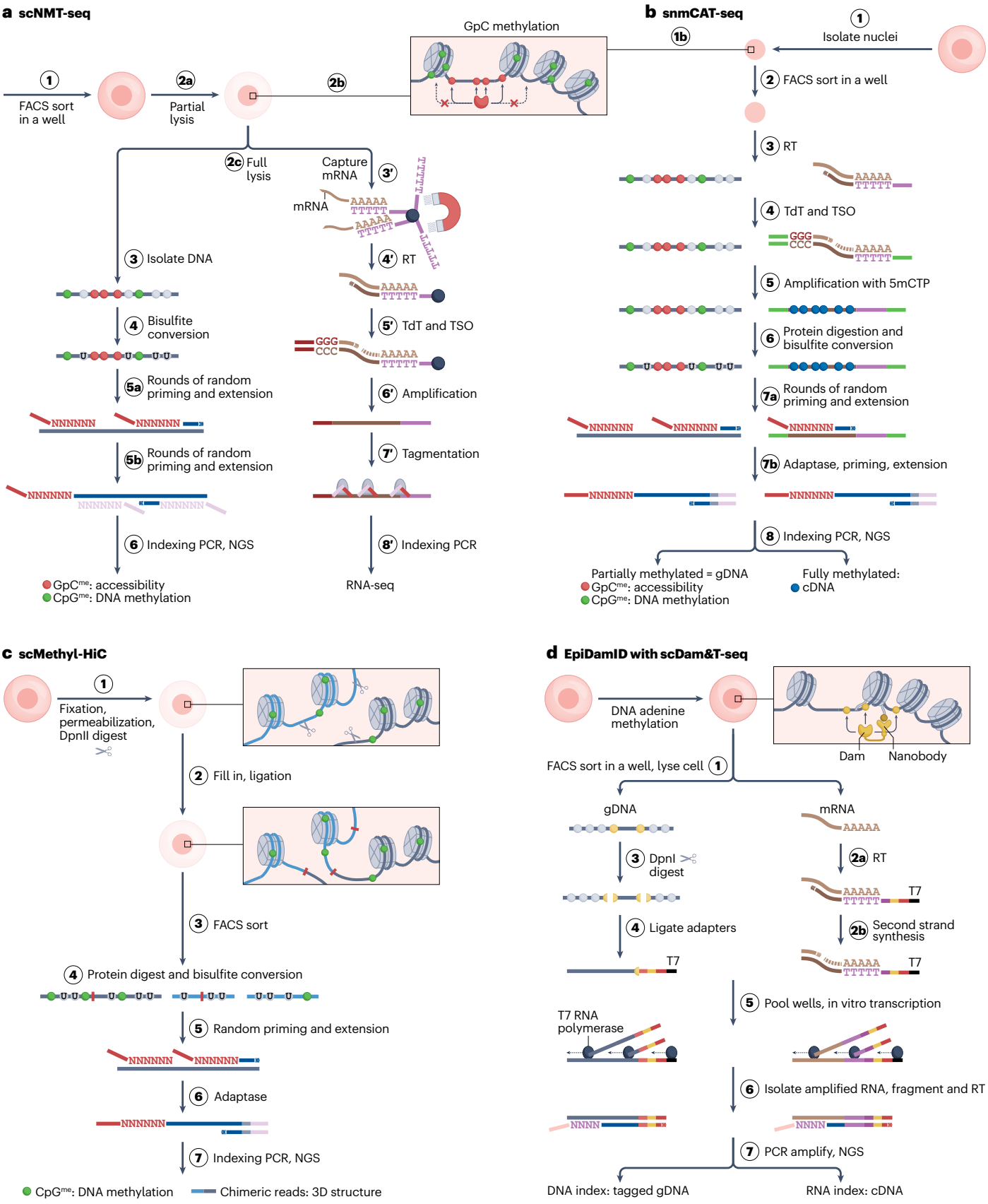

a scNMT-seq
1
FACS sort in a well
2a
Partial lysis
2b
GpC methylation
2c Full lysis
Capture mRNA
3'
mRNA
AAAAA
TTTTT
4' RT
5' TdT and TSO
GGG
CCC
6' Amplification
7' Tagmentation
8' Indexing PCR
RNA-seq
3 Isolate DNA
4 Bisulfite conversion
5a Rounds of random priming and extension
NNNNNN
5b Rounds of random priming and extension
6 Indexing PCR, NGS
GpC^me: accessibility
CpG^me: DNA methylation
b snmCAT-seq
1b
1 Isolate nuclei
2 FACS sort in a well
3 RT
4 TdT and TSO
5 Amplification with 5mCTP
6 Protein digestion and bisulfite conversion
7a Rounds of random priming and extension
7b Adaptase, priming, extension
8 Indexing PCR, NGS
Partially methylated = gDNA
GpC^me: accessibility
CpG^me: DNA methylation
Fully methylated:
cDNA
c scMethyl-HiC
1
Fixation, permeabilization, DpnII digest
2 Fill in, ligation
3 FACS sort
4 Protein digest and bisulfite conversion
5 Random priming and extension
6 Adaptase
7 Indexing PCR, NGS
CpG^me: DNA methylation
Chimeric reads: 3D structure
d EpiDamID with scDam&T-seq
DNA adenine methylation
Dam
Nanobody
FACS sort in a well, lyse cell 1
gDNA
mRNA
3 DpnI digest
4 Ligate adapters
T7
2a RT
2b Second strand synthesis
5 Pool wells, in vitro transcription
T7 RNA polymerase
6 Isolate amplified RNA, fragment and RT
NNNN
7 PCR amplify, NGS
DNA index: tagged gDNA
RNA index: cDNA

Fig. 4 | Selected DNA methylation-based methods for single-cell multi-omic analyses. Summaries of experimental workflows highlighting how and in what order different modalities are probed and separated for analysis, while retaining single-cell information. Shown are scNMT-seq[59] (part **a**), snmCAT-seq[65] (part **b**), scMethyl-HiC[67] (part **c**) and EpiDamID with scDam&T-seq[73] (part **d**). In all methods shown, several types of epigenetic information can be discerned from a single sequencing library: DNA methylation and accessibility (part **a**), DNA methylation, accessibility and gene expression (part **b**), DNA methylation and chromatin conformation (part **c**), and histone modifications and transcription (part **d**). These layers of information can be discriminated by analysing DNA methylation patterns (parts **a**,**b**), DNA methylation and read-pair mapping (part **c**) or read-associated barcode tags (part **d**). Only scNMT-seq (part **a**) involves physical separation of modalities to be probed (mRNA from intact cells). Each of these methods is plate-based and therefore restricted in its throughput. DNA methylation patterns can be read using bisulfite conversion of unmethylated cytosines to uracil (U) (parts **a**–**c**) or methylation-sensitive restriction digestion (part **d**), and reflect endogenous methylation alone (part **c**), endogenous methylation as well as exogenously added methylation reflecting chromatin accessibility (parts **a**,**b**), or endogenous adenine methylation added after genetic transformation (part **d**). cDNA, complementary DNA; Dam, DNA adenine methyltransferase; FACS, fluorescence-activated cell sorting; gDNA, genomic DNA; 5mCTP: 5-methyl-deoxycytidine triphosphate; NGS, next-generation sequencing; RT, reverse transcription; T7, T7 promoter; TdT, terminal deoxynucleotidyl transferase; TSO, template-switching oligonucleotide.

setting and their data to be integrated computationally (Fig. 5a). This requires sample collection and preparation procedures compatible with all assays of interest, well-thought-out experimental design and good assay coordination. This strategy is mostly required when multi-omics readouts of interest are not possible yet or suboptimal on the same section. Mass spectrometry imaging (MSI) methods[97], for example, can be used for spatial metabolome characterization of small biomolecules such as lipids (Box 1), but cannot easily be combined with other spatial genome, transcriptome or proteome readouts on the same section owing to specific sample preparations and limitations (for example, the need for matrix deposition in MSI[97], OCT compound embedding complicating mass spectrometry[98] and inherent breakdown of lipids during potential upfront assays). Conversely, MSI is compatible with H&E staining on the same section, which can also be leveraged for computational integration with other spatial assays. Additional serial sections can also be dissociated into single cells or nuclei for generating matched single-cell sequencing data that can be used for optimal deconvolution of the spatial data and additional data integration (Fig. 5a). As such, a combination of different single-cell sequencing and spatial transcriptomics and proteomics approaches was applied and integrated to assemble an atlas of healthy and obese murine and human livers and to map hepatic macrophage niches[99]. Disadvantages of this approach are that not all assays are compatible with all sample types (for example, fresh-frozen versus FFPE), that it suffers from sample heterogeneity, as even adjacent sections will differ slightly as to structure and cellular composition, and that different spatial assays may have different resolutions. Although computational tools exist that can compensate for these challenges (see below), the concordance between sections is never unambiguous. Hence, innovative methods enabling spatial multi-omics measurements on the same tissue sections have started to emerge.

Spatial (epi)genomics-plus-transcriptomics

The possibility for simultaneous unbiased profiling of chromatin accessibility or specific histone modifications and gene expression on the same tissue cryosections has been described for spatial ATAC&RNA-seq and spatial CUT&Tag-RNA-seq, respectively[100]. These methods are based on combining microfluidic deterministic barcoding in tissue (DBiT) strategies (Box 1) for spatial-ATAC-seq[101] or spatial-CUT&Tag[102] with DBiT-seq poly(A) transcript profiling[103] (Fig. 5b). For capturing chromatin accessibility, accessible gDNA is first tagmented in situ with a universal ligation linker. For capturing specific histone modifications, the tissue is first incubated with primary antibodies against the epigenetic marks, followed by a secondary antibody that allows protein A-transposome tethering for tagmentation of the DNA at these specific locations. Combined mapping of the transcriptome is achieved by hybridizing a biotinylated oligo-dT that also contains a universal ligation linker and primes RT in situ[100] (Fig. 5b). DBiT-seq barcoding with serial attachment to the tissue of two microfluidics chips with equidistant channels perpendicular to each other is then used to administer two sets of channel-specific barcodes (Box 1), with the first set 'A' ligating with the universal ligation linkers present on the tagmented DNA and poly(A) mRNA-derived cDNA, and the second set 'B' ligating to the set 'A' barcodes, resulting in an in situ 2D grid with uniquely barcoded AB crossroads (20–25-µm pixels)[100]. The spatially barcoded gDNA and cDNA fragments are collected by reverse crosslinking, cDNA is enriched with streptavidin-coated magnetic beads, gDNA is retained in the supernatant, and NGS libraries are constructed separately for sequencing (Fig. 5b). Based on the spatial barcodes, sequencing reads are combined with microscopy images of the tissue section, allowing the multi-omics sequence information to be mapped spatially[100]. Limitations of these assays are the near-single-cell resolution (20–25-µm pixel size), the small analysable area (2,500–10,000 pixels), the uncharacterized spaces in between adjacent pixels (depending on channel distances) and the expertise that is required in fabricating and handling microfluidics chips for implementation. Nevertheless, these methods were successfully applied to developing mouse and adult human brains, revealing how epigenetic states or modifications regulate cell type, states and dynamics[100].

Alternatively, microscopy-based methods can enable spatial profiling of genome or epigenome information together with gene expression by directly imaging DNA loci, chromosomal and nuclear structures, and transcripts within single cells at up to subcellular resolution. Several methods showing imaging-based multi-omic measurements for a limited number of combined analytes exist[104–106], but we focus on those that can characterize at least one analyte layer more comprehensively. Multiplex single-molecule fluorescent in situ hybridization (smFISH) methods, such as MERFISH[107,108] and seqFISH[109,110] approaches (Box 1), originally designed for targeted high-sensitivity spatial profiling of thousands of transcripts at subcellular resolution, were adapted to allow genome-scale chromatin tracing as in DNA-MERFISH[111] and DNA-seqFISH+[112]. Combined RNA, chromatin and nuclear body imaging in the same fibroblast cells was enabled by sequentially staining for >1,100 nascent transcripts using RNA-MERFISH, followed by DNA-MERFISH for >1,000 genomic loci, and finally fluorescent antibody readouts for cell-cycle-state determination and landmark nuclear structures, including nuclear speckles and nucleoli[111] (Fig. 5c). This approach allows for the characterization of chromatin domains, compartments and trans-chromosomal interactions and their relationship to transcription in single cells[111]. Similarly, DNA-seqFISH+ imaging of up to 3,660 chromosomal loci in mouse embryonic stem cells was shown to be compatible with RNA-seqFISH of 70 mRNAs and intron-seqFISH of 1,000 genes at their

Box 1

Principles of spatial mono-omics methods for spatial multi-omics developments

Methods for spatial mono-omics have progressed tremendously in the last decade with excellent reviews existing for spatial transcriptomics[165,171–173], (epi)genomics[166,174], proteomics[86,167] and metabolomics[168,175]. Here, we briefly highlight the basic principles of the approaches that currently form the basis for most ongoing spatial multi-omics developments.

Array-based spatial transcriptomics

These approaches allow for transcriptome-wide profiling through capturing of polyadenylated (poly(A)) RNA transcripts released from fixed and permeabilized tissue sections, making use of thousands of arrayed and spatially barcoded oligo-dT spots (for example, Spatial Transcriptomics[170,176] and 10x Genomics Visium[115]), uniquely DNA-barcoded beads (for example, Slide-seq[177] and HDST[178]) or even barcoded DNA nanoballs (for example, Stereo-seq[179]) on a slide surface. Following reverse transcription into barcoded cDNA and library preparation, gene expression is spatially profiled by addressing the barcoded next-generation sequencing (NGS) reads to specific locations within the imaged tissue section.

Microfluidic deterministic barcoding strategies

These methods (including DBiT-seq[103], spatial-ATAC-seq[101] and spatial-Cut&Tag[102]) allow high-resolution NGS-based spatial (multi-)omics profiling. By serially attaching two polydimethylsiloxane (PDMS) microfluidics chips with equidistant microchannels perpendicular to each other on a pre-fixed tissue section, two sets of channel-specific DNA barcodes (A1-N and B1-N) can be delivered to the tissue surface for barcoding of captured poly(A) transcripts, proteins or epigenomic information. This yields a 2D map of pixels on the tissue with unique barcode combinations (AB) at the channel crossflows to spatially allocate barcoded NGS reads for the analytes of interest.

DNA antibody tags

This approach allows protein targeting and spatial mapping via NGS by using polyadenylated antibody-derived tag-conjugated antibodies, derived from single-cell CITE-seq[76]. These antibodies linked to DNA sequences with an antibody-specific barcode and a poly(A) tail are compatible with the capture-based array or microfluidic deterministic barcoding approaches above, for spatial (co-)profiling of proteins by NGS. Alternatively, next to the imaging-based methods that rely on cyclic fluorescent antibody stains for the identification of up ~60 protein targets in tissue sections (such as seqIF[180], CycIF[181] and MILAN[182]), antibodies labelled with DNA barcodes can be identified through cyclic hybridization and imaging of fluorescent readout sequences or with fluorescent detection during oligosequence amplification (for example, CODEX[183] and immuno-SABER[184]).

Multiplex smFISH

Multiplex single-molecule fluorescent in situ hybridization (smFISH) methods allow imaging of thousands of gene transcripts and genomic loci in single cells with high accuracy and subcellular resolution, to enable single-molecule detection of RNAs in their native tissue context and characterize chromosomal structure and organization. With the most advanced methods (such as MERFISH[107,111] and seqFISH+[112,185]), predefined optical barcoding schemes are assigned and imprinted onto oligonucleotide targets using a library of encoding probes. Each encoding probe contains a region targeting the sequence of interest and variable distinct readout sequences in multiple copies. After encoding probe binding, detection of the imprinted barcodes is achieved through multiple rounds of fluorescent readout probe hybridization, high-resolution imaging and signal quenching.

In situ sequencing

In situ sequencing (ISS) methods allow high-throughput spatial mapping of transcripts and genomic loci in single cells. Both targeted and untargeted approaches (for example, ISS[186], ExSeq[125] and FISSEQ[187,188]) exist for transcript identification, in which highly specific padlock probes are hybridized to in situ-synthesized cDNA sequences, followed by probe ligation and rolling circle amplification (RCA) of either the barcoded probe sequences or short sequences of the cDNAs, to generate micrometre-sized RCA products within cells that are decoded using in situ imaging-based sequencing-by-ligation. OligoFISSEQ[113] methods also allow genome-wide targeting of genomic loci.

Mass spectrometry imaging

Mass spectrometry imaging (MSI) is an alternative to the antibody-based approaches for spatially characterizing not only proteins or peptides but also other small biomolecules, such as lipids, metabolites or sugars, in tissue sections. Samples are systematically scanned by light or particle beams ionizing the surface biomolecules and making them accessible to time-of-flight (TOF) mass spectrometry for identification. Matrix-assisted laser desorption/ionization (MALDI) MSI methods enable spatial metabolome characterization in tissue sections at (near) single-cell resolution[189].

nascent transcription active sites, as well as sequential immunofluorescence targeting of 17 nuclear structures, including the nuclear lamina, nucleolus and histone modification marks, using primary antibodies conjugated with DNA-oligonucleotides detectable by fluorescently labelled readout probes[112] (Fig. 5c). In addition to identifying heterogeneity in chromosome structure, this study found that many active gene loci reside at the surface of nuclear bodies and the presence of persistent global chromatin states[112]. Also, OligoFISSEQ methods allow rapid in situ sequencing-based visualization (Box 1) of multiple genomic loci in single cells, with the potential for genome-wide application and being compatible with immunofluorescence and other FISSEQ-based methods for protein and RNA characterization, respectively[113].

Although these microscopy-based methods are important for studying chromosomal and nuclear structures and organization linked to gene expression regulation at a single-cell level, they require expertise in complex optical barcoding schemes and high-resolution imaging modalities, are challenging to apply in complex tissue samples, are costly and usually can only characterize a limited area.

Alternatively, by using laser capture microdissection (LCM)-based isolation of specific (single) cells from tissue sections, single-cell sequencing-based genome-plus-transcriptome or epigenome-plus-transcriptome profiling methods as described above can be applied at spatial resolution, as was done for analysing tumour development, metastasis and prognosis in patients with triple-negative breast cancer[114].

Spatial transcriptomics-plus-proteomics

Methodologies allowing parallel spatial interrogation of both the transcriptome and proteome are currently still limited, are often based on serial characterization of both modalities, mostly allow co-characterization of only a limited number of proteins and often lack single-cell resolution. For example, the commercial array-based 10x Genomics Visium technology for poly(A) RNA capture and spatial barcoding at 55-µm resolution followed by NGS identification (Box 1) currently supports immunofluorescence protein detection of one or two targets on the same fresh-frozen or FFPE tissue section, although at the cost of the otherwise applied H&E staining used for spatial mapping[115] (Fig. 5d). However, Spatial PrOtein and Transcriptome Sequencing (SPOTS) demonstrates that Visium is also compatible with co-profiling of a larger number of proteins using polyadenylated antibody-derived tag (ADT)-conjugated antibodies (Box 1), as was shown for 21 proteins in mouse spleen and breast cancer tissue cryosections[116]. Similarly, Spatial Multi-Omics (SM-Omics) shows Spatial Transcriptomics (the forerunner of Visium with 100-µm resolution) to be compatible with DNA-barcoded antibody and/or immunofluorescence co-mapping of six proteins, with the added concept of full automation using liquid-handling robotics achieving increased throughput and performance[117].

Alternatively, NanoString GeoMx Digital Spatial Profiling (DSP) allows quantification of the abundance of RNAs and/or proteins by counting unique indexing oligonucleotides, which are covalently linked via a UV-photocleavable linker with probes or antibodies that target transcripts or proteins of interest, respectively[118,119] (Fig. 5e). Whole-transcriptome DSP assays or DSP assays for >100 protein targets are possible, although coupled to image-based selection of specific regions of interest following fluorescent antibody and/or RNAscope[120] marker stainings. Selected regions of interest are illuminated by directed UV light, which cleaves the photocleavable linkers and releases oligonucleotide indices, enabling their collection and identification using NGS. Achieving single-cell resolution is challenging with this technology, but the combination of both transcriptome and protein readouts on the same sample was shown in FFPE cell pellet array sections stained with the GeoMx Human Whole Transcriptome Atlas (WTA) probe set and a 59 GeoMx antibody panel[121].

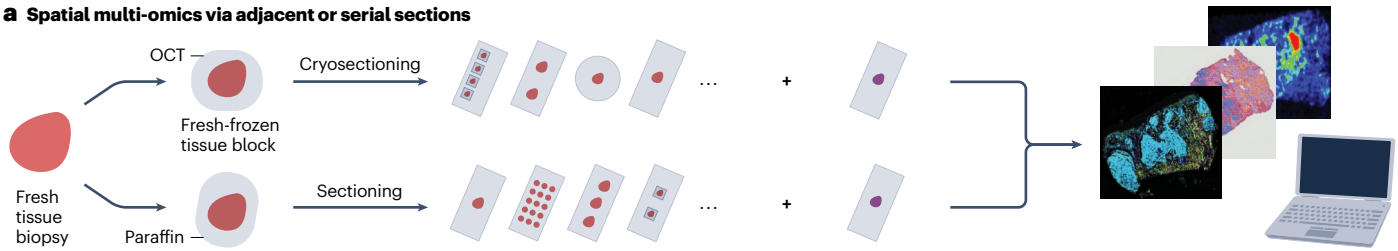
Chip-based DBiT-seq approaches (Box 1) also allow co-mapping of poly(A) mRNAs with proteins in tissue cryosections, as shown in the original DBiT-seq method with ~10-µm pixel resolution and tens of proteins co-profiled[103], but improved upon in spatial-CITE-seq (spatial co-indexing of transcriptomes and epitopes), in which whole-transcriptome and co-mapping of ~200–300 proteins was achieved at 20-µm pixel resolution[122] (Fig. 5b). Both these methods employ cocktails of ADT-conjugated antibodies (Box 1) each bearing a poly(A) oligonucleotide with an antibody-specific barcode. The channel-specific set 'A' barcodes, delivered by the first chip with equidistant channels placed on the tissue, consist of barcoded poly(T) DNA adapters also containing a universal ligation linker that hybridizes to both poly(A) ADT and mRNA molecules, and are then incorporated during in situ RT. The second chip introduces perpendicularly a set of channel-specific set 'B' barcodes that are ligated to the 'A' barcodes and also contain a PCR handle functionalized with biotin. The specific 'AB' barcoded cDNA and ADTs at the channel crossroads (pixels) are collected, purified with streptavidin-coated magnetic beads and prepared for NGS, after which reads for both mRNAs and proteins can be superimposed on an image of the section[103] (Fig. 5b). Spatial-CITE-seq was applied for improved tissue mapping to several mouse tissue types, as well as human tonsillar and skin biopsy tissue at the COVID-19 mRNA-vaccine injection site[122]. Although limited by the same aspects as previously mentioned for DBiT-based methods, including lack of single-cell resolution, to date spatial-CITE-seq is the method that allows for the highest number of proteins to be simultaneously spatially profiled together with the poly(A) transcriptome, with potential for further expansion.

As to microscopy-based approaches, several of the established seqFISH-based, MERFISH-based and in situ sequencing (ISS)-based methods for spatial transcript profiling of hundreds to thousands of targeted genes (Box 1) are also (sequentially) compatible with immunofluorescence or DNA-conjugated antibody protein readouts in the same sample (Fig. 1). For the latter, oligonucleotide sequences compatible with the specific fluorescent readout approaches of the transcriptome profiling methods are mostly used (Fig. 5c)[107,111,113,123–125]. Staining and localization of cell boundaries or other cellular, nuclear or subnuclear markers in these ways can enable more accurate cell segmentation, transcript allocation or resolution of nuclear organization. Additionally, combined spatial transcriptome and protein readouts on the same tissue section are crucial to correlate transcript with protein expression, localization and interactions, to help unravel the cellular mechanisms that govern specific cell types and states. Several platforms that recently became commercially available for automated imaging-based spatial transcriptome profiling of tissue sections at single-cell resolution, including the NanoString CosMx (smFISH-based)[126], Vizgen MERSCOPE (MERFISH-based)[127] and 10x Genomics Xenium (ISS-based)[128] platforms, will allow for imaging-based co-profiling of a few to tens of proteins, making these technologies more accessible.

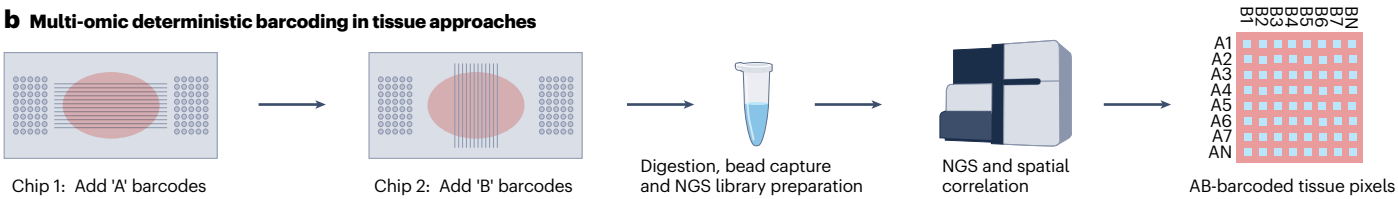
Data integration

A wide variety of algorithms have been described for the tailored data analysis of individual analytes in single cells. Here, we focus on computational advances that allow the integration of measurements across several modalities, as these algorithms leverage the multi-omic nature of these technologies to enhance our understanding of complex cellular states. Each modality covers different aspects of cellular identity and has its strengths and weaknesses. A principal goal of multi-omics data integration is to achieve robust and sensitive cell type or cell state identification[129]. This integrated multi-omics view of cellular identity can improve our understanding of differentiation trajectories, their underlying gene regulatory networks, cell–cell interactions, microenvironmental spatial organization, cellular lineages and clonal dynamics. Ultimately, a holistic view of cellular identity disentangles causal relationships between the different molecular layers that give rise to the observed cellular phenotypes. However, to achieve meaningful integration of high-dimensional data modalities, computational and statistical models need to be developed that consider the technical and biological intricacies of these technologies.

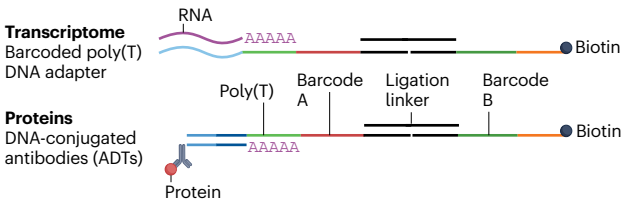
a Spatial multi-omics via adjacent or serial sections



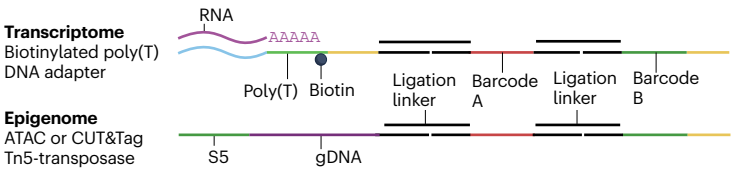
b Multi-omic deterministic barcoding in tissue approaches



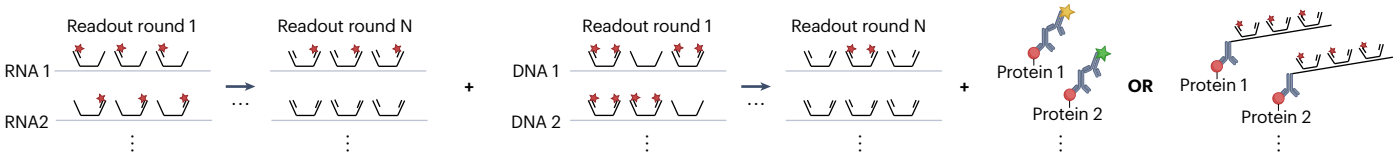
DBiT-seq and Spatial CITE-seq



ATAC&RNA-seq and CUT&Tag-RNA-seq



c Multi-omic single-molecule fluorescent in situ hybridization approaches



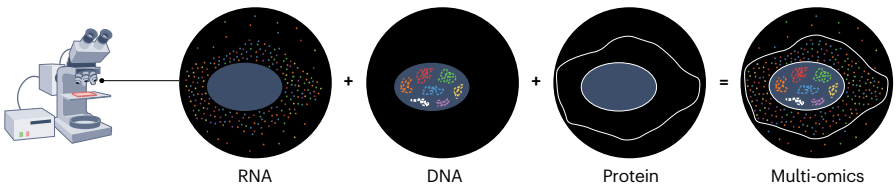
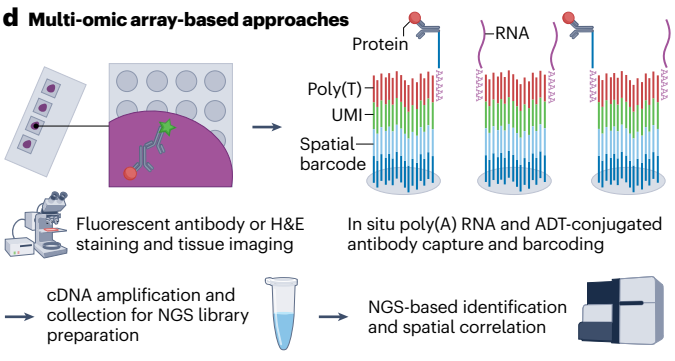


Image registration and decoding of optical barcodes

Readout round	1	2	3	4	5	...	N_{-2}	N_{-1}	N
Target analyte 1	1	0	0	0	1	...	0	0	1
Target analyte 2	1	0	0	0	0	...	0	0	0
⋮									

d Multi-omic array-based approaches



e Multi-omic Digital Spatial Profiling approach

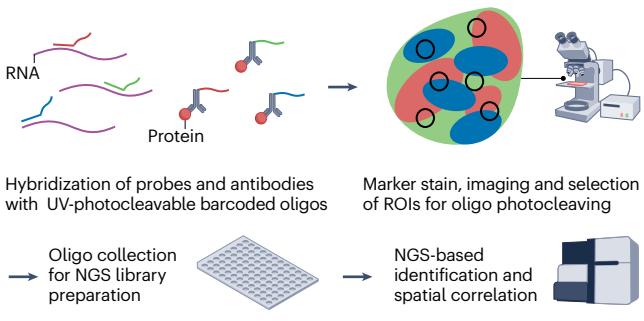


Fig. 5 | Methods for spatial multi-omics. Spatial multi-ome profiling of tissue samples can be achieved by applying spatial mono-omics assays separately on adjacent or serial tissue sections (part **a**) or in a combined way on the same tissue section (parts **b**–**e**). **a**, Serial fresh-frozen or formalin-fixed paraffin-embedded (FFPE) tissue sections can be analysed using different spatial mono-omic assays, potentially also combining with morphological stainings and annotations on the same or adjacent sections, followed by computational data integration. **b**, Microfluidic deterministic barcoding strategies in tissue allow next-generation sequencing (NGS)-based spatial multi-omics profiling of transcriptome-plus-proteins, as in DBiT-seq[103] and Spatial-CITE-seq[122], and epigenome-plus-transcriptome, as in ATAC&RNA-seq and CUT&Tag-RNAseq[100]. Using dual microfluidic chip-based spatial barcoding of poly(A) RNAs together with proteins or epigenome information at the crossroads of chip channels, a spatially barcoded 2D pixel map of the tissue is created. **c**, Advanced fluorescence in situ hybridization (FISH)-based methods, including MERFISH[107,108,111] and seqFISH+[109,110,112], allow microscopy-based identification of thousands of transcripts together with genomic loci in single cells, in addition to being compatible with limited protein readouts using fluorescent or DNA-conjugated antibody readout strategies. These high-resolution imaging methods leverage predefined optical barcoding schemes and complex encoding and readout probe designs. **d**, Array-based assays, including Spatial Transcriptomics[170] (ST) and 10x Genomics Visium[115], make use of slides with arrayed oligo-dT spots for capturing and spatial barcoding of poly(A) RNAs followed by NGS profiling. This can be combined with upfront haematoxylin and eosin (H&E) staining or limited protein antibody staining and tissue imaging for spatial mapping. In SM-Omics[117] and SPOTS[116], these technologies have also been shown to be compatible with antibody-derived tag (ADT)-conjugated antibody-based co-profiling of a larger number of proteins. **e**, NanoString GeoMx digital spatial profiling (DSP)[118,119,121] allows quantification of RNAs and proteins in specific regions of interest (ROIs) by counting uniquely barcoded oligonucleotides that are covalently linked through a UV-photocleavable linker with probes or antibodies. Tissue marker staining, imaging, ROI selection and illumination by directed UV light causes disintegration of the photocleavable linkers that are collected and profiled by NGS, followed by spatial mapping to the ROIs. cDNA, complementary DNA; gDNA, genomic DNA; OCT, optimal cutting temperature compound; UMI, unique molecular identifier.

Recently, Argelaguet et al.[130] defined three categories of data integration strategies depending on the anchor used to link the different data modalities. Horizontal integration strategies use common data features measured across different datasets to integrate independently assayed groups of cells, such as when assaying different batches with the same technology or when integrating across different technologies measuring the same molecular analyte. Inversely, vertical integration strategies use the cell as the anchoring unit to integrate non-overlapping data features, such as when measuring multiple omics layers of the same cells in parallel (Fig. 6a). The hardest integration problem emerges when neither cells nor common data features can be used as anchors. In this case, diagonal integration strategies are used to map separate groups of cells profiled by different molecular assays (Fig. 6b). Many computational methods have already been developed to tackle horizontal integration problems, as cell atlas efforts require the integration of many batches of single-cell data. For a comparison of these approaches, we refer to recent reviews and benchmark studies[130–133]. Here, we focus primarily on vertical integration approaches as these are applicable to most multi-omics technologies in which parallel measurements are made of individual cells.

Linking molecular layers

When paired data modalities can be unambiguously assigned to individual cells, several integration strategies can be applied depending on the end goal of the analysis. For finding significant cross-modal associations between specific features (coined local integration methods by Argelaguet et al.[130]), classic regression-based statistical models can be fitted in a supervised way[134]. These models can be expanded to handle common confounders (such as population stratification biases, sample-specific effects and sequence context-dependent effects), which might inflate the resulting association statistics, by explicitly modelling these confounders using linear hierarchical mixed models[135]. However, hypothesis testing the very large number of possible correlations of analytes between different molecular layers would require unattainable sample sizes after multiple testing correction. Therefore, the hypothesis search space is often constrained by previous biological information (for example, the distance from the epigenetic marker or accessible regulatory element to the expressed gene). These approaches can be used to identify, for example, cell-type-specific allelic effects of genetic variation on gene expression (expression quantitative trait loci (eQTLs)) in differentiating induced pluripotent stem cells (iPS cells)[136] or the effect of genetic perturbations on gene expression using multiplexed CRISPR technologies to identify *cis*-acting enhancer–gene pairs[137]. Local integration approaches can also be used to facilitate the inference of gene regulatory networks (GRNs). Classic GRN inference approaches, using only scRNA-seq data, identify sets of co-expressed genes enriched with particular upstream regulatory motifs[138]. However, this approach is prone to false-positive gene associations as the presence of these motifs does not necessarily guarantee the activity of the associated TF. Using methods for simultaneous analysis of open chromatin and gene expression with sequencing, it has been shown that chromatin accessibility at regulatory domains precedes gene expression[38] and may thus outperform methods for GRN inference from scRNA-seq alone.

In contrast to local approaches, global integration strategies aim to identify larger-scale patterns (hundreds to thousands of features) of covariation across modalities to identify global shifts in the cellular state in an unsupervised way. To this end, linear matrix decomposition approaches such as principal component analysis (PCA), canonical correlation analysis (CCA)[139] and non-negative matrix factorization (NMF)[140,141] have proven successful to identify latent factors underlying cellular heterogeneity. For example, MOFA+[142] extends on this principle and builds on the Bayesian Group Factor Analysis framework to jointly model variation across spatial and temporal covariates and employs a sparse hierarchical prior architecture that handles both shared sources of variation and private variation that is particular to individual modalities. This approach was used to integrate the different modalities in scNMT-seq data to reveal lineage-specific enhancers during embryo development[68]. Although the linearity of these approaches makes them inherently robust and results in readily interpretable factors of variation, they suffer from a lack of explanatory power when dealing with non-linearities that are commonly found in complex biological systems. Nonlinear multi-view neural network architectures[143,144] such as multimodal autoencoders have been proposed to tackle this limitation. These neural networks aim to learn simple (low-dimensional) representations of the data, also known as an encoding, from complex (high-dimensional) input datasets. Instead of learning individual representations for each modality, these representations can be forced to be

shared across different data modalities, thus forcing the neural network to implicitly learn common links between them. Several alternative architectures have been proposed that optimize for different criteria such as robustness to dropouts and batch effects, disentanglement of the learned latent factors for improved interpretability and cross-modal translation for missing modality imputation[145,146]. For further details we refer the reader to a recent overview of proposed deep learning approaches[147].

Data integration for spatial multi-omics

Spatial mono-omics and multi-omics technologies also benefit from various data integration strategies. Before applying

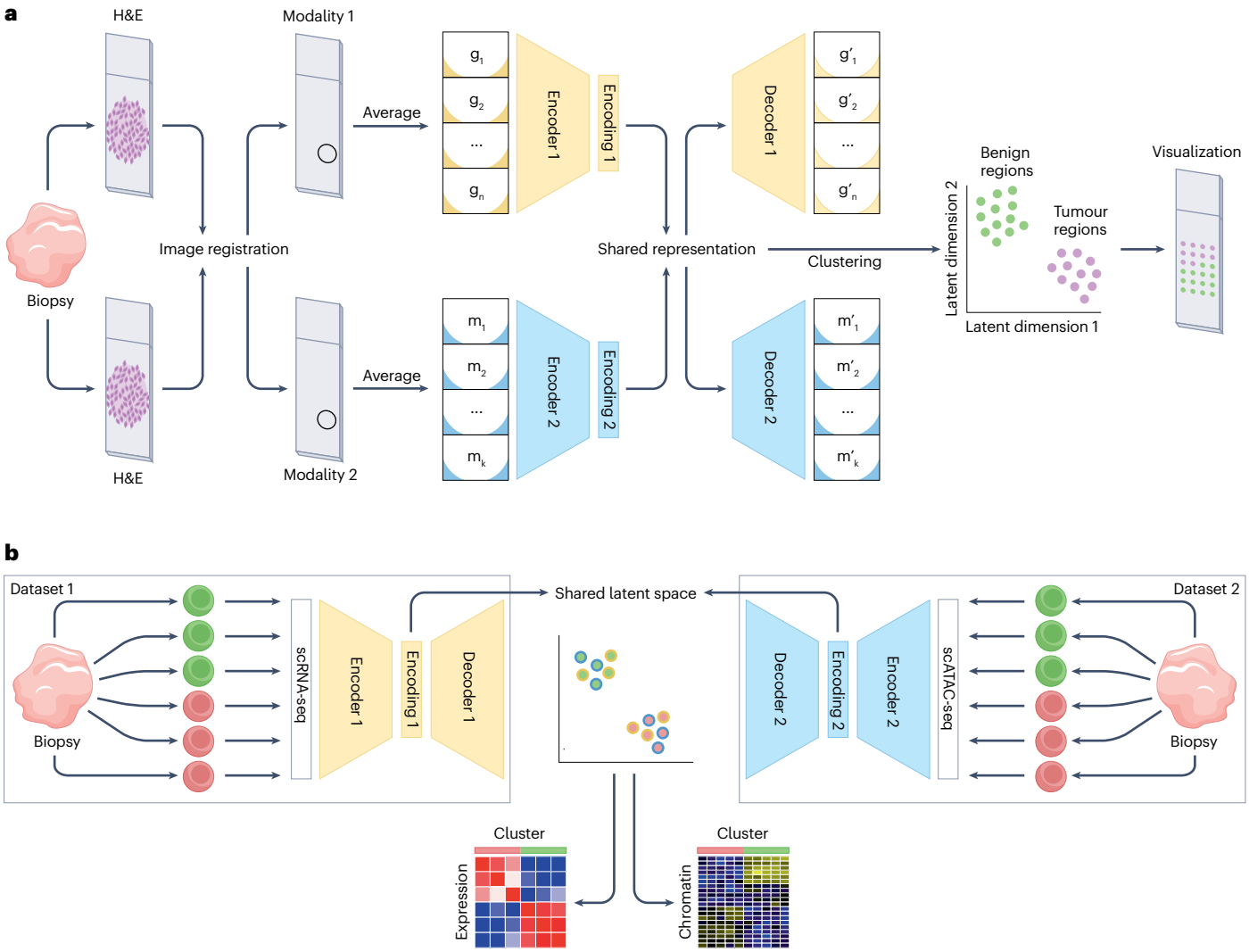


Fig. 6 | Data integration strategies. Examples of different scenarios in which various types of data integration strategies can be used. **a**, Vertical integration strategies aim to integrate information from paired molecular layers to obtain holistic representations of biological systems, at the single-cell level or at the tissue region level. Here, we illustrate an example of a spatial multi-omics experiment in which mirrored tissue slices have been assayed by two different spatial modalities (yellow and blue). To integrate both data sources, haematoxylin and eosin (H&E) staining images of each modality are first registered to account for deformations during sample preparation. Subsequently, to account for differences in resolutions across modalities, data points are averaged in windows of a predetermined region size[154]. For every region the averaged regional profiles can be used as paired inputs for either linear or nonlinear vertical integration approaches. In this example, we illustrate integration through the use of a multi-view autoencoder neural network[143]. Each modality is used as input into a dedicated encoder–decoder network that learns a shared data representation, effectively integrating both data modalities. This shared representation can be used for downstream analysis and/or visualization. The terms g_i and m_i correspond to the measurements of region *i* for modality 1 and modality 2, respectively, with g'_i and m'_i being the molecular profiles reconstructed from the shared representation by the decoder networks. **b**, Integration of datasets from different experiments with independent observations of individual cells and non-overlapping molecular features is the hardest integration problem and requires diagonal integration approaches. Here, we illustrate this problem for the integration of independently acquired single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin (scATAC-seq) datasets through the use of autoencoder neural networks with a probabilistic coupling to map the different data modalities to a shared latent space[144]. Although single-cell resolution is lost in this coupling, clusters obtained in this shared latent space can be used to ascertain correlations between molecular layers, discover multimodal biomarkers and/or translate between the different modalities.

Glossary

3C or Hi-C
A class of methods uncovering the higher-order chromosome conformation and chromatin interactions in the nucleus.

Autoencoders
A class of architectures of neural networks that take a highly dimensional input and encode into a low-dimensional representation via an encoder network. These networks are trained in a self-supervised way by reconstructing the original input from the encoding using a decoder network.

Bisulfite sequencing
A method in which bisulfite treatment of DNA before sequencing converts unmethylated cytosines to uracil, while 5-methylcytosines are protected from conversion. This method hence discloses sites in the genome where DNA methylation is found.

Cell plasticity
The ability of a cell to remodel its epigenomic, transcriptomic and proteomic landscape, leading to new phenotypic features.

Cell state
A specific stable or unstable functional condition of a cell belonging to a particular cell type.

Cell type
A categorization of cells by specific morphological and/or phenotypic characteristics.

Chromatin accessibility
The structures of DNA wrapped around histone octamer proteins (nucleosomes) that make up chromosomes. Chromatin can be highly condensed with DNA inaccessible or open with DNA accessible.

Displacement preamplification followed by PCR
(DA-PCR). A class of methods for whole-genome amplification of single-cell DNA using multiple rounds of strand displacement amplification, generating amplicons with PCR-handles at both ends, enabling molecules to be further amplified by PCR.

DNA methylation
The deposition of methyl groups onto DNA. In mammals, methylation is predominantly found at cytosines as 5-methylcytosine, and mostly in a cytosine–guanine dinucleotide (CpG) context.

Epigenome
The ensemble of modifications to DNA and DNA-associated proteins that signal and regulate gene expression and other DNA-related processes. Examples include DNA methylation, histone post-translational modifications and chromatin remodelling proteins.

Exome sequencing
A targeted next-generation sequencing approach that enables sequencing the coding exons, which can be supplemented with flanking intronic sequences as well as (part of) the gene regulatory sequences, approximating 1.1% or more of the human genome.

FFPE
Formalin-fixed paraffin-embedded (FFPE) tissues are preserved for long-term archival storage by formalin fixation followed by embedding in paraffin wax blocks.

Genome
The genetic blueprint or DNA of an organism established at fertilization, which for humans normally consists of 23 pairs of chromosomes contained within the nucleus, plus the mitochondrial DNA molecules present at multiple copies in the mitochondria of the cell.

Hierarchical mixed models
Statistical regression models that model nested hierarchies in the data taking into account complex covariate data structures.

Human Cell Atlas
The Human Cell Atlas (HCA) aims to create a reference map of all human cells, providing a basis for understanding human health and for diagnosing, monitoring and treating diseases.

Linear matrix decomposition
Numerical analysis algorithms that factorize observations into a product of smaller matrices. These matrices usually represent factors or components of variation that are easier to interpret (such as biological processes) and their presence within the observations (cells).

Metabolome
The ensemble of metabolites present within a cell or tissue.

Multiple displacement amplification
(MDA). An isothermal DNA amplification method that applies DNA polymerase(s) with strand displacement activity.

Neural networks
A class of algorithms that use densely connected networks of artificial neurons that are non-linearly activated given a combination of input values from connected input neurons, mimicking the human brain.

OCT compound
Optimal cutting temperature (OCT) compound is a standard type of sample-embedding medium used to embed fresh-frozen tissue samples for optimal storage and cryosectioning.

Phenome
The set of traits or characteristics expressed by an organism or cell.

Proteome
The ensemble of proteins translated from transcribed genes contained within the genome, which can be highly variable among different cell types.

Smart-seq2
Switching mechanism at the end of the 5′ end of the RNA transcript (Smart)-based method for single-cell cDNA generation, amplification and sequencing library conversion.

Tagmentation
A method in which double-stranded DNA is cleaved and tagged with adapter sequences in a single step by using a transposase complex loaded with these adapter sequences.

Template switching
A method in which Moloney murine leukaemia virus (MMLV)-type reverse transcriptases add non-templated nucleotides (CCC) to first-strand cDNA near the 5′ end of the transcript, enabling annealing of an rGrGrG-containing template-switching oligonucleotide and the reverse transcriptase to switch templates and copy the template-switching oligonucleotide sequence to the cDNA.

Transcriptome
The ensemble of transcripts or RNA molecules transcribed from genes contained within the genome, which can be highly variable among different cell types.

Unique molecular identifier
Short sequences that barcode each molecule individually and are added as tags to DNA fragments in next-generation sequencing-based approaches to identify molecules of interest with increased accuracy.

vertical integration, horizontal integration algorithms can be used to correct for technical variability between different sections of the same or different samples, and can also be expanded to fully utilize spatial correlations between sections[148]. Additionally, for spatial omics technologies that do not achieve single-cell resolution, deconvolution methods can be used to integrate non-spatial

single-cell data to identify the cellular constituents of a particular tissue region[149–151].

Vertical integration of non-spatial single-cell multi-omics data followed by spatial deconvolution using the RNA modality allows other molecular modalities to be indirectly spatially mapped. This approach was used, for example, by Foster et al., using ArchR[152] for scRNA-seq and scATAC-seq integration, to identify spatial fibroblast epigenome patterns during tissue repair[153]. Spatial multi-omics technologies also allow for vertical integration, although few currently available technologies allow for true multimodal assaying at single-cell resolution on the same tissue section. Hence, generating spatial multi-omics data often still relies on the analysis of adjacent tissue sections and is thus limited to studying correlations across modalities of tissue regions (Fig. 6a). Linear approaches are suitable for the inference of these correlations, although special care needs to be taken in the statistical analysis as spatial autocorrelation can violate independence assumptions of the observed datapoints[154,155]. Although more spatial multi-omics technologies are emerging, spatial mono-omics assays can often also be combined with simple histological stains, such as H&E staining, allowing for the integration of cellular morphology and gene expression data. Nonlinear neural networks have been proposed to map discrete spot-based expression profiles onto high-resolution morphology images[156] or to learn a joint cell morphology-plus-expression representation for the identification of novel cellular subpopulations missed by individual modalities[157]. As more spatial multi-omics technologies become available that can characterize many analytes at single-cell resolution, we expect rapid developments of vertical integration strategies based on adaptations of currently available non-spatial single-cell methodologies. By leveraging spatial information linking a cellular state to its respective micro-environments and macro-environments (for example, through the use of graph neural networks), more fine-grained multimodal representations of cellular state should be obtainable.

Finally, approaches for the harder problem of diagonal integration of unpaired spatial and non-spatial data modalities have also been proposed using autoencoder neural network architectures with a probabilistic coupling between the different data modalities. For these approaches, however, previous knowledge is generally required to constrain the resulting shared representations to biologically accurate alignments[144].

Current challenges in data integration

Despite extensive research, several challenges remain for computational data integration. These strategies make implicit assumptions about the expected similarity of cellular states captured across different experiments and/or modalities. In the case of horizontal data integration across different batches, this can lead to overcorrection of true biological variation[158], especially when studying subtle shifts in cellular state in different experimental conditions. This problem is further exacerbated in vertical integration strategies when correspondence between features is not immediately obvious and/or when non-linearities between the different layers might be biologically relevant[132]. For example, gene expression changes have been shown to be foreshadowed by chromatin priming, as seen in cell cycling via SHARE-seq[38]. Also, other biological phenomena such as alternative splicing and/or PTMs are likely to influence the correlation structure in cell state across the different molecular layers. Therefore, integration methods need to be developed that can not only identify the common anchors between the different samples and/or molecular layers but can also account for sample-specific and modality-specific variation in a readily interpretable way. Such models have already been proposed in other fields of multi-domain learning but have not yet been adapted to cellular multi-omics datasets[159]. The expansion of these models to integrate paired, unpaired and multimodal spatial assays will require robust data standards[160,161] for scalable analysis of multimodal data in addition to well-established benchmarks to evaluate their efficacy.

Perspectives

In the decades to come, multi-omics at single-cell and spatial resolution will innovate further, leading to a more holistic understanding of cell biology. Advances can be anticipated on multiple fronts, including improvements in throughput, reduction of cost and the incorporation of more modalities in a single assay. Additionally, we expect improvements of sensitivity and specificity in the detection and characterization of each modality as part of multi-omic measurements. For example, on the genome level, full and error-free characterization of all genetic variants is still a challenge, which currently limits opportunities for comprehensive somatic mutation profiling at single-cell resolution and reconstructing phylogenetic cell lineages from naturally acquired mutations. Similarly, measurements of the epigenome are severely limited for co-detection of the range of epigenomic features co-regulating gene expression and other DNA-related processes. For instance, histone PTMs can currently be detected only a single or a few marks at a time; hence, these methods would benefit from drastic increments in the number of PTMs that can be co-detected, also together with other epigenomic features. The characterization of the transcriptome is often limited to poly(A) RNA rather than total RNA measurements; hence, it would be beneficial to include both coding and (small) non-coding RNAs as well as concurrent isoform detection. Proteome assays are still antibody-based and thus limited in how many proteins can be profiled simultaneously. Unbiased low-input methods such as mass-spectrometry-based approaches may circumvent this, but cannot currently be combined with assays for other molecular layers. This lack of multimodal integration also holds true for metabolome and lipidome assays. In addition to such improvements for characterizing modalities in multi-omics methods, we also anticipate the development of multimodal assays that incorporate entirely new modalities that currently remain uncharted, such as the epitranscriptome (consisting of base modifications to transcripts, some of which are known to affect gene expression)[162]. Furthermore, the field is likely to see a continued strong technological push for spatial multi-omics, avoiding the need for tissue dissociation and enabling concurrent multi-omic profiling of cell-intrinsic and -extrinsic molecular features defining cell types and states. Moreover, integration of phylogenetic cell lineages reconstructed from naturally acquired or artificially induced DNA mutations with other spatial or single-cell multi-ome information will transform our understanding of organismal development, cell migration routes and stem cell biology in health and disease. Finally, it will be necessary to develop methods that capture not only transient phenotypes but also ancestral states, apply multi-omic technologies to serial measurements in live cells[163], and computationally improve the accuracy of data extraction from each molecular layer, as well as undertake integrative analyses across modalities to unravel dependencies within and across different data sources.

To arrive at such holistic multi-ome profiling of single cells dissociated from tissue or at spatial resolution will require many challenges to be overcome in the following years. For a view on how challenges associated with each modality specifically could be tackled, we refer to previous reviews on single-cell and spatial mono-omic profiling[11,86,164–168].

Optimization of low-input bulk molecular analysis tools to single-cell and/or spatial resolution, like amplification-free long-read single-molecule sequencing approaches, will enable the incorporation of novel molecular readouts. In parallel, to enable integration of more modalities per assay will require innovation in the principles presented in Fig. 2, including expansion and combination of the different principles. Making various single-cell and spatial multi-omics assays commercially available will also make them more accessible and applicable for the wide research community, but we are already starting to see this evolution[169].

All these technological and computational developments will translate into a better understanding of development, organismal function and functional decline of organs with ageing, and will be key in unravelling the cellular pathogenesis of diseases, identifying more effective stratification strategies of disease processes, devising novel therapeutic modalities and guiding precision medicine.

Published online: 02 March 2023

References

1. Regev, A. et al. The Human Cell Atlas. *eLife* **6**, e27041–e27041 (2017).
2. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
3. Grün, D. & van Oudenaarden, A. Design and analysis of single-cell sequencing experiments. *Cell* **163**, 799–810 (2015).
4. Elmentaite, R., Domínguez Conde, C., Yang, L. & Teichmann, S. A. Single-cell atlases: shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.* **23**, 395–410 (2022).
5. Wen, L. & Tang, F. Recent advances in single-cell sequencing technologies. *Precis. Clin. Med.* **5**, pbac002 (2022).
6. Macaulay, I. C. et al. Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nat. Protoc.* **11**, 2081–2103 (2016).
7. Macaulay, I. C. et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522 (2015).
   This paper presents one of the first demonstrations of physically separating DNA and poly(A) RNA from a single cell for whole-genome sequencing and full-length transcript sequencing.
8. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
9. Li, W., Calder, R. B., Mar, J. C. & Vijg, J. Single-cell transcriptogenomics reveals transcriptional exclusion of ENU-mutated alleles. *Mutat. Res. Mol. Mech. Mutagen.* **772**, 55–62 (2015).
10. Macaulay, I. C. & Voet, T. Single cell genomics: advances and future perspectives. *PLoS Genet.* **10**, e1004126 (2014).
11. Gawad, C., Koh, W. & Quake, S. R. Single-cell genome sequencing: current state of the science. *Nat. Rev. Genet.* **17**, 175–188 (2016).
12. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).
13. Han, K. Y. et al. SIDR: simultaneous isolation and parallel sequencing of genomic DNA and total RNA from single cells. *Genome Res.* **28**, 75–87 (2018).
14. Zachariadis, V., Cheng, H., Andrews, N. & Enge, M. A highly scalable method for joint whole-genome sequencing and gene-expression profiling of single cells. *Mol. Cell* **80**, 541–553.e5 (2020).
    This paper presents the feasibility of separating cytoplasmic RNA from nuclear DNA, enabling separate RNA-seq and DNA-seq library construction through direct tagmentation of the DNA, thus avoiding upfront WGA.
15. Valecha, M. & Posada, D. Somatic variant calling from single-cell DNA sequencing data. *Comput. Struct. Biotechnol. J.* **20**, 2978–2985 (2022).
16. Han, L. et al. Co-detection and sequencing of genes and transcripts from the same single cells facilitated by a microfluidics platform. *Sci. Rep.* **4**, 6485 (2014).
17. van Strijp, D. et al. Complete sequence-based pathway analysis by differential on-chip DNA and RNA extraction from a single cell. *Sci. Rep.* **7**, 11030 (2017).
18. Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* **33**, 285–289 (2015).
    This paper presents one of the first demonstrations of joint preamplification of single-cell gDNA and poly(A) RNA-derived cDNA, followed by splitting of the reaction for DNA-seq and RNA-seq library preparation.
19. Rodriguez-Meira, A. et al. Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. *Mol. Cell* **73**, 1292–1305.e8 (2019).
20. Yu, L. et al. scONE-seq: a single-cell multi-omics method enables simultaneous dissection of phenotype and genotype heterogeneity from frozen tumors. *Sci. Adv.* **9**, eabp8901 (2023).
21. Yin, Y. et al. High-throughput single-cell sequencing with linear amplification. *Mol. Cell* **76**, 676–690.e10 (2019).
    This paper presents one of the first demonstrations of combinatorial indexing for joint single-cell DNA and RNA sequencing, enabling massive throughput.
22. Rambow, F. et al. Toward minimal residual disease-directed therapy in melanoma. *Cell* **174**, 843–855.e19 (2018).
23. Miller, M. B. et al. Somatic genomic changes in single Alzheimer's disease neurons. *Nature* **604**, 714–722 (2022).
24. Bizzotto, S. & Walsh, C. A. Genetic mosaicism in the human brain: from lineage tracing to neuropsychiatric disorders. *Nat. Rev. Neurosci.* **23**, 275–286 (2022).
25. Proukakis, C. Somatic mutations in neurodegeneration: an update. *Neurobiol. Dis.* **144**, 105021 (2020).
26. Voet, T. & Vermeesch, J. R. Mutational processes shaping the genome in early human embryos. *Cell* **168**, 751–753 (2017).
27. van Neerven, S. M. & Vermeulen, L. Cell competition in development, homeostasis and cancer. *Nat. Rev. Mol. Cell Biol.* https://doi.org/10.1038/s41580-022-00538-y (2022).
28. Ren, P., Dong, X. & Vijg, J. Age-related somatic mutation burden in human tissues. *Front. Aging* **3**, 1018119 (2022).
29. Bekaert, B. et al. CRISPR/Cas gene editing in the human germline. *Semin. Cell Dev. Biol.* **131**, 93–107 (2022).
30. Alanis-Lobato, G. et al. Frequent loss of heterozygosity in CRISPR-Cas9–edited early human embryos. *Proc. Natl Acad. Sci. USA* **118**, e2004832117 (2021).
31. Fogarty, N. M. E. et al. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **550**, 67–73 (2017).
32. Sheth, R. U. & Wang, H. H. DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* **19**, 718–732 (2018).
33. Bhattarai-Kline, S. et al. Recording gene expression order in DNA by CRISPR addition of retron barcodes. *Nature* **608**, 217–225 (2022).
34. Choi, J. et al. A time-resolved, multi-symbol molecular recorder via sequential genome editing. *Nature* **608**, 98–107 (2022).
35. Liu, L. et al. Deconvolution of single-cell multi-omics layers reveals regulatory heterogeneity. *Nat. Commun.* **10**, 470 (2019).
36. Cheng, H. et al.Smart3-ATAC: a highly sensitive method for joint accessibility and full-length transcriptome analysis in singlecells. Preprint at *bioRxiv* https://doi.org/10.1101/2021.12.02.470912 (2021).
37. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).
38. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116.e20 (2020).
    SHARE-seq is a powerful method to profile gene expression and chromatin accessibility in the same cell; combining both layers of information shows that changes in enhancer activity can precede gene activation.
39. Plongthongkum, N., Diep, D., Chen, S., Lake, B. B. & Zhang, K. Scalable dual-omics profiling with single-nucleus chromatin accessibility and mRNA expression sequencing 2 (SNARE-seq2). *Nat. Protoc.* **16**, 4992–5029 (2021).
40. Zhu, C. et al. An ultra high-throughput method for single-cell joint analysis of open chromatin and transcriptome. *Nat. Struct. Mol. Biol.* **26**, 1063–1070 (2019).
41. Chen, S., Lake, B. B. & Zhang, K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. *Nat. Biotechnol.* **37**, 1452–1457 (2019).
42. Xing, Q. R. et al. Parallel bimodal single-cell sequencing of transcriptome and chromatin accessibility. *Genome Res.* **30**, 1027–1039 (2020).
43. Xu, W. et al. ISSAAC-seq enables sensitive and flexible multimodal profiling of chromatin accessibility and gene expression in single cells. *Nat. Methods* **19**, 1243–1249 (2022).
44. No authors listed. Sensitive, flexible and modular single-cell multi-omics profiling with ISSAAC-seq. *Nat. Methods* **19**, 1183–1184 (2022).
45. Tedesco, M. et al. Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. *Nat. Biotechnol.* **40**, 235–244 (2022).
46. Pan, L. et al. scPCOR-seq enables co-profiling of chromatin occupancy and RNAs in single cells. *Commun. Biol.* **5**, 678 (2022).
47. Xiong, H., Luo, Y., Wang, Q., Yu, X. & He, A. Single-cell joint detection of chromatin occupancy and transcriptome enables higher-dimensional epigenomic reconstructions. *Nat. Methods* **18**, 652–660 (2021).
48. Zhu, C. et al. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat. Methods* **18**, 283–292 (2021).
49. Sun, Z. et al. Joint single-cell multiomic analysis in Wnt3a induced asymmetric stem cell division. *Nat. Commun.* **12**, 5941 (2021).
50. Janssens, D. H. et al. CUT&Tag2for1: a modified method for simultaneous profiling of the accessible and silenced regulome in single cells. *Genome Biol.* **23**, 81 (2022).
51. Gopalan, S., Wang, Y., Harper, N. W., Garber, M. & Fazzio, T. G. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol. Cell* **81**, 4736–4746.e5 (2021).
52. Liu, Y. et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* **37**, 424–429 (2019).
53. Vaisvila, R. et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res.* **31**, 1280–1289 (2021).
54. Hu, Y. et al. Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol.* **17**, 88 (2016).
55. Angermueller, C. et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13**, 229–232 (2016).

56. Hou, Y. et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res.* **26**, 304–319 (2016).
57. Bian, S. et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060–1063 (2018).
58. Gu, H. et al. Smart-RRBS for single-cell methylome and transcriptome analysis. *Nat. Protoc.* **16**, 4004–4030 (2021).
59. Clark, S. J. et al. scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat. Commun.* **9**, 781 (2018).
60. Li, L. et al. Single-cell multi-omics sequencing of human early embryos. *Nat. Cell Biol.* **20**, 847–858 (2018).
61. Gu, C., Liu, S., Wu, Q., Zhang, L. & Guo, F. Integrative single-cell analysis of transcriptome, DNA methylome and chromatin accessibility in mouse oocytes. *Cell Res.* **29**, 110–123 (2019).
62. Pott, S. Simultaneous measurement of chromatin accessibility, DNA methylation, and nucleosome phasing in single cells. *eLife* **6**, e23203 (2017).
63. Wang, Y. et al. Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nat. Commun.* **12**, 1247 (2021).
64. Yan, R. et al. Decoding dynamic epigenetic landscapes in human oocytes using single-cell multi-omics sequencing. *Cell Stem Cell* **28**, 1641–1656.e7 (2021).
65. Luo, C. et al. Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genomics* **2**, 100107 (2022).
66. Lee, D.-S. et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019).
67. Li, G. et al. Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* **16**, 991–993 (2019).
68. Argelaguet, R. et al. Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).
By applying scNMT-seq to the early mouse development, this paper highlights the potential of this multi-omic method to disentangle hierarchies of transcriptional and epigenetic changes beyond mono-omic analyses.
69. Liu, X. et al. N^6-methyladenine is incorporated into mammalian genome by DNA polymerase. *Cell Res.* **31**, 94–97 (2021).
70. Rooijers, K. et al. Simultaneous quantification of protein–DNA contacts and transcriptomes in single cells. *Nat. Biotechnol.* **37**, 766–772 (2019).
71. Borsos, M. et al. Genome–lamina interactions are established de novo in the early mouse embryo. *Nature* **569**, 729–733 (2019).
72. Aughey, G. N., Estacio Gomez, A., Thomson, J., Yin, H. & Southall, T. D. CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. *eLife* **7**, e32341 (2018).
73. Rang, F. J. et al. Single-cell profiling of transcriptome and histone modifications with EpiDamID. *Mol. Cell* **82**, 1956–1970.e14 (2022).
EpiDamID profiles histone modifications and transcription in the same cell; histone modification profiling occurs orthogonally to classic antibody-based approaches, and as the presence of modifications can be written on DNA over a longer period, such methods may generate time records.
74. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S. & Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).
75. Peterson, V. M. et al. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* **35**, 936–939 (2017).
76. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
77. Chung, H. et al. Joint single-cell measurements of nuclear proteins and RNA in vivo. *Nat. Methods* **18**, 1204–1212 (2021).
78. Reimegård, J. et al. A combined approach for single-cell mRNA and intracellular protein expression analysis. *Commun. Biol.* **4**, 624 (2021).
79. Mimitou, E. P. et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).
80. Gerlach, J. P. et al. Combined quantification of intracellular (phospho-)proteins and transcriptomics from fixed single cells. *Sci. Rep.* **9**, 1469 (2019).
81. Mimitou, E. P. et al. Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat. Biotechnol.* **39**, 1246–1258 (2021).
82. Swanson, E. et al. Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *eLife* **10**, e63632 (2021).
83. Zhang, B. et al. Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. *Nat. Biotechnol.* **40**, 1220–1230 (2022).
84. Chen, A. F. et al. NEAT-seq: simultaneous profiling of intra-nuclear proteins, chromatin accessibility and gene expression in single cells. *Nat. Methods* **19**, 547–553 (2022).
85. Fiskin, E. et al. Single-cell profiling of proteins and chromatin accessibility using PHAGE-ATAC. *Nat. Biotechnol.* **40**, 374–381 (2022).
86. Mund, A., Brunner, A.-D. & Mann, M. Unbiased spatial proteomics with single-cell resolution in tissues. *Mol. Cell* **82**, 2335–2349 (2022).
87. Hu, W. et al. Single-cell transcriptome and translatome dual-omics reveals potential mechanisms of human oocyte maturation. *Nat. Commun.* **13**, 5114 (2022).
88. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
89. Jaitin, D. A. et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-Cell RNA-Seq. *Cell* **167**, 1883–1896.e15 (2016).
90. Dixit, A. et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016).
91. Rubin, A. J. et al. Coupled single-cell CRISPR screening and epigenomic profiling reveals causal gene regulatory networks. *Cell* **176**, 361–376.e17 (2019).
92. Liscovitch-Brauer, N. et al. Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. *Nat. Biotechnol.* **39**, 1270–1277 (2021).
93. Pierce, S. E., Granja, J. M. & Greenleaf, W. J. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. *Nat. Commun.* **12**, 2969 (2021).
94. Replogle, J. M. et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell* **185**, 2559–2575.e28 (2022).
By providing a first description of the full compendium of transcriptional changes induced by inactivating any expressed gene, this study demonstrates the potential of single-cell multi-omic methods for hypothesis-free genome-wide functional screening.
95. Norman, T. M. et al. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science* **365**, 786–793 (2019).
96. Eisenstein, M. Seven technologies to watch in 2022. *Nature* **601**, 658–661 (2022).
97. Yagnik, G., Liu, Z., Rothschild, K. J. & Lim, M. J. Highly multiplexed immunohistochemical MALDI-MS imaging of biomarkers in tissues. *J. Am. Soc. Mass. Spectrom.* **32**, 977–988 (2021).
98. Truong, J. X. M. et al. Removal of optimal cutting temperature (O.C.T.) compound from embedded tissue for MALDI imaging of lipids. *Anal. Bioanal. Chem.* **413**, 2695–2708 (2021).
99. Guilliams, M. et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379–396.e38 (2022).
100. Fan, R. et al. Spatially resolved epigenome-transcriptome co-profiling of mammalian tissues at the cellular level. *Res. Sq.* https://doi.org/10.21203/rs.3.rs-1728747/v1 (2022).
101. Deng, Y. et al. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609**, 375–383 (2022).
102. Deng, Y. et al. Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. *Science* **375**, 681–686 (2022).
103. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681.e18 (2020).
This is the first paper showing the possibility for spatial multi-omics (transcriptome+protein) profiling of tissue sections using a microfluidics-based deterministic barcoding-in-tissue approach.
104. Cardozo Gizzi, A. M. et al. Microscopy-based chromosome conformation capture enables simultaneous visualization of genome organization and transcription in intact organisms. *Mol. Cell* **74**, 212–222.e5 (2019).
105. Mateo, L. J. et al. Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
106. Liu, M. et al. Multiplexed imaging of nucleome architectures in single cells of mammalian tissue. *Nat. Commun.* **11**, 2907 (2020).
107. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl Acad. Sci. USA* **116**, 19490–19499 (2019).
108. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
109. Shah, S. et al. Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* **174**, 363–376.e16 (2018).
110. Shah, S., Lubeck, E., Zhou, W. & Cai, L. seqFISH accurately detects transcripts in single cells and reveals robust spatial organization in the hippocampus. *Neuron* **94**, 752–758.e1 (2017).
111. Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* **182**, 1641–1659.e26 (2020).
This paper shows combined RNA, chromatin and nuclear body imaging in the same single cells by applying RNA-MERFISH and DNA-MERFISH for >1,000 nascent transcripts and genomic loci together with fluorescent antibody readouts for cell-cycle-state markers and nuclear structures.
112. Takei, Y. et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
This paper shows combined RNA, chromatin and nuclear body imaging in the same single cells by applying DNA-seqFISH+ to thousands of genomic loci with RNA-seqFISH of 70 mRNAs and nascent transcriptionally active sites of 1,000 genes, together with antibody targeting of 17 nuclear structures.
113. Nguyen, H. Q. et al. 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* **17**, 822–832 (2020).
114. Zhu, Z. et al. Genome profiles of pathologist-defined cell clusters by multiregional LCM and G&T-seq in one triple-negative breast cancer patient. *Cell Rep. Med.* **2**, 100404 (2021).
115. 10x Genomics. Spatial proteogenomics. *10x Genomics* https://www.10xgenomics.com/products/spatial-proteogenomics (2023).
116. Ben-Chetrit, N. et al. Integrated protein and transcriptome high-throughput spatial profiling. Preprint at *bioRxiv* https://doi.org/10.1101/2022.03.15.484516 (2022).
117. Vickovic, S. et al. SM-Omics is an automated platform for high-throughput spatial multi-omics. *Nat. Commun.* **13**, 795 (2022).
118. NanoString. GeoMx DSP overview. *NanoString* https://nanostring.com/products/geomx-digital-spatial-profiler/geomx-dsp-overview/ (2023).

119. Merritt, C. R. et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.* **38**, 586–599 (2020).
120. Wang, F. et al. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).
121. Bonnett A. S. et al. The development and performance of a GeoMx® based proteogenomic workflow for the detection of both RNA and protein on a single FFPE slide (nanostring.com, 2022).
122. Fan, R. et al. Spatial-CITE-seq: spatially resolved high-plex protein and whole transcriptome co-mapping. *Res. Sq.* https://doi.org/10.21203/rs.3.rs-1499315/v1 (2022). **The spatial-CITE-seq method shows the highest level of spatial proteomics multiplexing in combination with unbiased transcriptome profiling to date using a microfluidics-based deterministic barcoding-in-tissue approach.**
123. Wang, G., Moffitt, J. R. & Zhuang, X. Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci. Rep.* **8**, 4847 (2018).
124. Kohman, R. E. & Church, G. M. Fluorescent in situ sequencing of DNA barcoded antibodies. Preprint at *bioRxiv* https://doi.org/10.1101/2020.04.27.060624 (2020).
125. Alon, S. et al. Expansion sequencing: spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, eaax2656 (2021).
126. He, S. et al. High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat. Biotechnol.* **40**, 1794–1806 (2022).
127. Liu, J. et al. Concordance of MERFISH spatial transcriptomics with bulk and single-cell RNA sequencing. *Life Sci. Alliance* **6**, e202201701 (2022).
128. Janesick, A. et al. High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue. Preprint at *bioRxiv* https://doi.org/10.1101/2022.10.06.510405 (2022).
129. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
130. Argelaguet, R., Cuomo, A. S. E., Stegle, O. & Marioni, J. C. Computational principles and challenges in single-cell data integration. *Nat. Biotechnol.* **39**, 1202–1215 (2021).
131. Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
132. Rautenstrauch, P., Vlot, A. H. C., Saran, S. & Ohler, U. Intricacies of single-cell multi-omics data integration. *Trends Genet.* **38**, 128–139 (2022).
133. Efremova, M. & Teichmann, S. A. Computational methods for single-cell omics across modalities. *Nat. Methods* **17**, 14–17 (2020).
134. Packer, J. & Trapnell, C. Single-cell multi-omics: an engine for new quantitative models of gene regulation. *Trends Genet. TIG* **34**, 653–665 (2018).
135. Vallejos, C. A., Marioni, J. C. & Richardson, S. BASiCS: Bayesian analysis of single-cell sequencing data. *PLoS Comput. Biol.* **11**, e1004333 (2015).
136. Cuomo, A. S. E. et al. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* **11**, 810 (2020).
137. Gasperini, M. et al. A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 377–390.e19 (2019).
138. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
139. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
140. Welch, J. D. et al. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177**, 1873–1887.e17 (2019).
141. Liu, J. et al. Jointly defining cell types from multiple single-cell datasets using LIGER. *Nat. Protoc.* **15**, 3632–3662 (2020).
142. Argelaguet, R. et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 111 (2020).
143. Lotfollahi, M., Litinetskaya, A. & Theis, F. J. Multigrate: single-cell multi-omic data integration. Preprint a *bioRxiv* https://doi.org/10.1101/2022.03.16.484643 (2022).
144. Yang, K. D. et al. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. *Nat. Commun.* **12**, 31 (2021).
145. Wu, K. E., Yost, K. E., Chang, H. Y. & Zou, J. BABEL enables cross-modality translation between multiomic profiles at single-cell resolution. *Proc. Natl Acad. Sci. USA* **118**, e2023070118 (2021).
146. Martinez-de-Morentin, X. et al. Adaptative machine translation between paired single-cell multi-omics data. Preprint at *bioRxiv* https://doi.org/10.1101/2021.01.27.428400 (2022).
147. Stanojevic, S., Li, Y., Ristivojevic, A. & Garmire, L. X. Computational methods for single-cell multi-omics integration and alignment. *Genomics Proteomics Bioinformatics* https://doi.org/10.1016/j.gpb.2022.11.013 (2022).
148. Zeira, R., Land, M., Strzalkowski, A. & Raphael, B. J. Alignment and integration of spatial transcriptomics data. *Nat. Methods* **19**, 567–575 (2022).
149. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
150. Elosua-Bayes, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Res.* **49**, e50 (2021).
151. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
152. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
153. Foster, D. S. et al. Integrated spatial multiomics reveals fibroblast fate during tissue repair. *Proc. Natl Acad. Sci. USA* **118**, e2110025118 (2021).
154. Ravi, V. M. et al. Spatially resolved multi-omics deciphers bidirectional tumor-host interdependence in glioblastoma. *Cancer Cell* **40**, 639–655.e13 (2022).
155. Velten, B. et al. Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO. *Nat. Methods* **19**, 179–186 (2022).
156. Bergenstråhle, L. et al. Super-resolved spatial transcriptomics by deep data fusion. *Nat. Biotechnol.* **40**, 476–479 (2022).
157. Bao, F. et al. Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. *Nat. Biotechnol.* **40**, 1200–1209 (2022).
158. Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods* **16**, 43–49 (2019).
159. Lee, M. & Pavlovic, V. Private-shared disentangled multimodal VAE for learning of hybrid latent representations. Preprint at https://doi.org/10.48550/arXiv.2012.13024 (2020).
160. Bredikhin, D., Kats, I. & Stegle, O. MUON: multimodal omics analysis framework. *Genome Biol.* **23**, 42 (2022).
161. Palla, G. et al. Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).
162. Wiener, D. & Schwartz, S. The epitranscriptome beyond m^6A. *Nat. Rev. Genet.* **22**, 119–131 (2021).
163. Chen, W. et al. Live-seq enables temporal transcriptomic recording of single cells. *Nature* **608**, 733–740 (2022).
164. Preissl, S., Gaulton, K. J. & Ren, B. Characterizing *cis*-regulatory elements using single-cell epigenomics. *Nat. Rev. Genet.* **24**, 21–43 (2022).
165. Moffitt, J. R., Lundberg, E. & Heyn, H. The emerging landscape of spatial profiling technologies. *Nat. Rev. Genet.* **23**, 741–759 (2022).
166. Bouwman, B. A. M., Crosetto, N. & Bienko, M. The era of 3D and spatial genomics. *Trends Genet.* **38**, 1062–1075 (2022).
167. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019).
168. Taylor, M. J., Lukowski, J. K. & Anderton, C. R. Spatially resolved mass spectrometry at the single cell: recent innovations in proteomics and metabolomics. *J. Am. Soc. Mass. Spectrom.* **32**, 872–894 (2021).
169. Zawistowski, J. S. et al. Unifying genomics and transcriptomics in single cells with ResolveOME amplification chemistry to illuminate oncogenic and drug resistance mechanisms. Preprint at *bioRxiv* https://doi.org/10.1101/2022.04.29.489440 (2022).
170. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
171. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–546 (2022).
172. Liao, J., Lu, X., Shao, X., Zhu, L. & Fan, X. Uncovering an organ's molecular architecture at single-cell resolution by spatially resolved transcriptomics. *Trends Biotechnol.* **39**, 43–58 (2021).
173. Asp, M., Bergenstråhle, J. & Lundeberg, J. Spatially resolved transcriptomes — next generation tools for tissue exploration. *BioEssays* **42**, 1900221 (2020).
174. Xu, J. & Liu, Y. A guide to visualizing the spatial epigenome with super-resolution microscopy. *FEBS J.* **286**, 3095–3109 (2019).
175. Ganesh, S. et al. Spatially resolved 3D metabolomic profiling in tissues. *Sci. Adv.* **7**, eabd0957 (2021).
176. Salmén, F. et al. Barcoded solid-phase RNA capture for spatial transcriptomics profiling in mammalian tissue sections. *Nat. Protoc.* **13**, 2501–2534 (2018).
177. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
178. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
179. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).
180. Wählby, C., Erlandsson, F., Bengtsson, E. & Zetterberg, A. Sequential immunofluorescence staining and image analysis for detection of large numbers of antigens in individual cell nuclei. *Cytometry* **47**, 32–41 (2002).
181. Lin, J.-R., Fallahi-Sichani, M., Chen, J.-Y. & Sorger, P. K. Cyclic Immunofluorescence (CycIF), a highly multiplexed method for single-cell imaging. *Curr. Protoc. Chem. Biol.* **8**, 251–264 (2016).
182. Cattoretti, G., Bosisio, F. M., Marcelis, L. & Bolognesi, M. M. Multiple iterative labeling by antibody neodeposition (MILAN). *Protoc. Exch.* https://doi.org/10.1038/protex.2018.106 (2018).
183. Black, S. et al. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).
184. Saka, S. K. et al. Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* **37**, 1080–1090 (2019).
185. Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
186. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
187. Lee, J. H. et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).
188. Lee, J. H. et al. Highly multiplexed subcellular RNA sequencing in situ. *Science* **343**, 1360–1363 (2014).
189. Zhu, X., Xu, T., Peng, C. & Wu, S. Advances in MALDI mass spectrometry imaging single cell and tissues. *Front. Chem.* **9**, 782432 (2022).

Acknowledgements
T.V., K.V., A.S. and B.T. are supported by Aligning Science Across Parkinson's (ASAP-000430) through the Michael J. Fox Foundation for Parkinson's Research (MJFF). T.V., K.V. and A.S. are supported by a grant from The Leona M. and Harry B. Helmsley Charitable Trust to KU Leuven (1903-03785) and by KU Leuven (IDN/19/039; IDN/21/006). T.V. and K.V. are supported by KU Leuven (C14/18/092 and C14/22/125), the Research Foundation Flanders (FWO: G081318N, G0C6120N, G088621N and I001818N), the Flemish government (iBOF/21/043), the Chan Zuckerberg Foundation (CZF, CZF2019-002436, 6000511-5500001380) and Opening the Future (OtF). B.T. is supported by FWO (G0C7519N and S003422N), the Foundation against Cancer (F/2020/1544) and KU Leuven (METH/21/06). A.S. is supported by KU Leuven (C24M/19/075) and FWO (G005923N).

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
T.V. is co-inventor on licensed patents WO/2011/157846 (Methods for haplotyping single cells), WO/2014/053664 (High-throughput genotyping by sequencing low amounts of genetic material) and WO/2015/028576 (Haplotyping and copy number typing using polymorphic variant allelic frequencies).

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41576-023-00580-2.

Peer review information *Nature Reviews Genetics* thanks Martin Enge, Kun Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Journal of Medicinal Chemistry

pubs.acs.org/jmc

Editorial

A New Era of Artificial Intelligence (AI): Transforming Drug Discovery and Development

Cite This: J. Med. Chem. 2025, 68, 23643−23652

Read Online

ACCESS | Metrics & More | Article Recommendations

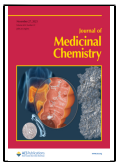
Artificial Intelligence (AI) has emerged as a transformative technology in contemporary drug discovery and development, fundamentally reshaping the processes by which novel drugs are identified, optimized, and translated into clinical applications. The inclusion of AI into drug discovery and development thus holds tremendous potential to change the landscape of therapeutics for various human diseases, enabling unprecedented precision, efficiency, and innovation. Drug development is a complex process, which encompasses numerous critical stages such as target identification, drug discovery, preclinical development, clinical trials, drug approval, and postmarket monitoring. Traditional drug discovery is an expensive, lengthy, and complex process with the low success rate.[1,2] It is estimated that the average cost of developing a novel therapeutic agent has risen to about US $2.6 billion, typically with a development timeline of 12 to 15 years.[3] Moreover, the overall success rate of novel drugs remains below 10%, even at the stage of clinical trials. By leveraging advanced computational techniques, including machine learning (ML), deep learning (DL), and predictive analytics, AI accelerates drug discovery and development process through automated data analysis, molecular modeling, target identification, virtual screening, lead optimization and clinical development. The applications of AI-based methods across all stages of drug discovery and development pipeline, from target identification to postmarket monitoring, are outlined in Figure 1.

Conventional approaches, including whole-genome knockdown screening and affinity pull-down assays, are typically employed for target identification. However, these methods are often constrained by extended timelines, substantial labor requirements, and relatively high failure rates. AI has shown great potential in identifying disease-related molecular patterns and causal relationships by generating multiomics data, thereby accelerating the drug target discovery.[4−8] Graph deep learning combines graph structures with DL to efficiently identify novel drug targets. Moreover, convergence of scientific and medical literature with multiomics data into knowledge graphs help AI to predict relationship between disease pathways and genes.[9−11] Integrating biomedical large language models (LLMs) with knowledge graphs or biological network provides precise and efficient means to link biological processes, diseases, and genes.[12] For instance, the PandaOmics platform identified TRAF2- and NCK-interacting kinase as a promising target for the treatment of fibrosis through multiomics data and biological networks analysis that culminated in the development of **INS018-055** (**14**) (Figure 2 and Table 1), a specific inhibitor of TRAF2- and NCK-interacting kinase.[13]

Virtual screening (VS) is a robust and promising computer-aided drug discovery (CADD) approach, which analyzes a vast chemical library to identify compounds with a high probability of interacting with a specific target. VS is broadly divided into two categories: (1) structure-based virtual screening (SBVS), which predicts the interaction between large libraries of compounds and a protein target using the 3D structure of the target along with detailed knowledge of its binding site, and (2) ligand-based virtual screening (LBVS), which is typically employed when information about the target structure is unavailable. Quantitative structure−activity relationship (QSAR), pharmacophore modeling and similarity matching are commonly employed in LBVS. AI-based ligand−receptor docking models are capable of predicting ligand spatial conformations and directly creating complex atomic coordinates through algorithms such as equivariant neural networks.[14,15] Furthermore, these powerful models can learn the probability density distributions of ligand−receptor distances, allowing them to form, rearrange, and refine binding poses with improved accuracy.[16,17] Recent DL-based ligand−receptor cofolding models, including AlphaFold3 and RoseTTAFold, have shown great potential for predicting complex structures directly from sequence information.[18,19]

AI-based models have tremendous potential in generating novel chemical architectures with optimally anticipated molecular properties, thereby expediting drug discovery and development. Traditional approaches to drug discovery such as structure-based, ligand-based and pharmacophore designs rely on existing chemical libraries, expert knowledge, and predefined rules. In contrast, AI, especially DL, enables the automated generation of new molecular entities optimized for specific biological activities.[20] For instance, reinforce learning (RL) is a powerful method that iteratively improves compounds design based on feedbacks from each iteration's results in meeting specific criteria such as stability, binding affinity, and pose.[21,22] Generative models such as variational autoencoders (VAEs) and generative adversarial networks

Published: November 27, 2025



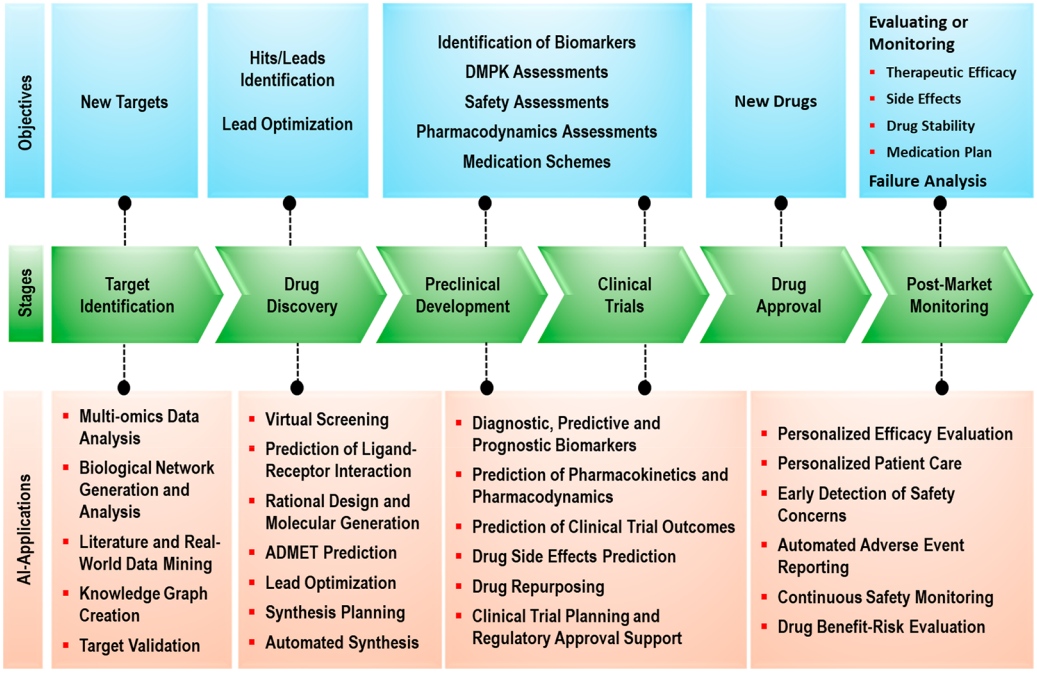


Figure 1. Applications of AI in distinctive stages of drug discovery and development. The drug development process comprises of many crucial stages such as target identification, drug discovery, preclinical development, clinical trials, drug approval by regulatory authorities, and postmarket monitoring. AI-based approaches hold promise in facilitating all stages of drug development.

(GANs) are capable of generating new compounds based on existing data.[23] Although VAEs and GANs are valuable methods in creating novel chemical space, these are accompanied by several shortcomings, including that newly generated compounds may not be synthetically viable, stable or biologically active.

Absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties play a pivotal role in determining the safety and efficacy of a drug candidate. Inferior ADMET profiles are responsible for the failure of many clinical trials. Early stage ADMET assessment could help alleviate the risk of failure due to poor properties, thereby enhancing the likelihood of successful drug development.[24] Traditional wet-lab methods for ADMET evaluation have several shortcomings such as extended timelines and high costs, which impede drug development. To circumvent these limitations and accelerate the process, AI holds great potential for predicting ADMET properties employing descriptors or molecular fingerprints. The *in silico* ADMET platform at Bayer employs ML tools, including random forest (RF) and support vector machines (SVM), which utilize descriptors such as extended connectivity fingerprints.[25] DL-driven ADMET prediction methods allow automatically extracting feature representations from simple input data. It is reported that deep neural networks (DNNs) demonstrate either equal or a little higher effectiveness than the traditional ML models when trained on ADMET data sets.[26−28] In addition, GNNs of DL model have emerged as the powerful tools for predicting ADMET properties. For instance, GeoGNN model has shown super performance in prediction of ADMET using geometric information.[29]

AI is also transforming the landscape of clinical trials by comprehensively analyzing patient data such as clinical history, genetic profiles, and lifestyle factors. Analysis of such data using AI tools may lead to discovery of biomarkers and characteristics of patients, thereby rendering clinical trials more informative and efficient. AI can increase the success of clinical trials and expedite the translation of drugs into medical use by optimizing treatment regimens, patient selection, and result measurements.

AI models may facilitate the discovery of diagnostic biomarkers, providing predictive insights and valuable guidance to support clinical diagnosis.[30] For instance, the 'nuclei.io' digital pathology framework[31] helps pathologists in diagnosing by integrating AI, active learning, and real-time feedback, significantly enhancing the efficiency and accuracy of diagnostics. Further, AI has demonstrated its potential in identifying prognostic biomarkers, opening avenues for personalized and targeted therapies. For example, the DL models can describe the morphology of $CD8^{+}T$ in blood samples as effective prognostic indicators of sepsis.[32] Moreover, the DL models hold potential to identify proteomic biomarkers, predicting liver disease consequences accurately.[33]

AI-based approaches play an instrumental role in refining therapeutic windows, improving the safety profiles of drugs, and adjusting dose−response relationships, thereby addressing key pharmacometrics challenges in precision treatments. For example, a study involving 442 small molecule kinases and 2,145 adverse events (AEs) used a ML-based model to identify novel kinase-AEs associations, providing valuable insights to reduce side effects and guide the development of small molecule kinase inhibitors with improved safety profiles.[34] The PharmBERT model can automatically extract the critical pharmacokinetic information from prescription labels, facilitating the identification of adverse drug reactions and drug−drug interactions.[35]

AI technologies have expedited the development of effective medications for various human diseases by repurposing existing drugs through the analysis of large-scale medical data sets. For example, AI has been effectively harnessed to repurpose existing drugs for treating coronavirus infection,

https://doi.org/10.1021/acs.jmedchem.5c03159
J. Med. Chem. 2025, 68, 23643−23652

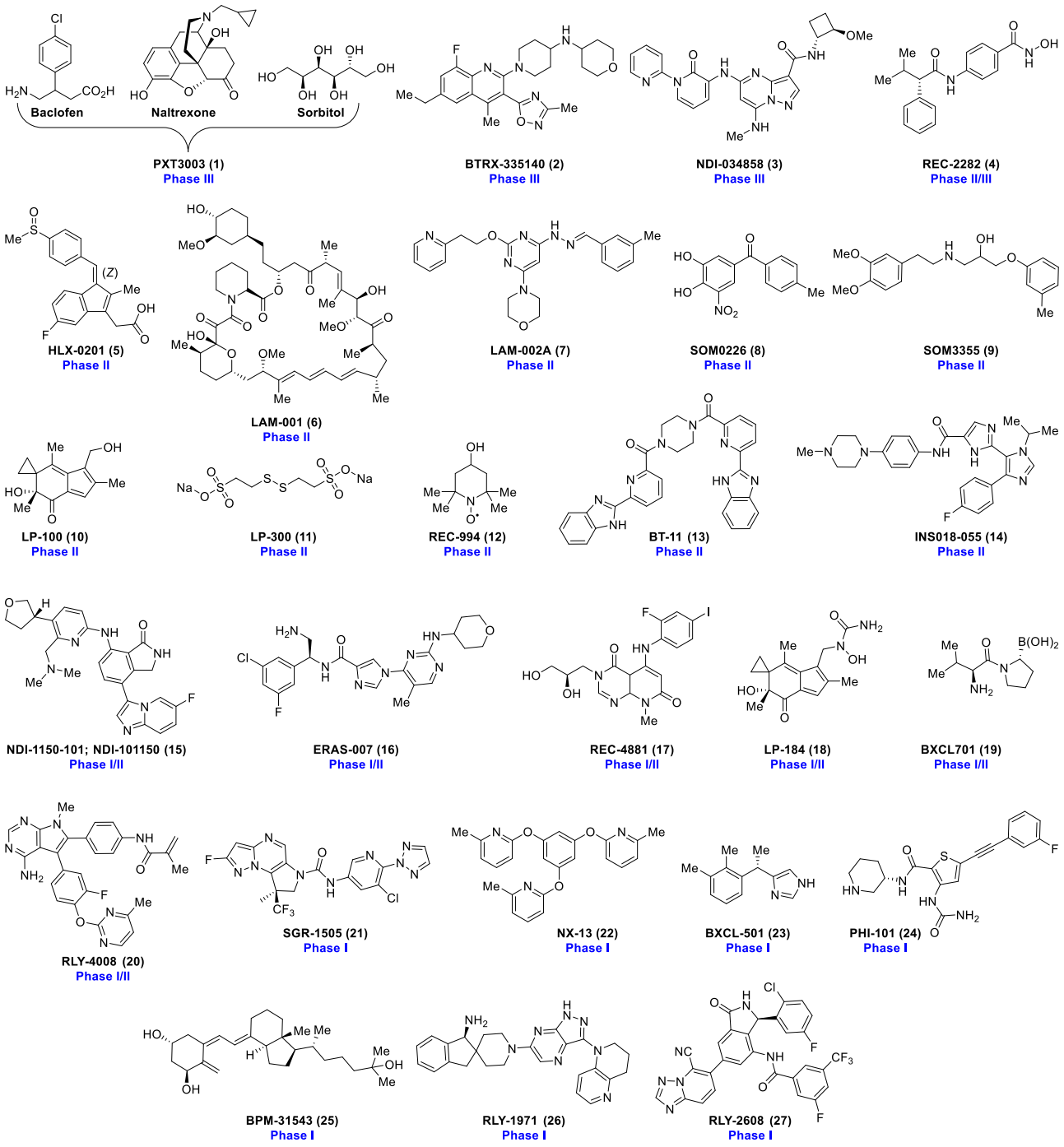


Figure 2. Chemical structures of representative AI-discovered clinical drug candidates in distinct phases of human clinical trials.

underscoring its critical role in discovering new therapeutic applications for approved treatments.[36] Additionally, AI-driven drug repurposing can be facilitated through real-world data such as electronic health record and insurance claims. For instance, a DL-based model has been utilized on a large cohort comprising millions of individuals with coronary artery disease to find drugs and drug combinations that significantly ameliorate the disease outcomes.[37]

The postmarket monitoring involves the reporting of AEs of drugs, with the aim of ensuring and supporting their continued efficacy, safety and quality. The regulatory bodies obtain, identify, assess, and process AEs from across the world.[38] Consequently, these agencies encounter a significant challenge in monitoring the safety of their medicines. Given the excellent ability of AI technology to analyze large data sets, it may help regulatory agencies streamline the processing of AEs. The Center for Drug Evaluation and Research (CDER) of the FDA is utilizing AI-based approaches in handling and assessing Individual Case Safety Reports (ICSRs) submitted to the FDA Adverse Event Reporting System (FAERS).[39] For instance, the ML models have been developed by office of surveillance and epidemiology (OSE) of CDER to automatically classify AE reports via supervised ML and text engineering approaches.[40] In addition, the Information Visualization Platform (InfoViP) has been created, featuring AL/ML (NLP) and advanced visualization capabilities. Its objective is to analyze reported

https://doi.org/10.1021/acs.jmedchem.5c03159
J. Med. Chem. 2025, 68, 23643−23652

Table 1. An Overview of Representative AI-Discovered Drugs in Clinical Trials[a]

Drug	Sponsor	Indication	Phase	NCT Identifier
PXT3003 (1)	Pharnext	Charcot−Marie−Tooth type 1A	III	NCT03023540
BTRX-335140 (2)	Neumora	Major depressive disorder	III	NCT06058013
NDI-034858 (3)	Takeda	Plaque psoriasis	III	NCT06973291
REC-2282 (4)	Recursion	Neurofibromatosis type 2	II/III	NCT05130866
HLX-0201 (5)	Healx	Male fragile X syndrome	II	NCT04823052
LAM-001 (6)	Steven Hays	Bronchiolitis obliterans syndrome	II	NCT06018766
LAM-002A (7)	OrphAI Therapeutics	Amyotrophic lateral sclerosis	II	NCT05163886
SOM0226 (8)	SOM Innovation Biotech	Transthyretin amyloidosis	II	NCT02191826
SOM3355 (9)	SOM Innovation Biotech	Chorea in Huntington's disease	II	NCT05475483
LP-100 (10)	Allarity	mCRPC	II	NCT03643107
LP-300 (11)	Lantern Pharma	Advanced lung adenocarcinoma	II	NCT05456256
REC-994 (12)	Recursion	Cerebral cavernous malformation	II	NCT05085561
BT-11 (13)	NImmune	Ulcerative colitis	II	NCT03861143
INS018-055 (14)	InSilico Medicine	Idiopathic pulmonary fibrosis	II	NCT05975983
NDI-1150-101 (15)	Nimbus	Solid tumor	I/II	NCT05128487
ERAS-007 (16)	Erasca	Advanced gastrointestinal malignancies	I/II	NCT05039177
REC-4881 (17)	Recursion	Familial adenomatous polyposis	I/II	NCT05552755
LP-184 (18)	Lantern	Solid tumor	I/II	NCT05933265
BXCL701 (19)	BioXcel Therapeutics	mCRPC	I/II	NCT03910660
RLY-4008 (20)	Elevar Therapeutics	Solid tumor	I/II	NCT04526106
SGR-1505 (21)	Schrödinger	Relapsed/refractory B-cell lymphomas	I	NCT05544019
NX-13 (22)	Landos Biopharma	Ulcerative colitis	I	NCT04862741
BXCL-501 (23)	BioXcel Therapeutics	Agitation associated with pediatric schizophrenia and bipolar disorder	I	NCT05025605
PHI-101 (24)	Seoul National University Hospital	Acute myelogenous leukemia	I	NCT04842370
BPM-31543 (25)	BPGbio	Alopecia	I	NCT01588522
RLY-1971 (26)	Hoffmann-La Roche	Metastatic solid tumor	I	NCT04252339
RLY-2608 (27)	Relay Therapeutics	Advanced solid tumor or breast cancer	I	NCT05216432

[a]Data collected from https://www.clinicaltrials.gov. (Accessed on October 25, 2025).

AEs by identifying duplicate ICSRs, classifying ICSRs based on label information, and visualizing the timeline of clinical events.[41]

To date, a significant number of AI-discovered drugs are currently being evaluated in distinct phases of human clinical trials, highlighting the growing importance of AI in drug discovery and development. The chemical structures of some representative drugs with their development facilitated by AI technologies are shown in Figure 2, and an overview of their clinical trials is provided in Table 1. PXT3003 (1) was discovered by AI-assisted repurposing of three already approved drugs: baclofen, a muscle relaxant; naltrexone, a therapy for opioid dependence; and sorbitol, a glucose-lowering agent utilized as a laxative.[42] PXT3003 (1) is being evaluated in phase III to assess its long-term tolerability and safety in patients with Charcot−Marie−Tooth type 1A (NCT03023540). BTRX-335140 (2), a.k.a. navacaprant, is a selective small molecule κ opioid receptor (KOR) antagonist (IC_{50} = 0.8 nM).[43] It was discovered by BlackThorn Therapeutics in a collaboration with the Scripps Research Institute. The company has employed its computational psychiatry platform, which integrates bioinformatics, neuro-informatics, and AI/ML technologies, to determine each patient's distinct neuroprint, thereby enabling the selection of optimal, on-target treatment options for biologically defined patient subtypes.[44] BTRX-335140 (2) is under study in phase III to examine its effects against placebo in patients with major depressive disorder (NCT06058013). NDI-034858 (3), a.k.a. zasocitinib or TAK-279, is a potent and highly selective tyrosine kinase 2 (TYK2) inhibitor, displaying remarkable selectivity for the human TYK2 JH2 (K_d = 0.0038 nM) over TYK1 JH2 (K_d = 4,975 nM) and TYK3 JH2 (K_d = 23,000 nM).[45] An AI-assisted SBVS strategy has been utilized to design molecules for enhanced selectivity and efficacy, ultimately leading to the identification of NDI-034858 (3).[46] It is under investigation in phase III to test its efficacy against deucravacitinib in adult patients with plaque psoriasis (NCT06973291). REC-2282 (4) is pan-HDAC inhibitor, discovered by Recursion Pharmaceuticals using its AI/ML platform.[47] It is being evaluated in phase II/III to determine its safety and efficacy in patients with progressive neurofibromatosis type 2 (NCT05130866). HLX-0201 (5), previously known as sulindac, is an approved nonsteroidal anti-inflammatory drug (NSAID). HLX-0201 (5) has been identified through the repurposing of sulindac using a novel omic-based drug-matching approach developed by Healx. This method holds potential for discovering new therapeutic connections and disease pathways by comparing the gene expression profile of a disease with gene expression profiles from Healx's curated drug database.[48] It is being evaluated in phase II to assess its safety, efficacy, and tolerability in patients with fragile X syndrome (NCT04823052). LAM-001 (6), a.k.a. sirolimus or rapamycin, was first approved in 1999 for the prevention of kidney transplant rejection. AI Therapeutics has repurposed sirolimus or rapamycin utilizing its AI-platform, which leverages DL tools, to identify its new therapeutic use for the treatment of pulmonary arterial hypertension (PAH).[49] LAM-001 (6) is designed for direct supply to the lungs, thereby mitigating systematic exposure and minimizing the toxicities typically associated with oral sirolimus. LAM-001 (6) is under investigation in phase II to test its safety and efficacy in patients with bronchiolitis obliterans syndrome

https://doi.org/10.1021/acs.jmedchem.5c03159
J. Med. Chem. 2025, 68, 23643−23652

(NCT06018766). AI Therapeutics has identified LAM-002A (**7**) as the potential treatment for amyotrophic lateral sclerosis (ALS) by the repurposing of apilimod using its AI-platform, Guardian Angel.[50] LAM-002A (**7**) is in phase II studies to evaluate its safety, efficacy, and tolerability in patients with C9ORF72-associted ALS (NCT05163886). SOM0226 (**8**) has been identified by SOM Biotech through the repurposing of an approved drug (tolcapone). Using its AI-based VS platform, the company has discovered a new therapeutic application of tolcapone as a promising treatment for the transthyretin amyloidosis.[51] SOM0226 (**8**) is being tested in phase II to assess its efficacy for the treatment of transthyretin amyloidosis (NCT02191826). SOM3355 (**9**), a.k.a. bevantolol, is a potent vesicular monoamine transporter type 2 (VMAT2) inhibitor that has been used for the treatment of hypertension for years. Using its AI-based SOM^{AI}PRO technology, SOM Biotech has repurposed bevantolol and identified a new clinical use as a potential therapy for chorea in Huntington's disease (HD).[52] The phase II studies of SOM3355 (**9**) are being conducted to test its safety and efficacy of two doses in patients with chorea in HD (NCT05475483). LP-100 (**10**), a.k.a. irofulven, has been discovered by Lantern Pharma using its AL/ML-based platform, RADR, and the NCI's CellMiner cross database.[53] Through this approach, the company has uncovered biological insights and potential new targets indications for irofulven. LP-100 (**10**) is being evaluated in a phase II study to assess its antitumor efficacy in combination with prednisolone in patients with metastatic castration-resistant prostate cancers (mCRPC) who have been previously treated with androgen receptor-targeted therapy, and docetaxel (NCT03643107). LP-300 (**11**) is a small molecule developed using the AI-driven platform, RADR, designed for the treatment of patients suffering from nonsmall cell lung cancer (NSCLC) who have never smoked, and their cancer has grown following treatment with tyrosine kinase inhibitors (TKIs).[54] The RADR is assisting the HARMONIC trial, evaluating LP-300 (**11**) in combination with pemetrexed and carboplatin in nonsmokers with advanced lung adenocarcinoma for whom treatment with TKIs is ineffective (NCT05456256).[54] REC-994 (**12**) is a small molecule, for the treatment of cerebral cavernous malformation (CCM), discovered employing Recursion's AI-based drug discovery platform, which leverages basic ML tools.[55] A phase II study is being conducted to assess the safety, efficacy, and pharmacokinetics (PK) of REC-994 (**12**) in patients with symptomatic CCM, compared to placebo (NCT05085561). BT-11 (**13**), a.k.a. omilancor, is a first-in-class, orally available, gut-restricted activator of lanthionine synthetase C-like 2 (LANCL2) with therapeutic potential for treating inflammatory bowel disease (IBD).[56] It was discovered using Landos Biopharma's advanced modeling and AI-based LANCE precision medicine platform, which helps in identifying immunometabolic targets and designing new molecules that modulate these conserved pathways, delivering therapeutic effects.[57] BT-11 (**13**) is under investigation in phase II to examine its safety and efficacy in patients with ulcerative colitis (UC) (NCT03861143). INS018-055 (rentosertib) (**14**) is a first-in-class, small molecule TNIK (TRAF2 and NCK-interacting protein kinase) inhibitor for the treatment of idiopathic pulmonary fibrosis (IPF), identified by generative AI.[58] It is being investigated in phase II to test its safety and efficacy in patients with IPF compared to placebo (NCT05975983). NDI-1150-101 or NDI-101150 (**15**) is a potent and selective inhibitor of hematopoietic progenitor cell kinase 1 (HPK1). It was discovered by Nimbus Therapeutics using computational tools integrated with ML-based predictive modeling approaches.[59] A phase I/II study is being carried out to evaluate the safety, efficacy, PK and pharmacodynamic (PD) of NDI-101150 as monotherapy or in combination with pembrolizumab in patients with advanced solid tumors (NCT05128487). ERAS-007 (**16**), a.k.a. ASN007, is a potent and selective inhibitor of extracellular signal-regulated kinases ERK1 and ERK2 (ERK1/2), with an IC_{50} value of 2 nM.[60] It was discovered by Erasca, a biotechnology company that employes AI-based platforms to facilitate drug discovery and development. ERAS-007 (**16**) is being investigated in phase I/II to examine its safety and tolerability at higher doses in combination with other cancer treatments in patients with advanced gastrointestinal malignancies (NCT05039177). REC-4881 (**17**) is a non-ATP-competitive small molecule mitogen-activated protein kinase 1 and 2 (MAPK1 and MAPK2) inhibitor, being developed for the treatment of familial adenomatous polyposis (FAP). It was identified by Recursion utilizing its AI/ML-based drug discovery platform, the Recursion Operating System, which maps relationship between biological targets and small molecules.[61] REC-4881 (**17**) is being evaluated in phase I/II to test its safety, efficacy, PK and PD in patients with FAP (NCT05552755). LP-184 (**18**) is a synthetic small molecule, belonging to a class of naturally occurring anticancer compounds. It exerts its therapeutic effects by preferentially destroying DNA in cancer cells. Lantern is using its AI platform, RADR, to accelerate the development of LP-184 (**18**) by discovering its mechanisms of actions for the treatment of challenging cancers and providing valuable insights into specific populations of patients.[62] LP-184 (**18**) is being examined in phase I//II to assess its safety, efficacy, and maximum tolerable dose in patients suffering from advanced solid tumors (NCT05933265). BXCL-701 (**19**) is a nonselective inhibitor targeting dipeptidyl peptidase 8 and 9 (DPP8/9) as well as fibroblast activation protein (FAP). It was earlier investigated in phase II and III clinical trials; however, it failed to demonstrate sufficient efficacy in its intended therapeutic indication. BXCL701 (**19**) has been repurposed by BioXcel Therapeutics using its AI-based platform. Through this approach, the company has identified that BXCL701 (**19**) may act synergistically with checkpoint inhibitors, potentially providing therapeutic advantages for patients who do not respond to checkpoint inhibitor monotherapy.[63] A phase II study of BXCL701 (**19**) is being conducted to assess its safety and efficacy as monotherapy and in combination with pembrolizumab in patients with mCRPC either small cell neuroendocrine prostate cancer or adenocarcinoma phenotype (NCT03910660). RLY-4008 (**20**), a.k.a. lirafugratinib, is a highly selective small molecule fibroblast growth factor receptor 2 (FGFR2) inhibitor.[64] It was discovered by Relay Therapeutics employing its AI-based platform.[65] RLY-4008 (**20**) is being evaluated in phase I/II to assess its safety, tolerability, PK, PD, and antineoplastic activity in patients with metastatic cholangiocarcinoma (CCA) and other solid tumors (NCT04526106). SGR-1505 (**21**) is a potent and orally available small molecule inhibitor of mucosa-associated lymphoid tissue lymphoma translocation protein 1 (MALT1). It was identified by the Schrödinger using integrated advanced physics-based modeling methods, free energy calculations with ML tools and chemistry-aware compound enumeration workflow.[66] SGR-1505 (**21**) is under study in phase I to assess its safety and tolerability in

https://doi.org/10.1021/acs.jmedchem.5c03159
J. Med. Chem. 2025, 68, 23643−23652

patients with relapsed/refractory B-cell lymphomas. NX-13 (**22**) is an orally active, gut-restricted, novel small molecule inhibitor of nucleotide-binding oligomerization domain, leucine rich repeat containing X1 (NLRX1) for the treatment of UC.[67] Landos Biopharma identified NLRX1 using its advanced AI-based integrated computational and experimental precision medicine platform, led to the discovery of NX-13 (**22**).[68] A phase I study of NX-13 (**22**) is being carried out to assess its safety, tolerability, and PD in patients with active UC (NCT04862741). BXCL-501 (**23**) was discovered through the repurposing of dexmedetomidine, an approved drug, by BioXcel Therapeutics using AI-based technologies.[69] BXCL-501 (**23**) is being evaluated in phase I to assess its safety and efficacy in treating agitation associated with pediatric schizophrenia and bipolar disorder (NCT05025605). PHI-101 (**24**) is an orally available, next-generation FLT3 tyrosine kinase inhibitor, with a novel mechanism of action designed to circumvent resistance to acute myeloid leukemia (AML).[70] It was discovered and developed by the Pharos iBio, using its AI-based platform, Chemiverse, which is capable to deal with all aspects of the drug discovery process, from novel drug target identification to clinical candidate selection.[71] PHI-101 (**24**) is being investigated in phase I to examine its safety, tolerability, PK, and PD in patients with relapsed or refractory AML (NCT04842370). BPM-31543 (**25**) was identified by BERG, a biopharmaceutical company, using its AI-powered, Interrogative Biology platform for the treatment of chemotherapy-induced alopecia (CIA).[72] It is understudy in phase I to assess its safety in patients with CIA (NCT01588522). RLY-1971 (**26**) is a small molecule protein tyrosine phosphatase SHP2 inhibitor, being developed for the treatment of solid tumors.[73] It was discovered by Relay Therapeutics using its AI-driven platform, Dynamo, which analyzes protein dynamic to identify disease-relevant conformational states.[74] RLY-1971 (**26**) is under investigation in phase I to determine its safety, efficacy, tolerability, and PK in patients with advanced or metastatic solid tumors (NCT04252339). RLY-2608 (**27**) is first-in-class, mutant-selective, allosteric inhibitor of PI3Kα, that dissociates antitumor activity from hyperinsulinemia.[75] It was discovered by Relay Therapeutics using its AI-driven platform and targets a novel allosteric pocket distinct from the primary active site of the protein that is preferentially favored in certain mutant forms of PI3Kα.[76] RLY-2608 (**27**) is being evaluated in phase I to test its safety, tolerability, PK and PD as monotherapy and in combination with endocrine therapy with or without a CDK4/6 or CDK4 inhibitor in patients with advanced solid tumors or breast cancer (NCT05216432).

Despite the immense potential of AI technologies, as exemplified by the milestone successes of the aforementioned drug candidates in human clinical trials, it is accompanied by numerous significant limitations to fully revolutionize drug discovery and development. The dearth of high-quality data required for training renders AI less effective in discovery of novel targets, biomarkers, and other related applications. In addition, the presence of incomplete information, biases, and errors in available data or data with potential ethical concerns further attenuate the reliability of AI. The lack of published negative data, including failed experiments and trials with unfavorable results, impedes the insights into efficacy, drug−target−disease relationships, and broader clinical characteristics.[77,78] Current generative approaches hold considerable promise for generating novel chemical architectures; many of these structures are likely to be synthetically unfeasible or unstable. While molecular generation methods complemented by detailed reaction knowledge have demonstrated significant potential, they still require further improvement.[79,80] The traditionally undruggable targets such as transcription factors or scaffolding proteins with flexible, dynamic structures or lacking well-defined binding sites pose significant challenges for AI-based drug discovery methods. These hurdles may be addressed by integrating AI tools with high-content screening to search the conformational space and discover suitable binding sites. In addition, many AI platforms or systems, often referred to as "black boxes," lack transparency in their decision-making process, raising a potential barrier to regulatory and clinical trust. Moreover, the FDA is still developing clear guidelines for AI-driven drug development, creating uncertainty for AI-assisted pharmaceuticals. The unique characteristics of AI continue to pose challenges for both drug developers and regulators.

While AI is facilitating drug discovery and development process, future efforts are required to address existing challenges to fully realize the potential of AI approaches and become standard practice. Viable strategies should be devised to determine data standards, increase data sharing, and generate new AI algorithms capable of making accurate predictions from limited data. Multimodal utilizing chemical and textual information has shown potential in solving data lacking concern.[81] AI can uncover the therapeutic potential of existing drugs for treating orphan diseases through repurposing by integrating data, including protein−protein interactions, multiomics profiles, disease-specific molecular pathways, and clinical records.[82] Multimodal such as DL-based classification of drugs, can identify underlying mechanism of actions, predict efficacy, and evaluate toxicity, demonstrating the future potential of AI-based approaches in drug development.[83,84] AI-based models are anticipated to change the landscape of medical modeling and simulation. AI models will enable the generation of comprehensive virtual human simulations,[85] providing valuable insights into disease mechanisms,[86] drug actions, and biological diversity among individuals.[87] AI can improve clinical trial design by simulating multiple scenarios to optimize selection criteria,[88] expedite patient recruitment, and ensure more diverse, representative populations.[89]

In conclusion, AI holds immense potential in addressing the challenges associated with traditional drug discovery and development by accelerating timelines, reducing costs, and increasing the probability of success. Advanced AI methods are capable of analyzing large data sets for novel target identification, augmenting the efficiency of VS, performing lead optimization, and facilitating preclinical development. Moreover, AI plays a pivotal role in early clinical trials by assisting in patient recruitment and predicting outcomes, thereby attenuating the likelihood of trial failure. Further, AI can distinguish between simple prognostic biomarkers and those predictive responses to medication, thereby advancing the development of personalized treatments. The power of AI can be harnessed, if applied judiciously, across all stages of the drug discovery and development pipeline, ultimately enabling the discovery of novel and precision medicines for treatment of various human diseases.

Saghir Ali orcid.org/0000-0002-7400-5512
Xiaochen Tian
Haiying Chen
Jia Zhou orcid.org/0000-0002-2811-1090

https://doi.org/10.1021/acs.jmedchem.5c03159
J. Med. Chem. 2025, 68, 23643−23652

AUTHOR INFORMATION

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jmedchem.5c03159

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.
The authors declare no competing financial interests.

ACKNOWLEDGMENTS

This work was partly supported by grants R01 DA038446, R01 DA040621, and R01 DA060228 from the National Institutes of Health, the John D. Stobo, M.D. Distinguished Chair Endowment Fund (J.Z.), the Edith & Robert Zinn Chair in Drug Discovery Endowment Fund (J.Z.).

REFERENCES

(1) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **2010**, *9*, 203−214.
(2) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **2019**, *18*, 463−477.
(3) Zhang, K.; Yang, X.; Wang, Y.; Yu, Y.; Huang, N.; Li, G.; Li, X.; Wu, J. C.; Yang, S. Artificial intelligence in drug development. *Nat. Med.* **2025**, *31*, 45−59.
(4) Chen, H.; King, F. J.; Zhou, B.; Wang, Y.; Canedy, C. J.; Hayashi, J.; Zhong, Y.; Chang, M. W.; Pache, L.; Wong, J. L.; Jia, Y.; Joslin, J.; Jiang, T.; Benner, C.; Chanda, S. K.; Zhou, Y. Drug target prediction through deep learning functional representation of gene signatures. *Nat. Commun.* **2024**, *15*, 1853.
(5) Tasaki, S.; Xu, J.; Avey, D. R.; Johnson, L.; Petyuk, V. A.; Dawe, R. J.; Bennett, D. A.; Wang, Y.; Gaiteri, C. Inferring protein expression changes from mRNA in Alzheimer's dementia using deep neural networks. *Nat. Commun.* **2022**, *13*, 655.
(6) Rodriguez, S.; Hug, C.; Todorov, P.; Moret, N.; Boswell, S. A.; Evans, K.; Zhou, G.; Johnson, N. T.; Hyman, B. T.; Sorger, P. K.; Albers, M. W.; Sokolov, A. Machine learning identifies candidates for drug repurposing in Alzheimer's disease. *Nat. Commun.* **2021**, *12*, 1033.
(7) Hong, C.; Cao, Q.; Zhang, Z.; Tsui, S. K.-W.; Yip, K. Y. Reusability report: Capturing properties of biological objects and their relationships using graph neural networks. *Nat. Mach. Intell.* **2022**, *4*, 222−226.
(8) Ratajczak, F.; Joblin, M.; Hildebrandt, M.; Ringsquandl, M.; Falter-Braun, P.; Heinig, M. Speos: an ensemble graph representation learning framework to predict core gene candidates for complex diseases. *Nat. Commun.* **2023**, *14*, 7206.
(9) Youn, J.; Rai, N.; Tagkopoulos, I. Knowledge integration and decision support for accelerated discovery of antibiotic resistance genes. *Nat. Commun.* **2022**, *13*, 2360.
(10) Ge, Y.; Tian, T.; Huang, S.; Wan, F.; Li, J.; Li, S.; Wang, X.; Yang, H.; Hong, L.; Wu, N.; Yuan, E.; Luo, Y.; Cheng, L.; Hu, C.; Lei, Y.; Shu, H.; Feng, X.; Jiang, Z.; Wu, Y.; Chi, Y.; Guo, X.; Cui, L.; Xiao, L.; Li, Z.; Yang, C.; Miao, Z.; Chen, L.; Li, H.; Zeng, H.; Zhao, D.; Zhu, F.; Shen, X.; Zeng, J. An integrative drug repositioning framework discovered a potential therapeutic agent targeting COVID-19. *Signal Transduct. Target. Ther.* **2021**, *6*, 165.
(11) Gogleva, A.; Polychronopoulos, D.; Pfeifer, M.; Poroshin, V.; Ughetto, M.; Martin, M. J.; Thorpe, H.; Bornot, A.; Smith, P. D.; Sidders, B.; Dry, J. R.; Ahdesmaki, M.; McDermott, U.; Papa, E.; Bulusu, K. C. Knowledge graph-based recommendation framework identifies drivers of resistance in EGFR mutant non-small cell lung cancer. *Nat. Commun.* **2022**, *13*, 1667.
(12) Luo, R.; Sun, L.; Xia, Y.; Qin, T.; Zhang, S.; Poon, H.; Liu, T. Y. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Brief. Bioinform.* **2022**, *23*, bbac409.
(13) Ren, F.; Aliper, A.; Chen, J.; Zhao, H.; Rao, S.; Kuppe, C.; Ozerov, I. V.; Zhang, M.; Witte, K.; Kruse, C.; Aladinskiy, V.; Ivanenkov, Y.; Polykovskiy, D.; Fu, Y.; Babin, E.; Qiao, J.; Liang, X.; Mou, Z.; Wang, H.; Pun, F. W.; Torres-Ayuso, P.; Veviorskiy, A.; Song, D.; Liu, S.; Zhang, B.; Naumov, V.; Ding, X.; Kukharenko, A.; Izumchenko, E.; Zhavoronkov, A. A small-molecule TNIK inhibitor targets fibrosis in preclinical and clinical models. *Nat. Biotechnol.* **2025**, *43*, 63−75.
(14) Zhang, X.; Zhang, O.; Shen, C.; Qu, W.; Chen, S.; Cao, H.; Kang, Y.; Wang, Z.; Wang, E.; Zhang, J.; Deng, Y.; Liu, F.; Wang, T.; Du, H.; Wang, L.; Pan, P.; Chen, G.; Hsieh, C. Y.; Hou, T. Efficient and accurate large library ligand docking with KarmaDock. *Nat. Comput. Sci.* **2023**, *3*, 789−804.
(15) Qiao, Z.; Nie, W.; Vahdat, A.; Miller, T. F., III; Anandkumar, A. State-specific protein-ligand complex structure prediction with a multiscale deep generative model. *Nat. Mach. Intell.* **2024**, *6*, 195−208.
(16) Peng, Q.; Lei, Y.; Feng, P.; Jia, L.; Ma, J.; Zhao, D.; Zeng, J. Characterizing the interaction conformation between T-cell receptors and epitopes with deep learning. *Nat. Mach. Intell.* **2023**, *5*, 395−407.
(17) Mendez-Lucio, O.; Ahmad, M.; del Rio-Chanona, E. A.; Wegner, J. K. A geometric deep learning approach to predict binding conformations of bioactive molecules. *Nat. Mach. Intell.* **2021**, *3*, 1033−1039.
(18) Krishna, R.; Wang, J.; Ahern, W.; Sturmfels, P.; Venkatesh, P.; Kalvet, I.; Lee, G. R.; Morey-Burrows, F. S.; Anishchenko, I.; Humphreys, I. R.; McHugh, R.; Vafeados, D.; Li, X.; Sutherland, G. A.; Hitchcock, A.; Hunter, C. N.; Kang, A.; Brackenbrough, E.; Bera, A. K.; Baek, M.; DiMaio, F.; Baker, D. Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **2024**, *384*, No. eadl2528.
(19) Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A. J.; Bambrick, J.; Bodenstein, S. W.; Evans, D. A.; Hung, C. C.; O'Neill, M.; Reiman, D.; Tunyasuvunakool, K.; Wu, Z.; Zemgulyte, A.; Arvaniti, E.; Beattie, C.; Bertolli, O.; Bridgland, A.; Cherepanov, A.; Congreve, M.; Cowen-Rivers, A. I.; Cowie, A.; Figurnov, M.; Fuchs, F. B.; Gladman, H.; Jain, R.; Khan, Y. A.; Low, C. M. R.; Perlin, K.; Potapenko, A.; Savy, P.; Singh, S.; Stecula, A.; Thillaisundaram, A.; Tong, C.; Yakneen, S.; Zhong, E. D.; Zielinski, M.; Zidek, A.; Bapst, V.; Kohli, P.; Jaderberg, M.; Hassabis, D.; Jumper, J. M. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493−500.
(20) Schneider, G. Automating drug discovery. *Nat. Rev. Drug Discov.* **2018**, *17*, 97−113.
(21) Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **2018**, *4*, No. eaap7885.
(22) Fang, Y.; Pan, X.; Shen, H. B. De novo drug design by iterative multiobjective deep reinforcement learning with graph-based molecular quality assessment. *Bioinformatics* **2023**39.btad157.
(23) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360−365.
(24) Swanson, K.; Walther, P.; Leitz, J.; Mukherjee, S.; Wu, J. C.; Shivnaraine, R. V.; Zou, J. ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. *Bioinformatics* **2024**, *40*, btae416.
(25) Goller, A. H.; Kuhnke, L.; Montanari, F.; Bonin, A.; Schneckener, S.; Ter Laak, A.; Wichard, J.; Lobell, M.; Hillisch, A. Bayer's in silico ADMET platform: a journey of machine learning over the past two decades. *Drug Discov. Today* **2020**, *25*, 1702−1709.
(26) Postel-Vinay, S.; Lam, V. K.; Ros, W.; Bauer, T. M.; Hansen, A. R.; Cho, D. C.; Stephen Hodi, F.; Schellens, J. H. M.; Litton, J. K.; Aspeslagh, S.; Autio, K. A.; Opdam, F. L.; McKean, M.; Somaiah, N.; Champiat, S.; Altan, M.; Spreafico, A.; Rahma, O.; Paul, E. M.; Ahlers, C. M.; Zhou, H.; Struemper, H.; Gorman, S. A.; Watmuff, M.;

Yablonski, K. M.; Yanamandra, N.; Chisamore, M. J.; Schmidt, E. V.; Hoos, A.; Marabelle, A.; Weber, J. S.; Heymach, J. V. First-in-human phase I study of the OX40 agonist GSK3174998 with or without pembrolizumab in patients with selected advanced solid tumors (ENGAGE-1). *J. Immunother. Cancer* **2023**, *11*, No. e005301.
(27) Piperno-Neumann, S.; Carlino, M. S.; Boni, V.; Loirat, D.; Speetjens, F. M.; Park, J. J.; Calvo, E.; Carvajal, R. D.; Nyakas, M.; Gonzalez-Maffe, J.; Zhu, X.; Shirley, M. D.; Ramkumar, T.; Fessehatsion, A.; Burks, H. E.; Yerramilli-Rao, P.; Kapiteijn, E. A phase I trial of LXS196, a protein kinase C (PKC) inhibitor, for metastatic uveal melanoma. *Br. J. Cancer* **2023**, *128*, 1040−1051.
(28) Ortega-Paz, L.; Giordano, S.; Capodanno, D.; Mehran, R.; Gibson, C. M.; Angiolillo, D. J. Clinical Pharmacokinetics and Pharmacodynamics of CSL112. *Clin. Pharmacokinet.* **2023**, *62*, 541−558.
(29) Fang, X.; Liu, L.; Lei, J.; He, D.; Zhang, S.; Zhou, J.; Wang, F.; Wu, H.; Wang, H. Geometry-enhanced molecular representation learning for property prediction. *Nat. Mach. Intell.* **2022**, *4*, 127−134.
(30) Claudio Quiros, A.; Coudray, N.; Yeaton, A.; Yang, X.; Liu, B.; Le, H.; Chiriboga, L.; Karimkhan, A.; Narula, N.; Moore, D. A.; Park, C. Y.; Pass, H.; Moreira, A. L.; Le Quesne, J.; Tsirigos, A.; Yuan, K. Mapping the landscape of histomorphological cancer phenotypes using self-supervised learning on unannotated pathology slides. *Nat. Commun.* **2024**, *15*, 4596.
(31) Huang, Z.; Yang, E.; Shen, J.; Gratzinger, D.; Eyerer, F.; Liang, B.; Nirschl, J.; Bingham, D.; Dussaq, A. M.; Kunder, C.; Rojansky, R.; Gilbert, A.; Chang-Graham, A. L.; Howitt, B. E.; Liu, Y.; Ryan, E. E.; Tenney, T. B.; Zhang, X.; Folkins, A.; Fox, E. J.; Montine, K. S.; Montine, T. J.; Zou, J. A pathologist-AI collaboration framework for enhancing diagnostic accuracies and efficiencies. *Nat. Biomed. Eng.* **2025**, *9*, 455−470.
(32) Sung, M.; Kim, J. H.; Min, H. S.; Jang, S.; Hong, J.; Choi, B. K.; Shin, J.; Chung, K. S.; Park, Y. R. Three-dimensional label-free morphology of CD8 + T cells as a sepsis biomarker. *Light Sci. Appl.* **2023**, *12*, 265.
(33) Heckenbach, I.; Mkrtchyan, G. V.; Ezra, M. B.; Bakula, D.; Madsen, J. S.; Nielsen, M. H.; Oro, D.; Osborne, B.; Covarrubias, A. J.; Idda, M. L.; Gorospe, M.; Mortensen, L.; Verdin, E.; Westendorp, R.; Scheibye-Knudsen, M. Nuclear morphology is a deep learning biomarker of cellular senescence. *Nat. Aging* **2022**, *2*, 742−755.
(34) Gong, X.; Hu, M.; Liu, J.; Kim, G.; Xu, J.; McKee, A.; Palmby, T.; de Claro, R. A.; Zhao, L. Decoding kinase-adverse event associations for small molecule kinase inhibitors. *Nat. Commun.* **2022**, *13*, 4349.
(35) ValizadehAslani, T.; Shi, Y.; Ren, P.; Wang, J.; Zhang, Y.; Hu, M.; Zhao, L.; Liang, H. PharmBERT: a domain-specific BERT model for drug labels. *Brief. Bioinform.* **2023**, *24*, bbad226.
(36) Zhou, Y.; Wang, F.; Tang, J.; Nussinov, R.; Cheng, F. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit. Health* **2020**, *2*, e667−e676.
(37) Liu, R.; Wei, L.; Zhang, P. A deep learning framework for drug repurposing via emulating clinical trials on real-world patient data. *Nat. Mach. Intell.* **2021**, *3*, 68−75.
(38) Niazi, S. K. The Coming of Age of AI/ML in Drug Discovery, Development, Clinical Testing, and Manufacturing: The FDA Perspectives. *Drug Des. Devel. Ther.* **2023**, *17*, 2691−2725.
(39) Ball, R.; Dal Pan, G. "Artificial Intelligence" for Pharmacovigilance: Ready for Prime Time? *Drug Saf.* **2022**, *45*, 429−438.
(40) Kreimeyer, K.; Dang, O.; Spiker, J.; Munoz, M. A.; Rosner, G.; Ball, R.; Botsis, T. Feature engineering and machine learning for causality assessment in pharmacovigilance: Lessons learned from application to the FDA Adverse Event Reporting System. *Comput. Biol. Med.* **2021**, *135*, 104517.
(41) Office of Surveillance and Epidemiology 2023 Annual Report. Available at: https://www.fda.gov/media/180347/download. (Accessed on October 10, 2025).
(42) How AI Is Finding New Cures in Old Drugs. Availabe at: https://fortune.com/longform/ai-drugs-pharma-pharnext-cmt/. (Accessed on October 16, 2025).
(43) Guerrero, M.; Urbano, M.; Kim, E. K.; Gamo, A. M.; Riley, S.; Abgaryan, L.; Leaf, N.; Van Orden, L. J.; Brown, S. J.; Xie, J. Y.; Porreca, F.; Cameron, M. D.; Rosen, H.; Roberts, E. Design and Synthesis of a Novel and Selective Kappa Opioid Receptor (KOR) Antagonist (BTRX-335140). *J. Med. Chem.* **2019**, *62*, 1761−1780.
(44) BlackThorn Therapeutics Reports Data Supporting Phase 2 Development of KOR Antagonist, BTRX-335140, in Depression and Other CNS Disorders. Available at: https://www.biospace.com/blackthorn-therapeutics-reports-data-supporting-phase-2-development-of-kor-antagonist-btrx-335140-in-depression-and-other-cns-disorders. (Accessed on October 17, 2025).
(45) Leit, S.; Greenwood, J.; Carriero, S.; Mondal, S.; Abel, R.; Ashwell, M.; Blanchette, H.; Boyles, N. A.; Cartwright, M.; Collis, A.; Feng, S.; Ghanakota, P.; Harriman, G. C.; Hosagrahara, V.; Kaila, N.; Kapeller, R.; Rafi, S. B.; Romero, D. L.; Tarantino, P. M.; Timaniya, J.; Toms, A. V.; Wester, R. T.; Westlin, W.; Srivastava, B.; Miao, W.; Tummino, P.; McElwee, J. J.; Edmondson, S. D.; Masse, C. E. Discovery of a Potent and Selective Tyrosine Kinase 2 Inhibitor: TAK-279. *J. Med. Chem.* **2023**, *66*, 10473−10496.
(46) Zasocitinib: AI-Designed TYK2 Inhibitor Shows Promise in Psoriasis and Psoriatic Arthritis Treatment. Availabe at: https://trial.medpath.com/news/46255192148953e1/ai-powered-innovation-zasocitinib-s-journey-to-precision-inhibition-dermatology-times. (Accessed on October 17, 2025).
(47) Artificial Intelligence Lights a Beacon to New Medicine for Neurofibromatosis type 2. Available at: https://medium.com/recursion-pharmaceuticals/artificial-intelligence-provides-a-beacon-to-new-medicine-for-neurofibromatosis-type-2-d0ec41344f50. (Accessed on October 17, 2025).
(48) Healx receives IND and Orphan Drug Designation for fragile X clinical trial. Availabel at: https://healx.ai/ind-fragile-x-clinical-trial/#:~:text=News-,Healx%20receives%20IND%20and%20Orphan%20Drug%20Designation%20for%20fragile%20X,8000%20females%20across%20the%20globe. (Accessed on October 20, 2025).
(49) AI Therapeutics Announces Initiation of a Phase 2 Clinical Trial of LAM-001 for Treatment of Pulmonary Arterial Hypertension (PAH). Available at: https://www.orphai-therapeutics.com/news/ai-therapeutics-announces-initiation-of-a-phase-ii-clinical-trial-of-lam-001-for-treatment-of-pulmonary-arterial-hypertension-pah-2#:~:text=%E2%80%9CBy%20modulating%20both%20the%20mTOR,%2D001%20study%20in%20PAH.%E2%80%9D. (Accessed on October 20, 2025).
(50) AI-based Platform Suggests Cancer Candidate LAM-002 as Possible ALS Therapy, AI Therapeutics Says. Available at: https://alsnewstoday.com/news/ai-based-platform-indicates-candidate-cancer-therapy-lam-002-possible-treatment-als-ai-therapeutics-announces/. (Accessed on October 20, 2025).
(51) SOM Biotech develops a drug for Amyloidosis in collaboration with the Vall d'Hebron Research Institute. Available at: https://www.pcb.ub.edu/en/som-biotech-develops-a-drug-for-amyloidosis-in-collaboration-with-the-vall-dhebron-research-institute/#:~:text=The%20new%20therapeutic%20activity%20of%20the%20SOM0226,identify%20new%20therapeutic%20activities%20of%20already%2Dmarketed%20drugs. (Accessed on October 20, 2025).
(52) SOM Biotech's Phase IIb Clinical Trial in Huntington's Disease Achieves 80% Enrolment. Available at: https://www.clinicaltrialsarena.com/contractors/software-technology/som-biotech/pressreleases/huntingtons-disease-enrolment/#:~:text=About%20SOM3355,current%20standard%2Dof%2Dcare. (Accessed on October 20, 2025).
(53) McDermott, J.; Sturtevant, D.; Kathad, U.; Varma, S.; Zhou, J.; Kulkarni, A.; Biyani, N.; Schimke, C.; Reinhold, W. C.; Elloumi, F.; Carr, P.; Pommier, Y.; Bhatia, K. Artificial intelligence platform, RADR®, aids in the discovery of DNA damaging agent for the ultra-rare cancer Atypical Teratoid Rhabdoid Tumors. *Front. Drug Discov.* **2022**, *2*, 1033395.
(54) Lung Cancer Patient in Lantern Pharma's Harmonic Trial Shows Durable Complete Response in Target Cancer Lesions with

https://doi.org/10.1021/acs.jmedchem.5c03159
J. Med. Chem. 2025, 68, 23643−23652

Survival Continuing for Nearly Two Years. Available at: https://ir.lanternpharma.com/news-1/news/news-details/2025/Lung-Cancer-Patient-in-Lantern-Pharmas-Harmonic-Trial-Shows-Durable-Complete-Response-in-Target-Cancer-Lesions-with-Survival-Continuing-for-Nearly-Two-Years/default.aspx. (Accessed on October 21, 2025).

(55) Recursion Receives FDA Clearance of Investigational New Drug Application to Initiate First Clinical Trial of REC-994 in Cerebral Cavernous Malformation. Available at: https://ir.recursion.com/news-releases/news-release-details/recursion-receives-fda-clearance-investigational-new-drug?utm_source=chatgpt.com. (Accessed on October 21, 2025).

(56) Leber, A.; Hontecillas, R.; Zoccoli-Rodriguez, V.; Colombel, J. F.; Chauhan, J.; Ehrich, M.; Farinola, N.; Bassaganya-Riera, J. The Safety, Tolerability, and Pharmacokinetics Profile of BT-11, an Oral, Gut-Restricted Lanthionine Synthetase C-Like 2 Agonist Investigational New Drug for Inflammatory Bowel Disease: A Randomized, Double-Blind, Placebo-Controlled Phase I Clinical Trial. *Inflamm. Bowel Dis.* **2019**, *26*, 643−652.

(57) Landos Biopharma Announces Positive Results from a Phase 2 Trial of Oral BT-11 for Patients with Ulcerative Colitis. Available at: https://www.biospace.com/landos-biopharma-announces-positive-results-from-a-phase-2-trial-of-oral-bt-11-for-patients-with-ulcerative-colitis?utm_source=chatgpt.com. (Accessed on October 21, 2025).

(58) Xu, Z.; Ren, F.; Wang, P.; Cao, J.; Tan, C.; Ma, D.; Zhao, L.; Dai, J.; Ding, Y.; Fang, H.; Li, H.; Liu, H.; Luo, F.; Meng, Y.; Pan, P.; Xiang, P.; Xiao, Z.; Rao, S.; Satler, C.; Liu, S.; Lv, Y.; Zhao, H.; Chen, S.; Cui, H.; Korzinkin, M.; Gennert, D.; Zhavoronkov, A. A generative AI-discovered TNIK inhibitor for idiopathic pulmonary fibrosis: a randomized phase 2a trial. *Nat. Med.* **2025**, *31*, 2602−2610.

(59) Nimbus Therapeutics Presents Positive Preliminary Data from Clinical Trial of HPK1 Inhibitor in Solid Tumors at SITC Annual Meeting. Available at: https://www.nimbustx.com/2023/10/31/nimbus-therapeutics-presents-positive-preliminary-data-from-clinical-trial-of-hpk1-inhibitor-in-solid-tumors-at-sitc-annual-meeting/#:~:text=NDI%2D101150%20was%20found%20to,benefits%20to%20patients%20with%20cancer.%E2%80%9D&text=Nimbus%20Therapeutics%20is%20a%20clinical,visit%20www.nimbustx.com. (Accessed October 21, 2025).

(60) Portelinha, A.; Thompson, S.; Smith, R. A.; Da Silva Ferreira, M.; Asgari, Z.; Knezevic, A.; Seshan, V.; de Stanchina, E.; Gupta, S.; Denis, L.; Younes, A.; Reddy, S. ASN007 is a selective ERK1/2 inhibitor with preferential activity against RAS-and RAF-mutant tumors. *Cell Rep. Med.* **2021**, *2*, 100350.

(61) Recursion Initiates Two Additional Clinical Trials For a Total of Four in 2022. Available at: https://ir.recursion.com/news-releases/news-release-details/recursion-initiates-two-additional-clinical-trials-total-four?mobile=1. (Accessed on October 22, 2025).

(62) Lantern Pharma's LP-184 Gets Second FDA Fast Track for TNBC Treatment. Available at: https://synapse.patsnap.com/article/lantern-pharmas-lp-184-gets-second-fda-fast-track-for-tnbc-treatment. (Accessed on October 22, 2025).

(63) Facilitating Drug Re-Innovation with Artificial Intelligence. Available at: https://www.pharmasalmanac.com/articles/facilitating-drug-re-innovation-with-artificial-intelligence. (Accessed on October 22, 2025).

(64) Schonherr, H.; Ayaz, P.; Taylor, A. M.; Casaletto, J. B.; Toure, B. B.; Moustakas, D. T.; Hudson, B. M.; Valverde, R.; Zhao, S.; O'Hearn, P. J.; Foster, L.; Sharon, D. A.; Garfinkle, S.; Giordanetto, F.; Lescarbeau, A.; Kurukulasuriya, R.; Gerami-Moayed, N.; Maglic, D.; Bruderek, K.; Naik, G.; Gunaydin, H.; Mader, M. M.; Boezio, A. A.; McLean, T. H.; Chen, R.; Wang, Y.; Shaw, D. E.; Watters, J.; Bergstrom, D. A. Discovery of lirafugratinib (RLY-4008), a highly selective irreversible small-molecule inhibitor of FGFR2. *Proc. Natl. Acad. Sci. U. S. A.* **2024**, *121*, No. e2317756121.

(65) Arnold, C. Inside the nascent industry of AI-designed drugs. *Nat. Med.* **2023**, *29*, 1292−1295.

(66) Nie, Z.; Trzoss, M.; Placzek, A. T.; Trzoss, L.; Krilov, G.; Feng, S.; Lawrenz, M.; Ye, M.; Marshall, N.; Dingley, K. H.; Pelletier, R. D.; Lai, W. G.; Bell, J. A.; Tang, H.; Devine, P.; Liu, Z.; Skrdla, P.; Shimanovich, R.; Liu, M.; Wang, R.; Xu, X.; Bhat, S.; Bos, P. H.; Abel, R.; Akinsanya, K.; Yin, W. Accelerated In Silico Discovery of SGR-1505: A Potent MALT1 Allosteric Inhibitor for the Treatment of Mature B-Cell Malignancies. *J. Med. Chem.* **2025**, DOI: 10.1021/acs.jmedchem.5c01494.

(67) Leber, A.; Hontecillas, R.; Zoccoli-Rodriguez, V.; Ehrich, M.; Chauhan, J.; Bassaganya-Riera, J. Exploratory studies with NX-13: oral toxicity and pharmacokinetics in rodents of an orally active, gut-restricted first-in-class therapeutic for IBD that targets NLRX1. *Drug Chem. Toxicol.* **2022**, *45*, 209−214.

(68) Landos Biopharma Reports Positive Top-Line Results From NX-13 Phase 1b Trial. Available at: https://www.biospace.com/landos-biopharma-reports-positive-top-line-results-from-nx-13-phase-1b-trial#:~:text=About%20NX%2D13,cytokines%20in%20the%20gastrointestinal%20tract. (Accessed on October 23, 2025).

(69) riple combo: calming Alzheimer's agitation with AI, wearables and a novel drug. Available at: https://www.medicaldevice-network.com/features/wearable-ai-device-for-agitation/?cf-view. (Accessed on October 23, 2025).

(70) Shin, D. Y.; Yoon, S. S.; Hong, J.; Lee, J. H.; Jang, J. H.; Cheong, J. W.; Shin, H. J.; Lee, J. K.; Lee, Y. J.; Ahn, J. S.; Cho, B. S.; Kim, H. J.; Clarey, J.; Sung, G. J.; Im, J.; Nam, K. Y.; Han, J.; Kim, K. T.; Yoon, J. H.; Nguyen, B.; Li, L.; Small, D. HI-101, a Novel FLT3 TKI, Shows Clinical Efficacy in Relapsed/Refractory FLT3-Mutated AML. *Blood* **2024**, *1441495*, 1495.

(71) The Chemiverse platform: AI and big data-based drug discovery and development. Available at: https://www.nature.com/articles/d43747-021-00036-8. (Accessed on October 23, 2025).

(72) BERG Presents Key Findings From Two Important Clinical Initiatives At The 2019 ASCO Annual Meeting: BERG's Final Phase 1 Results For BPM 31543 In Chemotherapy-Induced Alopecia And A Clinical Study. Available at: https://www.biospace.com/berg-presents-key-findings-from-two-important-clinical-initiatives-at-the-2019-asco-annual-meeting-berg-s-final-phase-1-results-for-bpm-31543-in-chemotherapy-induced-alopecia-and-a-clinical-study. (Accessed on October 23, 2025).

(73) Taylor, A. M.; Williams, B. R.; Giordanetto, F.; Kelley, E. H.; Lescarbeau, A.; Shortsleeves, K.; Tang, Y.; Walters, W. P.; Arrazate, A.; Bowman, C.; Brophy, E.; Chan, E. W.; Deshmukh, G.; Greisman, J. B.; Hunsaker, T. L.; Kipp, D. R.; Saenz Lopez-Larrocha, P.; Maddalo, D.; Martin, I. J.; Maragakis, P.; Merchant, M.; Murcko, M.; Nisonoff, H.; Nguyen, V.; Nguyen, V.; Orozco, O.; Owen, C.; Pierce, L.; Schmidt, M.; Shaw, D. E.; Smith, S.; Therrien, E.; Tran, J. C.; Watters, J.; Waters, N. J.; Wilbur, J.; Willmore, L. Identification of GDC-1971 (RLY-1971), a SHP2 Inhibitor Designed for the Treatment of Solid Tumors. *J. Med. Chem.* **2023**, *66*, 13384−13399.

(74) Relay Therapeutics pays $85M for startup with a new AI tech for drug discovery. Available at: https://medcitynews.com/2021/04/relay-therapeutics-pays-85m-for-startup-with-a-new-ai-tech-for-drug-discovery/#:~:text=Meanwhile%2C%20Relay%20is%20making%20progress,American%20Association%20for%20Cancer%20Research. (Accessed on October 23, 2025).

(75) Varkaris, A.; Pazolli, E.; Gunaydin, H.; Wang, Q.; Pierce, L.; Boezio, A. A.; Bulku, A.; DiPietro, L.; Fridrich, C.; Frost, A.; Giordanetto, F.; Hamilton, E. P.; Harris, K.; Holliday, M.; Hunter, T. L.; Iskandar, A.; Ji, Y.; Larivee, A.; LaRochelle, J. R.; Lescarbeau, A.; Llambi, F.; Lormil, B.; Mader, M. M.; Mar, B. G.; Martin, I.; McLean, T. H.; Michelsen, K.; Pechersky, Y.; Puente-Poushnejad, E.; Raynor, K.; Rogala, D.; Samadani, R.; Schram, A. M.; Shortsleeves, K.; Swaminathan, S.; Tajmir, S.; Tan, G.; Tang, Y.; Valverde, R.; Wehrenberg, B.; Wilbur, J.; Williams, B. R.; Zeng, H.; Zhang, H.; Walters, W. P.; Wolf, B. B.; Shaw, D. E.; Bergstrom, D. A.; Watters, J.; Fraser, J. S.; Fortin, P. D.; Kipp, D. R. Discovery and Clinical Proof-of-Concept of RLY-2608, a First-in-Class Mutant-Selective Allosteric

PI3Kalpha Inhibitor That Decouples Antitumor Activity from Hyperinsulinemia. *Cancer Discov.* **2024**, *14*, 240−257.
(76) AI Drug Discovery Spotlight: Mutant-Selective PI3K-alpha Inhibitors. Available at: https://www.living.tech/data-visual/mutant-selective-pi3k-alpha-inhibitors#:~:text=Relay%20Therapeutics%20searched%20for%20molecules,building%20block%20for%20combination%20therapies. (Accessed on October 24, 2025).
(77) Sadybekov, A. V.; Katritch, V. Computational approaches streamlining drug discovery. *Nature* **2023**, *616*, 673−685.
(78) Schrader, M. L.; Schafer, F. R.; Schafers, F.; Glorius, F. Bridging the information gap in organic chemical reactions. *Nat. Chem.* **2024**, *16*, 491−498.
(79) Button, A.; Merk, D.; Hiss, J. A.; Schneider, G. Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis. *Nat. Mach. Intell.* **2019**, *1*, 307−315.
(80) Qiang, B.; Zhou, Y.; Ding, Y.; Liu, N.; Song, S.; Zhang, L.; Huang, B.; Liu, Z. Bridging the gap between chemical reaction pretraining and conditional molecule generation with a unified model. *Nat. Mach. Intell.* **2023**, *5*, 1476−1485.
(81) Liu, S.; Nie, W.; Wang, C.; Lu, J.; Qiao, Z.; Liu, L.; Tang, J.; Xiao, C.; Anandkumar, A. Multi-modal molecule structure-text model for text-based retrieval and editing. *Nat. Mach. Intell.* **2023**, *5*, 1447−1457.
(82) Banerjee, J.; Taroni, J. N.; Allaway, R. J.; Prasad, D. V.; Guinney, J.; Greene, C. Machine learning in rare disease. *Nat. Methods* **2023**, *20*, 803−814.
(83) Aliper, A.; Plis, S.; Artemov, A.; Ulloa, A.; Mamoshina, P.; Zhavoronkov, A. Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. *Mol. Pharmaceutics* **2016**, *13*, 2524−2530.
(84) Artemov, A. V.; Putin, A.; Vanhaelen, Q.; Aliper, A.; Ozerov, I. V.; Zhavoronkov, A. Integrated deep learned transcriptomic and structure-based predictor of clinical trials outcomes. *bioRxiv*, **2016**, DOI: 10.1101/095653.
(85) Izard, S. G.; Juanes Mendez, J. A.; Palomera, P. R. Virtual Reality Educational Tool for Human Anatomy. *J. Med. Syst.* **2017**, *41*, 76.
(86) Vanhaelen, Q.; Lin, Y. C.; Zhavoronkov, A. The Advent of Generative Chemistry. *ACS Med. Chem. Lett.* **2020**, *11*, 1496−1505.
(87) He, J.; Baxter, S. L.; Xu, J.; Xu, J.; Zhou, X.; Zhang, K. The practical implementation of artificial intelligence technologies in medicine. *Nat. Med.* **2019**, *25*, 30−36.
(88) Chebanov, D. K.; Misyurin, V. A. Predictive Modeling of Clinical Trial Outcomes for Novel Drugs using Digital Twin Patient Cohorts and GenerativeAI. *medRvix*, **2023**, DOI: 10.1101/2023.1109.1111.23295380.
(89) Liu, R.; Rizzo, S.; Whipple, S.; Pal, N.; Pineda, A. L.; Lu, M.; Arnieri, B.; Lu, Y.; Capra, W.; Copping, R.; Zou, J. Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature* **2021**, *592*, 629−633.

https://doi.org/10.1021/acs.jmedchem.5c03159
J. Med. Chem. 2025, 68, 23643−23652

AI IN HEALTHCARE DELIVERY AT VCU HEALTH PRE-READS

Enterprise Data & AI at VCUH

Executive Summary

Artificial Intelligence (AI) and advanced analytics are increasingly embedded in healthcare delivery, operations, research, and education. At VCU Health, AI is already in use across multiple domains—from clinical decision support and imaging to operational efficiency and workforce productivity.

As adoption accelerates, the opportunity is significant, but so are the risks. AI systems depend on high-quality data, introduce new safety and ethical considerations, and require clear accountability. To realize value responsibly and at scale, VCU Health has established an **enterprise data and AI operating model** built on three pillars:

1. **A strong enterprise data foundation**
2. **Practical AI tools delivering real-world impact**
3. **A clear governance structure that enables innovation while managing risk**

This paper provides an overview of that model and prepares board members for the accompanying presentation.

---

1. Why Data & AI Matter Now

Healthcare organizations are under increasing pressure to:

- Improve patient outcomes and safety
- Reduce clinician burden and burnout
- Operate more efficiently amid workforce and financial constraints
- Enable research and innovation at scale

AI offers the ability to:

- Surface insights earlier than traditional reporting
- Embed intelligence directly into workflows
- Augment, not replace, clinical and operational decision-making

However, AI systems can also:

- Produce incorrect or biased outputs
- Create patient safety risks if poorly governed
- Increase compliance and reputational exposure
- Scale costs rapidly if unmanaged

For these reasons, **AI cannot be treated as isolated tools or pilot projects**. It must be approached as an **enterprise capability**, grounded in data and governed intentionally.

---

2. The Enterprise Data Foundation

At the core of VCU Health's AI strategy is a **centralized enterprise data platform (data lake)** that consolidates data from across the organization.

This foundation enables:

- Consistent, trusted data for analytics and AI
- Reuse of data assets across multiple use cases
- Faster development of models, dashboards, and tools
- Secure access aligned with privacy and regulatory requirements

Role of Data Governance

Data governance operates primarily at this foundational layer and focuses on:

- Data quality and standardization
- Security, privacy, and regulatory compliance
- Data stewardship and lifecycle management
- Appropriate access and use controls

Strong data governance ensures that AI systems are built on **reliable and ethically managed data**, reducing downstream risk.

---

3. AI Tools in Practice: Where Value Is Delivered

AI at VCU Health is not limited to experimentation. It is already embedded in real workflows, delivering tangible value.

Examples include:

- **Clinical AI embedded in the EHR**, such as predictive models and decision support
- **Radiology AI tools** that assist with prioritization and detection
- **Generative AI and assistant tools** that help clinicians and staff interact with information more efficiently
- **Operational and administrative AI** supporting forecasting, throughput, and workforce processes

The emphasis is on:

- Workflow integration rather than standalone tools
- Measurable impact on care delivery, efficiency, or experience
- Human oversight and accountability for AI-supported decisions

---

4. Custom AI and Agentic Capabilities

Beyond vendor-provided tools, VCU Health is developing **custom AI capabilities** enabled by the enterprise data platform.

These include:

- Predictive models tailored to VCU Health populations and workflows
- Agentic and copilot-style solutions that assist users across tasks
- Domain-specific AI for clinical, operational, and research needs

Custom development allows:

- Greater alignment with organizational priorities
- Transparency into AI behavior
- Flexibility to evolve as needs change

These efforts are deliberately governed and prioritized to avoid fragmentation or unmanaged risk.

---

5. Governance: Enabling Innovation While Managing Risk

Governance is a core enabler—not a barrier—to responsible AI adoption.

AI Workgroup (Operational Governance)

The AI Workgroup serves as the **enterprise AI governance body**, responsible for:

- Reviewing all AI tools and initiatives prior to deployment
- Evaluating risk, value, compliance, and readiness
- Ensuring alignment with Responsible AI principles
- Recommending approval, revision, pilot, or rejection

The workgroup includes representation across clinical, IT, security, privacy, compliance, research, and operational domains.

Data & AI Governance Council (Executive Oversight)

The Data & AI Governance Council provides:

- Executive oversight of data and AI strategy
- Alignment with organizational priorities and risk tolerance
- Coordination across data governance and AI governance efforts
- Escalation and decision-making for system-level investments

Together, these bodies ensure AI is:

- Transparent
- Accountable
- Ethically deployed
- Strategically aligned

6. The End-to-End Operating Model

VCU Health's approach integrates all components into a single operating model:

1. **Data Foundation**
    - Trusted, governed, enterprise data
2. **AI Development & Deployment**
    - Vendor tools and custom solutions
3. **Embedded Governance**
    - Data governance and AI governance throughout the lifecycle
4. **Ongoing Monitoring**
    - Performance, risk, and impact tracked over time

This model allows VCU Health to:

- Move faster without sacrificing safety
- Scale successful AI initiatives
- Retire or adjust tools that do not deliver value
- Maintain trust with patients, staff, and regulators

Closing Perspective

VCU Health's enterprise approach—anchored in data, focused on real-world impact, and guided by clear governance—positions the organization to capture AI's benefits while managing its risks.

The accompanying presentation will summarize this model visually and highlight key examples currently in practice.