







# Ten challenges and opportunities in computational immuno-oncology

Riyue Bao <sup>1,2</sup>, Alan Hutson,<sup>3</sup> Anant Madabhushi <sup>4,5,6</sup>, Vanessa D Jonsson,<sup>7,8</sup> Spencer R Rosario,<sup>3</sup> Jill S Barnholtz-Sloan,<sup>9,10</sup> Elana J Fertig,<sup>11,12,13</sup> Himangi Marathe,<sup>3</sup> Lyndsay Harris,<sup>14</sup> Jennifer Altreuter,<sup>15</sup> Qingrong Chen,<sup>16</sup> James Dignam,<sup>17</sup> Andrew J Gentles,<sup>18</sup> Edgar Gonzalez-Kozlova,<sup>19,20,21</sup> Sacha Gnjjatic,<sup>19,20,21</sup> Erika Kim,<sup>22</sup> Mark Long <sup>3</sup>, Martin Morgan,<sup>3</sup> Eytan Rupp <sup>23</sup>, David Van Valen,<sup>24,25</sup> Hong Zhang,<sup>3</sup> Natalie Vokes <sup>26,27</sup>, Daoud Meerzaman,<sup>16</sup> Song Liu,<sup>3</sup> Eliezer M Van Allen <sup>28,29,30,31</sup>, Yi Xing<sup>32,33,34</sup>

**To cite:** Bao R, Hutson A, Madabhushi A, *et al.* Ten challenges and opportunities in computational immuno-oncology. *Journal for ImmunoTherapy of Cancer* 2024;**12**:e009721. doi:10.1136/jitc-2024-009721

RB, AH, AM, VD, J, SRR, JSB-S and EJJ contributed equally.

Accepted 23 September 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Alan Hutson;  
Alan.Hutson@roswellpark.org

Dr Riyue Bao; baor@upmc.edu

Dr Anant Madabhushi;  
anantm@emory.edu

Dr Daoud Meerzaman;  
meerzam@mail.nih.gov

Dr Song Liu;  
song.liu@roswellpark.org

Dr Eliezer M Van Allen;  
eliezerm\_vanallen@dfci.harvard.edu

Dr Yi Xing; xingyi@chop.edu

## ABSTRACT

Immuno-oncology has transformed the treatment of cancer, with several immunotherapies becoming the standard treatment across histologies. Despite these advancements, the majority of patients do not experience durable clinical benefits, highlighting the imperative for ongoing advancement in immuno-oncology. Computational immuno-oncology emerges as a forefront discipline that draws on biomedical data science and intersects with oncology, immunology, and clinical research, with the overarching goal to accelerate the development of effective and safe immuno-oncology treatments from the laboratory to the clinic. In this review, we outline 10 critical challenges and opportunities in computational immuno-oncology, emphasizing the importance of robust computational strategies and interdisciplinary collaborations amid the constantly evolving interplay between clinical needs and technological innovation.

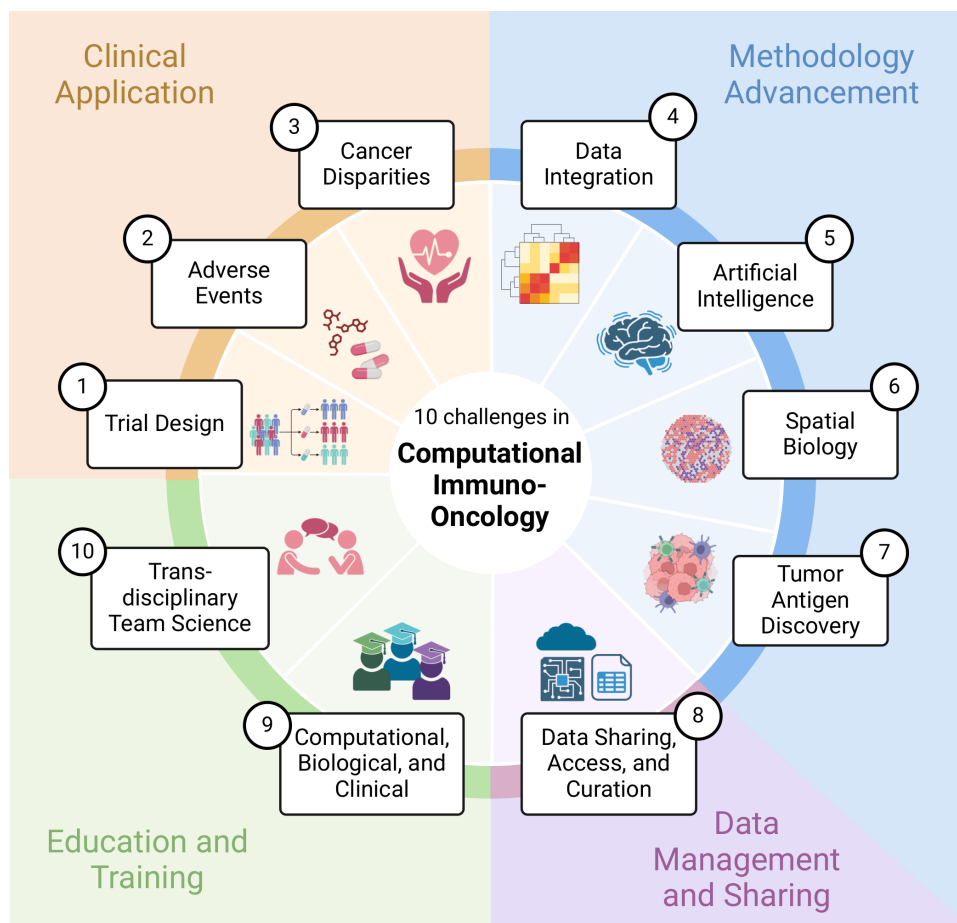
## INTRODUCTION

Immunotherapies have revolutionized treatment for patients with cancer,<sup>1</sup> with immuno-oncology (IO) emerging as one of the most significant advancements in recent years.<sup>2</sup> Unlike traditional treatments that target cancer cells directly, IO harnesses the innate and adaptive immune systems of the host to eliminate tumors. Core examples of IO include immune checkpoint inhibitors<sup>3</sup> (ICI), chimeric antigen receptor (CAR) T-cell therapy,<sup>4</sup> tumor-infiltrating lymphocyte (TIL) therapy,<sup>5</sup> and cancer vaccines.<sup>6</sup> However, despite these advancements, the majority of patients do not experience durable clinical benefits, underscoring the need for improved therapeutic strategies.

With recent advances in high-dimensional technologies and artificial intelligence (AI), new opportunities have emerged in leveraging data science to optimize clinical trial design, predict therapy response,<sup>7</sup> decipher

resistance mechanisms,<sup>8</sup> and establish new biomarkers.<sup>9</sup> Computational IO<sup>10</sup> (compIO) is an innovative, transdisciplinary field that stems from biomedical data science and interfaces with oncology, immunology, and clinical research, with the overarching goal of accelerating the move of effective and safe IO treatments from the laboratory to the clinic. To fully harness its potential, the field must embrace novel, data-driven strategies that systematically interrogate IO responses across data modalities, patient demographics, and socioeconomic backgrounds. Appropriately designed studies, standardized data protocols, and scalable analytical approaches are all vital to advancing IO therapies and to improving patient outcomes.

In this review, we describe 10 main challenges and opportunities in compIO (figure 1), highlighting the importance of robust computational strategies and interdisciplinary collaborations. We begin by discussing practical clinical applications such as trial design and predictive modeling of immune-related adverse events (irAEs), as well as strategies for addressing cancer disparities in under-represented groups. We then delve into methodological innovations in data integration and AI, examining the real-world challenges and pathways for implementing these strategies in clinical practice. Furthermore, we emphasize the critical role of establishing standardized data collection and analysis protocols, with ethical considerations in data-sharing practices. Lastly, we call for a unified effort among computational scientists, biologists, bioengineers, and medical researchers including oncologists, pathologists,



**Figure 1** Overview of 10 challenges and opportunities in computational immuno-oncology.

radiologists, and surgeons, to catalyze a paradigm shift in patient care.

## CHALLENGES AND OPPORTUNITIES

### Clinical trial design

The design of IO trials is critical to determining the best strategy for assessing the safety and efficacy of new therapies in patients. Given the complexity of interactions between the immune system and cancer, advanced computational and analytical methodologies have become indispensable in accelerating drug development and optimizing trial parameters, from preclinical studies to phase IV post-marketing studies.

The optimization of the dose and timing of candidate IO drugs is a crucial task in trial design. Quantitative systems pharmacology models, along with pharmacokinetic (PK) and pharmacodynamic and spatial-temporal models, are used to simulate and predict patient responses.<sup>11–12</sup> These methods are also increasingly used in the preclinical phase and to inform regulatory interactions in clinical research.<sup>13–15</sup> To optimize efficacy and minimize severe irAEs in trials, computational algorithms have been employed in trial design to simulate scenarios across combinations of different treatment modalities,<sup>16,17</sup> where the proper assessment of immunogenicity endpoints is increasingly recognized as crucial.<sup>18–19</sup> Classical phase I

dose-escalation trials, which generally assume that the highest dose meeting safety criteria will consequently be the most effective, and that a lower dose is inherently safer than a higher dose, have proven inadequate in the IO drug setting due to the non-linear relationship between dose and efficacy or toxicity.<sup>20</sup> New models that account for and continually reassess the safety of both long-term and short-term irAEs are essential to identify the safest and most effective dose. These models are multivariate in nature, incorporating measures of immune response and tumor activity.<sup>20</sup>

Cost-effective and expeditious phase II and phase I/II trial designs have become a priority, in part due to the need for receiving US Food and Drug Administration (FDA) rapid approval.<sup>21–23</sup> An optimal statistical design that employs early stopping points for efficacy and futility,<sup>24</sup> and that aims to obtain response profiles under both frequentist and Bayesian frameworks, is desirable.<sup>25–26</sup> Adaptive randomization schemes, which can reweigh treatment allocations, are now being employed in the phase II setting for IO.<sup>26</sup> Computational tumor dynamic models are also increasingly being applied to support trial readouts, go/no-go decision-making, and regulatory interactions during early clinical developments.<sup>27–28</sup> For large-scale phase III randomized trials, the design must consider delays in effects relative to

progression-free survival (PFS) and overall survival (OS). Modeling delayed PFS and OS beyond the median, or employing time-varying treatment effect estimates, may be appropriate when modeling IO drug outcomes.<sup>29</sup> Precision medicine approaches, such as large basket trials that leverage immune profile testing and other companion diagnostics to assign agents to subjects, are increasingly important.<sup>30</sup> This has been demonstrated in the iMATCH trial, which uses biological mechanisms of resistance to categorize tumors,<sup>31 32</sup> and the MyPathway phase IIa trial, which uses tumor mutational burden (TMB) as a predictor for the response to atezolizumab.<sup>33</sup>

A highly impactful component of phase II and III trials is the correlative studies that are crucial for elucidating the biological mechanisms driving treatment responses and resistance.<sup>34 35</sup> These studies, supported by comprehensive data repositories like the Genomic Data Commons,<sup>36</sup> Imaging Data Commons,<sup>37</sup> and the National Clinical Trials Network (NCTN) and the NCI Community Oncology Research Program (NCORP) Data Archive,<sup>38</sup> contribute to a robust understanding of the tumor-immune microenvironment. While profiling can be performed in all of these trials, neoadjuvant designs provide surgical specimens that are most amenable to high-throughput profiling studies to discover cellular drivers and pathways of therapeutic response in patients. Profiling studies are accompanied by the development of computational and mathematical algorithms, as well as multifaceted bioinformatics tools, to inform drug repurposing and development of future novel combination trials.<sup>39</sup> The National Cancer Institute (NCI) provides sources of clinical trial biospecimens and annotated clinical data that researchers may request through the NCTN Navigator resource<sup>40</sup> for novel assay development and biomarker discovery. In addition, the NCI Informatics Technology for Cancer Research program (ITCR) supports new methodology and software developments to enable wide adoption of these new methods throughout the IO research community.<sup>41</sup>

Throughout the entire drug development pipeline, from preclinical studies to phase III/IV trials, computational methodologies play a central role in accelerating medicines to market.<sup>42</sup> Early phase target discoveries in IO require accurate modeling of the complex interactions among tumor, immune, and stroma cells. High-dimensional data integration from genomics, proteomics, and transcriptomics<sup>43</sup> can identify targets however suffer from data heterogeneity and a lack of standardized protocols for target validation. Molecular docking<sup>44</sup> has been used to predict binding affinities and interaction sites for compound selection, however, in silico models often struggle to predict off-target effects and the overall PK profile of lead compounds, which can lead to failures in later stages of development. Similarly, AI-driven drug repurposing<sup>45 46</sup> approaches are limited by the dynamics of cell–cell interactions, which are often not well-characterized in existing drug databases. Combining virtual chemical screening with machine learning (ML)

techniques can better capture the complexity of ligand-receptor interactions and enhance prediction.<sup>42</sup>

### Adverse events

Patients undergoing IO treatment are vulnerable to developing irAE because the treatment itself can compromise the immune system, making it harder for the body to fight off infections or inflammation. Common irAEs include cytokine release syndrome, neurotoxicity, pneumonitis, and rash, among others. These conditions increase the disease burden and, in rare cases, can lead to fatality.<sup>47</sup> For example, immunotherapy-induced pneumonitis<sup>48</sup> is a type of lung inflammation that can cause symptoms such as coughing, shortness of breath, and chest pain; if untreated, it can lead to serious complications like respiratory failure. Hyperprogressive disease (HPD) is another phenomenon, wherein cancer grows at an accelerated rate after the initiation of IO treatment, leading to worse outcomes and reduced survival rates. The incidence of HPD varies between different tumor histologies and treatments. These adverse events limit the effectiveness of new IO treatments, as treatment may be discontinued due to a severe irAE.

Computational strategies to predict and monitor irAEs have become a priority in IO. Forecasting irAE incidences may help doctors adjust treatment plans to potentially prevent the worsening of the disease or design alternative therapies for patients. Radiomics from CT scans has been used to identify IO-induced pneumonitis<sup>50–52</sup>, and to predict treatment response and pneumo-toxicity from programmed cell death-1 pathway inhibition in patients with non-small cell lung cancer.<sup>52</sup> Radiomics can also distinguish between radiation- and IO-induced pneumonitis.<sup>53 54</sup> Identifying HPD is critical because it can influence the decision to continue or discontinue immunotherapy treatment, as well as the choice of alternative therapies that may be more effective for the patient. By analyzing radiologic images, AI algorithms have been demonstrated for predicting the risk of irAE such as pneumonitis and HPD.<sup>55</sup> This information can assist healthcare professionals in making more informed decisions about treatment plans, thereby improving patient outcomes and safety. Integrating clinical and radiomic parameters from baseline, pretreatment CT scans could facilitate the identification of HPD in patients with lung cancer being treated with immunotherapy.<sup>56</sup> The combination of radiomic parameters representing texture patterns of lung nodules and features related to vessel tortuosity of a nodule could distinguish between responders, non-responders, and hyperprogressors in immunotherapy.<sup>55</sup>

In addition to CT scans, recent studies have reported a correlation between specific tumor microenvironment (TME) characteristics, such as TIL density and irAE, with evidence suggesting organ-specific co-occurrence of toxicity.<sup>57 58</sup> Notably, predictive modeling has shown promise in anticipating drug combinations that minimize the risk of irAEs.<sup>57</sup> However, accurately predicting and understanding the incidence of irAE remains challenging

due to the complex nature of human immune responses, which vary significantly according to individual genetic backgrounds and environmental factors present across different hospital settings. Enhancing computational algorithms, as well as acquiring training data that encompasses diverse patient demographics and clinical contexts, should be priorities for the field. In particular, in the design of effective treatment strategies while minimizing the risk of irAE, developing interpretable models that provide insights into the underlying biological mechanisms driving irAE occurrence should be carefully considered.

### Cancer disparities

Disparities in IO are a significant concern to equitable and effective cancer care and can lead to differences in trial accrual, treatment outcomes, and survival rates.<sup>59</sup> These disparities are influenced by a variety of factors such as socioeconomic status, geographic location, race, ethnicity, sex, and underlying health conditions. To reduce health inequities and ensure that the benefits of IO treatments are accessible to all patient populations, it is essential to understand and address these factors. Here, we discuss two main aspects: the disparities in molecular profiling and trial access, and how compIO strategies may help overcome these barriers.

Disparities can lead to unequal access to healthcare, resulting in higher rates of disease and mortality in certain populations. This is also true for risk calculators developed to identify individual patient risk. For example, a report in *JAMA Oncology* indicated that the Oncotype DX assay, a 21-gene expression test designed to help identify which estrogen receptor positive, patients with lymph node negative breast cancer would benefit from adjuvant chemotherapy, was not prognostic when used in black women.<sup>60</sup> AI could help alleviate the issue of health disparities, especially in oncology, by providing more accurate and personalized diagnoses and treatments. However, to achieve this, an intentional focus is needed to ensure biases are not incorporated into AI algorithms.<sup>61</sup> AI could also assist in identifying potential morphologic and molecular differences between populations, allowing for the creation of more specific and accurate population-tailored risk prediction models. For instance, AI has shown promise in analyzing digital pathology images to identify morphologic differences in the appearance of prostate cancer between black and white patients, which in turn could aid in creating more population-tailored models that are more accurate in predicting post-surgical recurrence risk compared with population-agnostic models.<sup>62</sup> Such approaches could help reduce variability in cancer diagnosis and treatment, ensuring that all patients receive the best care possible. Additionally, AI can be used to develop targeted interventions to improve cancer outcomes in populations disproportionately affected by health disparities.

Financial toxicity refers to the economic burden experienced by patients with cancer due to the high cost of

cancer treatment,<sup>63</sup> which can lead to financial distress, psychological stress, and potentially compromise the quality of life for patients and their families. AI can play a crucial role in mitigating financial toxicity by analyzing routine data to identify patients who are likely to benefit from specific therapies. For example, AI could help predict patients who will, or will not, respond to ICI, which may cost nearly US\$200K per patient per year but only work in 20–25% of cases.<sup>64</sup> By analyzing routine CT or pathology images, AI can guide the use of these drugs through patient stratification towards those with a higher likelihood of clinical benefit, thereby enhancing treatment efficacy and reducing unnecessary financial strain.<sup>65 66</sup>

Addressing disparities in IO requires a multifaceted approach that includes policy changes, community engagement, and tailored healthcare strategies. Future efforts should focus on generating more data from molecular profiling and improving clinical trials accrual for under-represented populations, implementing education and outreach programs to increase awareness of IO treatments, building trust between doctors and the community, and developing policies that promote health equity. Research should also continue to explore the underlying causes of these disparities, with a particular focus on integrating social determinants of health into the design and implementation of IO therapies.

### Data integration

As discussed above, addressing computational and clinical challenges in IO requires multimodality data and proper strategies to identify meaningful signals without introducing bias. Multistudy and multimodal data integration is a pivotal aspect of compIO across diverse research domains. Careful integration of data from preclinical and clinical IO studies can enable (1) cross-trial correlative analyses for reverse translation and biological discovery, and (2) cross-assay studies that use multimodal data types for enhanced inference, which can guide biomarker development strategies.<sup>67</sup> Within these realms, multiple foundational studies have not only demonstrated the promise of data integration for IO discovery but have also exemplified the ongoing challenges.

Data integration across clinical trials is necessary because IO trials and the associated correlative analyses from trials and concomitant model systems are typically too small for biomarker identification, particularly with datasets that contain errors or inconsistencies. These errors can arise from a variety of factors, such as measurement errors, data entry errors, equipment malfunctions, or environmental factors. Noise in data can also be introduced by variability in sample collection and processing methods across different clinical sites, including sample handling, storage conditions, and preparation protocols. Additionally, differences in patient demographics, disease characteristics, and treatment regimens among clinical trials contribute to data heterogeneity. Thoughtful data integration across clinical trials helps advance knowledge so that



further studies are statistically powered to identify underlying molecular commonalities and differences between responding and non-responding patient populations. Such data reuse<sup>68</sup> requires careful statistical and analytic approaches, including conditioning on treatment regimens, harmonized clinical data elements and outcome associations, and standardization of molecular analysis pipelines. However, when done carefully, such analyses can reveal novel insights, especially in the context of ICI, which can induce durable treatment responses across different cancer histologies. For example, building on an initial cross-trial ICI molecular study of 249 patients across 5 histologies, a follow-up analysis<sup>69</sup> of over 1,000 samples from patients with solid tumors, including samples from multiple trials and contexts, revealed the role of different mutational signatures and candidate molecular targets that may be relevant to all tumors treated with ICI.

The aforementioned studies also highlight the difficulties that may be encountered when integrating data across trials. One of the major limitations in combining multiple studies is the harmonization of clinical data to enable patient selection and research discovery at scale following Common Data Model and FAIR (Findable, Accessible, Interoperable, Reusable) principles.<sup>70</sup> Variation in data resources such as electronic health records (EHR), pharmacy data, and patient registries, as well as nomenclature and annotations in healthcare systems can contribute to differences in data content and format.<sup>71</sup> Natural language processing and deep-learning techniques<sup>72</sup> such as large language models<sup>73 74</sup> can be leveraged to reduce such heterogeneity and enhance information extraction and standardization. Furthermore, it has been shown that even with thoughtful data integration, statistical power constraints cannot be overcome for cross-histology analysis, despite that the underlying data are properly harmonized.<sup>75</sup> Similarly, it has been demonstrated that after overcoming major challenges regarding harmonization through extensive computational techniques (eg, correction of batch effects stemming from numerous clinical and technical factors), no previously reported signatures of ICI response or those derived in the study could be validated across contexts.<sup>76</sup> While this demonstrates some of the challenges of cross-trial comparisons, it also highlights the critical need to validate data from one cohort in other independent and rigorously vetted cohorts with uniformly processed data. Companion findings in neoantigen discovery highlight similar standardization and reproducibility challenges.<sup>77</sup> Taken together, numerous ongoing efforts to harmonize the increasing number of cross-trial cohorts for uniform analysis and interpretation are critical to increase predictive power and minimize false positive signals within patient data in IO.

Similarly, computational innovations have enabled significant advances in understanding immune responses to cancer through the integration of data from different assays, such as gene expression profiling, flow cytometry, and imaging. For example, a combination of single-cell and spatial analysis was used to discover myeloid

cell-attracting hubs at the tumor-luminal interface in mismatch repair deficient tumors.<sup>78</sup> Transfer learning approaches can be used for annotating descriptions of immune cell function across reference atlases, which facilitates the standardized annotation of single-cell datasets,<sup>79</sup> and this need extends to the tracking of lymphoid cell state changes in the periphery across ICI studies.<sup>80</sup> Likewise, advances in deep learning (DL) are enabling the multimodal fusion of data types across many domains simultaneously, such as genomics, transcriptomics, and pathology images,<sup>81 82</sup> and these efforts may yield novel IO insights. However, similar to cross-trial considerations, innovations in statistical paradigms are critical to maximize the potential of multimodal high-dimensional data generated from the same specimens for inference. Furthermore, data heterogeneity and standardization issues must be addressed to ensure that data can be easily shared and integrated across different platforms.

Taken together, novel data integration and inference strategies may unlock generalizable insights for IO research programs. However, significant investments are necessary to harmonize data types across contexts as well as to innovate new methods for learning from complex, yet complementary, data types.

### Artificial intelligence

With the exponential growth in data volume and the complexity of cancer-immune biology, AI has transformed the methodologies we use to conduct research. As discussed in preceding sections, AI offers immense potential to enhance clinical trial design, improve data integration and fusion, and predict irAEs as well as therapy responses in IO.<sup>83</sup> However, alongside these promising advancements, it is important to approach AI strategies with a balanced focus on computational excellence, ethical considerations, and clinical relevance.

Traditionally, FDA-approved tissue clinical biomarkers for solid tumors, such as programmed death-ligand 1, microsatellite instability, and TMB, have been routinely used for patient stratification in ICI treatment.<sup>84</sup> Transcriptomic-based biomarkers have also been widely recognized for predicting ICI response, with available tests like NanoString's tumor inflammation signature (TIS) gene expression assay measuring suppressed adaptive immunity. Similarly, Tempus' Immune Profile Score (IPS) algorithmic test uses an AI-driven algorithm to integrate DNA and RNA data, providing a high/low IPS score that predicts response to ICI. Despite these advances, the accuracy of existing biomarkers is often limited, and their performance varies by cancer type and study cohort. Techniques such as dynamical network biomarkers (DNBs)<sup>85 86</sup> represents a promising approach by modeling multimodal and/or longitudinal data.<sup>85 86</sup> Unlike traditional biomarkers, which are often static and measured at a single time point, DNBs focus on capturing the dynamic changes in biological networks over time, potentially leading to improved reliability and reproducibility of biomarkers. For instance, DNB models have been

shown to determine a tipping point between immune control and immune evasion that can explain differential response in patients, and the distance of a patient from the tipping point may determine survival outcome.<sup>87</sup>

Radiology and pathology are areas where AI has demonstrated considerable progress in IO. AI algorithms have been used to analyze radiologic images and identify tumor characteristics, such as size, shape, and texture, and assess treatment responses. Radiomics, a technique that extracts quantitative features from images, has shown promise in predicting the response to immunotherapy in patients with melanoma.<sup>88</sup> The study found that radiomic features, such as tumor heterogeneity, entropy, and contrast, were significantly associated with clinical outcomes. In pathology, AI-driven analysis of histopathological images has enabled the identification of immune cells including location and density, predicting the response to immunotherapy. For example, one study used neural network to analyze histopathological images from patients with melanoma.<sup>89</sup> The study found that DL could accurately predict the response to immunotherapy based on the density and location of immune cells in the TME. AI algorithms can analyze large-scale genomic data, such as DNA sequencing data, to identify genetic mutations that may affect the response to immunotherapy. For instance, a study employed a convolutional neural network to combine genomic and clinical features to stratify patients with non-small cell lung cancer according to their response to immunotherapies.<sup>90</sup>

In recent years, the development of foundation models has catalyzed transformative breakthroughs across omics, imaging, and EHR in cancer research.<sup>91</sup> Using histopathology images and annotations sourced from social media and public databases, researchers have pioneered generative AI capable of identifying and annotating tumor, immune, and other cell types to assist with pathology education and diagnosis.<sup>92</sup> LMMs have also demonstrated utility in significantly enhancing knowledge accessibility by simplifying legal jargon into easily-understandable formats for medical patient consent forms.<sup>93</sup> Furthermore, the development of CONCH<sup>94</sup> and UNI,<sup>95</sup> pathology generative AI models trained on histopathology slides from academic medical centers and industry collaborators, has demonstrated high accuracy in recognizing cell types and generating pathology reports that resemble expert opinions.

AI has made remarkable advancements in clinical imaging however the progress in genomic, transcriptomic, and spatial data has been slower. This discrepancy arises from the inherent biological complexity, technical noise, different platforms, and smaller cohort sizes of omics data, which hinders model training and validation. Collaborative efforts to create large, standardized multi-omics datasets with harmonized clinical annotation, such as Immuno-Oncology Translational Network (IOTN), Cancer Immunoprevention Network, and Human Tumor Atlas Network (HTAN), are essential to accelerate AI advancements in these areas. Simultaneously, the rapid

evolution of AI in healthcare faces regulatory and policy challenges that have not kept pace with technological advancements. The practical integration of AI into clinical practice remains uncertain due, in part, to regulatory and accountability concerns. From a methodology perspective, ethical, representation, and technological biases are inherent from the data collection stage, and persist throughout model development and ultimately to downstream decision-making and applications. Strategies that make such biases more visible, accountable, and quantifiable should be prioritized, along with the development of AI solutions capable of effectively correcting for bias and validation of their performance in real-world datasets.

### Spatial biology

Moving beyond predictive biomarkers, it is important to understand the underlying biology that drives correlative patterns and causal relationships. Understanding why patients respond differently to treatments, why some develop irAEs while others do not, and how we can design new clinical trials to overcome resistance are all pressing issues in advancing IO. High-resolution, high-throughput spatial technologies offer a unique advantage in dissecting the tumor-immune ecosystem, systems immunity, and host-environmental interactions, ultimately linking these insights back to clinical outcomes.

With co-detection by indexing (CODEX),<sup>96</sup> tissue-based cyclic immunofluorescence microscopy,<sup>97</sup> and multiplexed ion beam imaging, researchers can now scale the number of traditionally measured biomarkers from 10 or fewer to upwards of 40–100 antibodies simultaneously for a single FFPE (formalin-fixed, paraffin-embedded) slide of tumor tissue. This expansion in data capturing capabilities is critical for in-depth investigations of cell-cell communications within and between distinct tumor niches. Notably, studies have used these technologies to identify unique subsets of cell populations associated with patients' responses to ICIs and CAR T-cell therapy.<sup>98–100</sup> On account of further achievements, single-molecule localization imaging by photoactivated localization microscopy (PALM) and stochastic optical reconstruction microscopy (STORM) can now provide intracellular levels of resolution for target transcripts, proteins, or metabolites of interest, as well as spatial data on subcellular organelles such as lysosomes and extracellular vesicles, which play important roles in cancer immunoediting and immunotherapy.

On a systems level, recent advances in computational reconstruction of tissues from sequentially sectioned H&E (hematoxylin and eosin)-stained images have revealed tumor-lymphocyte interfaces that contribute to cancer progression.<sup>101</sup> Because FFPE tissues are commonly stored by biorepositories, there exists a vast resource that can be used for high-resolution spatial analyses, such as CODEX,<sup>96</sup> Visium, GeoMX, CosMX,<sup>102</sup> and other MICCCS (Multiplex Immuno-Cohort Characterization in Cancer Studies) technologies. These

technologies are capable of analyzing multiple protein and RNA biomarkers in tissue samples simultaneously, thus providing deeper insights into the complexity of tumor biology. There is a growing need to develop computational tools that are accurate and scalable for the analysis of such spatial imaging data. Current approaches are designed to accommodate specific assays with potential to scale, such as STARmap,<sup>103</sup> HistomicsML,<sup>104</sup> and others. Of note, super-resolution spatial gene expression prediction from histopathology images<sup>105</sup> and three-dimensional spatial transcriptomics<sup>106 107</sup> have become a reality with the integration of spatial omics and sequential sectioned H&E slides, leveraging AI and the unique morphology of cells in tumor.

Omics technologies that augment the dimensionality of tissue imaging have enabled spatially resolved, large-scale profiling of the human transcriptome, proteome, and metabolome. For example, advances in digital spatial profiling have provided new insights into the architecture of tertiary lymphoid structures and responses to immunotherapies in melanoma.<sup>108 109</sup> Recent studies on cancer evolution have demonstrated that single-cell copy number alterations, inferred from 10× Visium spatial transcriptomics, can be used to detect early events that arise in normal tissues before malignancy<sup>110</sup> and intra-tumor heterogeneity in immunotherapy response.<sup>111</sup> The broader application of mass spectrometry (MS) imaging, such as matrix-assisted laser desorption/ionization, in IO studies continues to grow and is starting to uncover the underlying tumor metabolism-driven resistance mechanisms to ICIs. Multiple NCI-supported spatial and imaging initiatives, including those of the Human BioMolecular Atlas Program and the HTAN, are starting to untangle the spatiotemporal characteristics of the TME. Such studies have enabled the multifaceted discovery of spatial biomarkers and computational algorithms driven by ML.<sup>80 112</sup>

While spatial omics and imaging technologies offer invaluable biological insights, their clinical scalability is hindered by cost considerations. In contrast, H&E staining is a routine and readily available process in clinical practice, applicable to almost all patients without additional processing. However, H&E staining has its limitations: the interpretation can be subjective and vary between pathologists, leading to inconsistencies; and variations in tissue preparation and fixation protocols can affect staining quality. Prioritizing strategies that harness AI for stain normalization<sup>113</sup> as well as create spatially aware biomarkers and elucidate tumor heterogeneity in the context of therapy response should be the next direction for clinical translation efforts. For example, studies have shown that AI can analyze digital pathology images to accurately predict TIL levels,<sup>114</sup> which have been linked to the response to ICIs. Additionally, AI can analyze the spatial distribution and heterogeneity of TILs within the TME, which have been shown to be associated with the response to ICIs.<sup>66</sup> With these insights, clinicians and researchers can better identify patients who are

likely to respond to ICIs and potentially improve patient outcomes.

### Tumor antigen discovery

As spatial biology continues to evolve as an exploratory field, our focus shifts towards driving biology-informed advancements in CompIO that can translate into tangible changes in clinical practice. A prime example of this is discovering and characterizing tumor antigens for new and effective IO treatments. While cellular immunotherapies targeting somatic mutation-derived antigens (ie, “neoantigens”) or lineage specific antigens have been remarkably successful, suitable antigen targets remain elusive for many cancer types. Recent studies indicate that RNA-level dysregulation, such as cancer-specific RNA processing events, can generate a large catalog of tumor antigens for potential immunotherapy targeting.<sup>115</sup> However, various computational challenges exist for discovering and targeting this novel class of tumor antigens. Classic short-read RNA sequencing (RNA-seq)-based methods for interrogating cancer transcriptomes are limited in their ability to predict full-length transcript and protein products in cancer cells.<sup>116</sup> Newer long-read RNA-seq-based methods allow end-to-end sequencing of full-length transcripts, but their lower base-calling accuracy and modest sequencing yield represent a major bottleneck.<sup>117</sup> New computational tools that leverage short-read and long-read RNA-seq technologies as well as data resources, and address their relative technological limitations, are urgently needed. Such computational tools will enable a more comprehensive and accurate characterization of cancer transcriptomes and proteomes, thus improving antigen discovery.

Multiple factors need to be considered when prioritizing and selecting antigen targets for therapy development. The cellular heterogeneity of antigen targets across cancer and normal tissues may critically affect therapy outcomes, yet such heterogeneity remains to be comprehensively assessed. Recognizing the risks associated with invasive tumor specimen collection, patients’ safety must be a primary consideration when designing protocols for multisite sample collection. Another important topic is the choice between “public” versus “personalized” antigen targets. While public antigens shared by many patients may have broader clinical applicability, there may be safety concerns due to off-tumor expression in normal tissues. In contrast, personalized antigens may provide targets with higher tumor specificity and lower off-tumor toxicity, but concerns about cost and feasibility may hinder their adoption. Biases in training data, including the scarcity of ethnically diverse patient cohorts, may also limit the accuracy of antigen prediction in all populations requiring further research. Computational tools and large-scale data resources, representing a wide array of diverse immunogenetic backgrounds, may help address these issues. Specifically, computational tools for single-cell or spatial analysis of RNA and protein variation across diverse populations and cell types may pinpoint



antigen targets with high efficacy, low toxicity, and broad utility, and bring equitable therapeutic opportunities to a broader patient population.

Cancer immunotherapies such as ICIs are designed to unleash antitumor T cells, which in turn mediate antitumor immunity that provides long-term protection. However, open questions remain about the mechanisms that lead to a successful antitumor T-cell response: What determines the fate trajectories of therapeutically relevant T cells? How can we systematically identify tumor-reactive T cells and their cellular phenotypes? Insights into these questions will help characterize the specificity and strength of endogenous antitumor T-cell responses during immunotherapy, which may lead to significant advances in vaccine and immunotherapy engineering. Despite the increasing number of high-throughput, standardized technologies to experimentally identify tumor-antigen-specific T cells,<sup>118–121</sup> experiments are costly and few studies to date have successfully identified neoantigen-specific T cells.<sup>122–125</sup> In individuals with cancer, only a modest number of neoantigen-specific T cells are detected in the TME; this is confounded by the presence of bystander T cells with viral specificities.<sup>122 123 125–127</sup> Remarkably, neoantigen-specific T cells exhibit congruent transcriptional programs across cancer studies.<sup>122–125</sup> This supports the joint analysis<sup>128 129</sup> of existing single-cell transcriptomic datasets paired with matching TCR data as a basis to more effectively uncover transcriptomic and TCR sequence similarities that underlie tumor-reactive T cells. However, understanding the technology used to generate these datasets is essential. Single-nuclei RNA-seq often under-represents immune cells compared with single-cell RNA-Seq (scRNA-seq),<sup>130</sup> and unlike the 5' sequencing protocol, 3' sequencing protocol (10× Genomics) does not cover TCR or B-cell receptor sequences. These technical differences highlight the need for careful consideration in selecting single-cell datasets that provide comprehensive coverage of both the transcriptome and immune receptor repertoire to accurately identify tumor-reactive T cells.

Advances in AI for protein structure prediction have radically transformed the field of protein optimization and de novo design.<sup>131</sup> While significant progress has been made in ML-based methods for T-cell antigen specificity prediction, a hallmark problem in immune engineering, the complexity and variability of antigen interactions continue to present substantial challenges. Unsupervised ML methods have been developed to annotate T-cell antigen specificity based on cross-referencing with the known TCR-epitope databases<sup>132</sup> or by calculating similarities between TCR sequence properties and/or transcriptional profiles in single cells.<sup>133 134</sup> Supervised methods, including DL models, have been developed to use TCR-antigen pairs<sup>135</sup> as training data to predict epitope specificity, achieving some success in model training but with diminished predictability for unseen antigens.

One of the major barriers in building ML models to predict T-cell antigen specificity is the scarcity of paired

TCR-antigen specificity data. Single-cell studies have yielded insights into transcriptional programs of individual T-cell clones and offered clues about T-cell antigen specificity.<sup>136</sup> However, with the exception of a few single-cell studies to date,<sup>137</sup> TCR and cognate antigen pairs are largely unlabeled or lacking. Publicly available TCR-epitope databases (VDJdb, McPAS-TCR, MIRA, and IEDB) contain an abundance of viral antigen specificity data, and thus, bias can be introduced when using these data for model training. Hence, the paucity of TCR-tumor antigen data poses a significant challenge toward building robust ML models to predict T-cell specificity to tumor antigens.

### Data sharing, access, and curation

CompIO has the potential to transform cancer care by enabling more personalized treatments, improving patient outcomes, and reducing healthcare costs. However, challenges remain such as acquiring diverse datasets, standardizing data integration and analysis, and validating AI algorithms in clinical environments. The Surgery Committee of the Society for Immunotherapy of Cancer (SITC) delineates best practices for tissue procurement in IO trials, highlighting critical factors like tissue site, tumor representation, and preservation, which are essential for harvesting tumor-reactive T cells.<sup>138</sup> Moreover, accurate documentation of tissue resources and metastatic burden is crucial in IO, exemplified by the immunosuppressive effect of liver metastases.<sup>139</sup> In data management, standardization is key to reducing the heterogeneity in gene and protein markers defining immune cell subsets for increased dataset comparability. Efforts by the NCI's Cancer Immune Monitoring and Analysis Centers and Cancer Immunologic Data Commons (CIMAC-CIDC) Network<sup>140</sup> and Cancer Research Data Commons,<sup>141</sup> including the recent introduction of the Immuno-Oncology Data Commons, aim to standardize clinical data models across trials and studies and enhance the understanding of how treatments alter the tumor-immune landscape.

Resource sharing, including the sharing of clinical sample and data, is crucial for advancing research in IO. The National Institutes of Health's (NIH) Data Management and Sharing (DMS) Policy underscores the importance of de-identifying and securely managing human subject data.<sup>142–145</sup> Despite these guidelines, the implementation of data-sharing policies is often hindered by long processes like establishing Material Transfer Agreements and Data Use Agreements, which can significantly delay research. To overcome these obstacles, innovative approaches are needed to expedite data sharing while safeguarding privacy. Potential solutions could include streamlined institutional review processes for quicker data access, template consents for genomic data sharing, the generation of synthetic data that retains utility for result reproduction and model training, or the use of privacy-preserving methods like federated learning, which allows



for comprehensive model validation without direct data exchange.

Resource sharing through platforms like the NCI's NCTN Navigator<sup>142</sup> is mandatory across oncology NCTN groups, facilitating access to clinical trial biospecimens and annotated clinical data for assay and biomarker development. The application processes typically involve initial discussion based on preliminary data with the group that conducted the trial, followed by submission of a detailed proposal for review by the Core Correlative Sciences Committee, overseen by the NCTN Program and consisting of independent peer-reviewers from the cancer research community. Both academic and private industry entities use NCTN Navigator to obtain histopathology images and radiologic scans for AI development and validation in outcome and treatment response prediction.<sup>146 147</sup> The CIMAC-CIDC Network, by collecting biospecimens from immunotherapy trials,<sup>140</sup> empowers community-wide research efforts to identify new biomarkers and therapeutic targets to enhance treatment strategies for patients. This endeavor requires advanced computational, mathematical, and bioinformatics tools to drive innovations in drug development and combination therapies.

In CompIO, where ML and DL models are increasingly used in clinical applications including designing personalized cancer vaccines and T-cell therapies, maintaining the highest standards of reproducible research is essential. This requires rigorous preprocessing, quality control, and transparent code sharing. The complexity of ML/DL workflows in this domain presents unique challenges. Automation tools, including Nextflow<sup>148</sup> or Snakemake,<sup>149</sup> enable the creation of scalable, reproducible, and sustainable computational workflows that can be easily shared and adapted by the community. Centralized platforms such as Google Colab<sup>150</sup> and Code Ocean, along with containerization technologies like Docker and Singularity, facilitate data and code sharing by allowing researchers to rebuild the models and predictions within consistent environments. Open-source software enables community-wide assessment and reuse, ensuring the accuracy of these methods and providing additional validation.<sup>151</sup> Software ecosystems that offer peer-reviewed packages,<sup>152</sup> such as R/Bioconductor and Python's PyPI, provide an additional layer of quality assurance. The practice of reproducible research may require substantial overhead and additional personnel beyond the efforts of a single analyst, particularly given the complexity of integrating multiomics data to understand antitumor immunity. However, this investment is justified to ensure the reliability of findings that could directly influence patient care. Nonetheless, balancing the protection of intellectual property and adherence to best practices in reproducible research will be the key to advance both innovation and credibility in clinical applications.

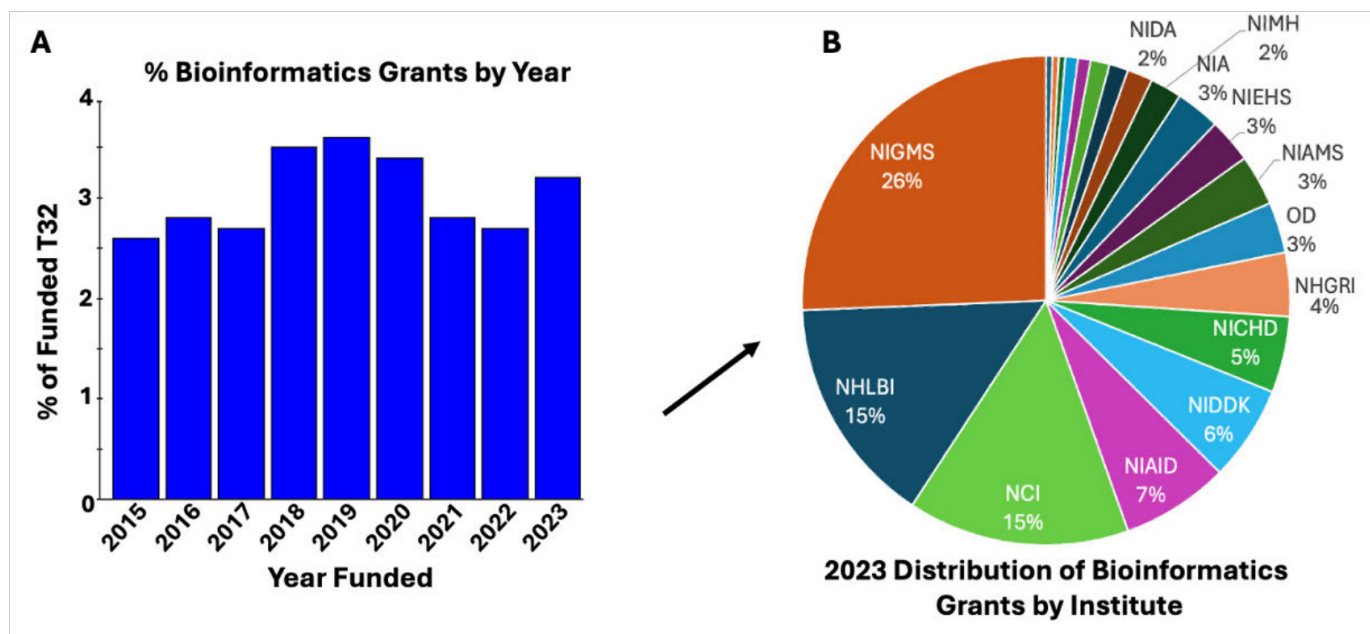
As new biomarkers are developed and technologies continue to evolve, the lack of universal standards for in silico validation, reproducibility, and mechanistic analysis

significantly hinders the clinical translation of computational discoveries in IO. Enforcing bidirectional bench-to-bedside research and developing community-driven best practices for AI in IO are imperative for ethical and effective patient care. Collaborative efforts, particularly within multidisciplinary teams, are vital to ensure that computational findings are relevant and rigorously validated with ethical consideration. Comprehensive real-world validation of new biomarkers and predictions is crucial, especially in diverse patient populations, to underpin biomarker-guided trials. The design of clinical trials must be grounded in a strong biological rationale, integrating functional studies to delineate causality and avoid oversimplified drug combinations. By adhering to these rigorous validation standards, the field can improve the efficiency and success rate of clinical translations.

### Education and training across computational, biological, and clinical domains

Breakthroughs in high-throughput sequencing, drug screening, and MS-based technologies have led to the rapid accumulation of large-scale omics data across a multitude of disciplines, including IO.<sup>153</sup> The increased access to and capability of deriving such datasets demand the proper training and retention of individuals equipped with the skill sets required to properly leverage and interpret them for novel biological insight. Therefore, a concerted effort to boost the number of skilled bioinformatics scientists, computational biologists, and data scientists to effectively handle this profusion of data is critical for advancing IO research.

Given the multidisciplinary nature of bioinformatics and the challenge in designing cross-discipline courses, developing such programs has been complicated by a limited consensus in the field regarding the essential knowledge for a successful bioinformatician, and the most effective means of delivering it.<sup>154</sup> This lack of agreement was compounded by the wide range of scientific and non-scientific fields in need of computational training, each with different backgrounds and intended applications. The Curriculum Task Force of the International Society for Computational Biology Education Committee has provided a framework curriculum, which outlines a set of bioinformatics core competencies that span various disciplines, backgrounds, and training programs. These core competencies include topics such as recognizing and critically reviewing the format, scope, and limitations of different biological data-generating platforms, technology applications, and methodology in consideration of the experimental material available and experimental objectives, data management plan, and reproducibility. Bioinformatics education programs range from informal workshops and training courses to structured, didactic learning in the form of certified degree programs, including doctoral-level education.<sup>155 156</sup> Many of these programs are offered at institutions with Comprehensive Cancer Centers in the USA. However, the training might not be cancer-specific and could benefit from a tailored,



**Figure 2** Limited growth in bioinformatics-focused institutional research training (T32) grants. Despite the significant increase of data in omics space, funded T32 grants targeted towards bioinformatics training have seen only a modest rise from 2.7% to 3.5% over the past 9 years (A). The distribution of these grants is uneven among NIH funding agencies (B). Our search results for funded T32 grants over the past 22 years used five search terms: “Bioinformatics”, “Computational Biology”, “Computational Oncology”, “Immune Bioinformatics”, and “Omics”. We acknowledge the potential incompleteness of awards from this search. NCI:National Cancer Institute; NIDA:National Institute on Drug Abuse; NIMH:National Institute of Mental Health; NIA:National Institute on Aging; NIEHS:National Institute of Environmental Health Sciences; NIAMS:National Institute of Arthritis and Musculoskeletal and Skin Diseases; OD: Office of the Director; NHGRI:National Human Genome Research Institute; NIGMS:National Institute of General Medical Sciences; NHLBI:National Heart, Lung, and Blood Institute; NIAID:National Institute of Allergy and Infectious Diseases; NIDDK:National Institute of Diabetes and Digestive and Kidney Diseases; NIH: National Institutes of Health; NICHD:Eunice Kennedy Shriver National Institute of Child Health and Human Development.

widespread, open-source curriculum provided by initiatives such as the NCI ITCR Training Network<sup>152</sup> and the SITC-NCI Computational Immuno-oncology Webinar Series.<sup>157</sup>

Currently, funded training grants in bioinformatics lag far behind their experimental counterparts, which support trainees producing vast amounts of data requiring advanced analytical attention. This discrepancy suggests a further widening gap and a diminishing capacity to meet computational demands. For instance, the number of T32 awards for bioinformatics training programs has increased only incrementally, despite the exponential growth of data in the omics space (figure 2A). Moreover, few NIH agencies have funded these awards to date, resulting in pronounced disparities among them (figure 2B). Further, none of the currently funded awards specifically focus on multidisciplinary training in areas of need, such as computational training within the IO space, where there are unique data challenges unmet by generic computational training programs and a rapidly rising clinical urgency. In total, the training of individuals in bioinformatics and computational biology has not kept pace with the advances in high-throughput omics techniques and big data production, leading to a bottleneck and dampened productivity in the translational research pipeline. Collectively, these examples highlight an urgent

need for educators and funding agencies to support a sufficient number and variety of programs in CompIO accessible to trainees at all levels.

Significant emphasis has been placed on providing bioinformatics training necessary to enable the implementation of bioinformatics research. As multidisciplinary teams are becoming the norm, it is important to acknowledge that many computationally trained investigators might have limited training and experience in clinical research. The issues of liability, patient safety, patient privacy, and clinical research ethics, which are second nature to clinical investigators, could be foreign concepts to bioinformatics scientists and computational biologists. Consequently, not all members of a research team may be equally comfortable with or capable of rapidly adapting to implementing these approaches in the context of clinical care. Therefore, training programs that also provide in-depth clinical research training to computational scientists play a vital role in the ethical use of data science and AI methods in translational IO research.

### Team science

Bringing together diverse expertise in team science is fundamental for accelerating progress in IO and solving its complex problems. Knowledge sharing and the establishment of new computational solutions have been key

drivers of new scientific discoveries.<sup>112 158 159</sup> As the scale and complexity of data continue to increase, collaboration with bioinformatics scientists, computational biologists, and data scientists—who develop and apply sophisticated algorithms for data evaluation, integration, and interpretation—has become essential.

Historically, bioinformatics scientists were seen as playing supportive roles in translational or clinical research. However, this outdated view has shifted to a modern, convergence science model. Today, research often begins with large-scale data production and in silico discoveries on human specimens, proceeds to in vitro and in vivo validations, and ultimately moves the needle in clinical practice, all in collaboration with experts across computational, biological, and clinical domains. This culture shift is reflected in a new trend of transdisciplinary research, where laboratories specialized in high-dimensional computational and biological techniques recruit bench scientists for functional experiments and seek clinical partners to translate discoveries into improved patient outcomes. To recruit and retain talent in academia, it is important to adequately recognize the contributions of bioinformatics scientists, computational biologists, and data scientists in team science projects. This requires innovative strategies beyond traditional tenure stream evaluations. Several institutions have implemented a new “team science” track for faculty, where collaborations are valued as equally important as single-lead projects. The general trend of multi-PI (principal investigator) grants as well as increasing the number of coauthors on transdisciplinary IO papers demonstrates the growing importance and impact of collaborations among experts from different fields.<sup>160</sup>

The NCI has implemented several programs and networks to incentivize team science. These initiatives encourage collaboration and coordination among researchers from different disciplines, institutions, and sectors. Some specific examples include the IOTN,<sup>161</sup> the Specialized Programs of Research Excellence, and the Cancer Center Support Grants. With the implementation of the new NIH DMS policy,<sup>162</sup> there is an opportunity to further promote team science, especially involving contributions from bioinformatics scientists, computational biologists, and data scientists. This policy requires that researchers make data from their NIH-funded studies publicly available and encourages collaboration and data sharing among researchers. This can facilitate the exchange of data and knowledge across fields, leading to new insights and discoveries, as well as new opportunities to promote the development of new data analysis tools and resources. Overall, the new DMS policy aims to break down silos and, hopefully, will bring us closer to the goal of building a more inclusive, diverse, and collaborative future for the next generation of scientists in cancer research.

## CONCLUSIONS

The journey through the 10 critical challenges and opportunities of compIO highlights the intertwined considerations between clinical needs and technological advancement. From the nuanced complexities of clinical trial design to irAE prediction, the need for robust computational methodologies is evident. Addressing disparities and socioeconomic factors emphasizes the importance of equity in trial access and treatment, which will require a global approach to data collection, algorithm development, and resource sharing. The call for standardization of validation, reproducibility, and code transparency demands universal benchmarks to ensure the ethical and effective implementation of AI-driven strategies. The integration of multidisciplinary efforts remains key to navigating the landscape of IO research, where collaborative innovation can lead to breakthroughs in improving patient outcomes through personalized immunotherapies.

Looking ahead, we have started to witness the transformative changes that AI might bring to healthcare, particularly at the early stage of clinical development and implementation. Future directions involve leveraging AI not only to predict but also to prevent adverse events, democratize access across patient populations, and enhance the accuracy and efficiency of both prediction and biological interpretation. In addition, AI approaches are being embedded as part of prospective IO-related clinical trials.<sup>163</sup> The establishment of global standards for data and resource sharing will pave the way for a more cohesive and inclusive approach to IO discoveries. Ultimately, the collective endeavor in compIO seeks to break the conventional boundaries of science and technology, toward building a future where every patient can benefit from the promise of immunotherapies.

## Author affiliations

<sup>1</sup>UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

<sup>2</sup>Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

<sup>3</sup>Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA

<sup>4</sup>Emory University, Atlanta, Georgia, USA

<sup>5</sup>Georgia Institute of Technology, Atlanta, Georgia, USA

<sup>6</sup>Atlanta Veterans Affairs Medical Center, Atlanta, Georgia, USA

<sup>7</sup>Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, California, USA

<sup>8</sup>Genomics Institute, University of California Santa Cruz, Santa Cruz, California, USA

<sup>9</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA

<sup>10</sup>Center for Biomedical Informatics & Information Technology, National Cancer Institute, Bethesda, Maryland, USA

<sup>11</sup>Department of Oncology, Johns Hopkins University, Baltimore, Maryland, USA

<sup>12</sup>Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA

<sup>13</sup>Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, Maryland, USA

<sup>14</sup>Cancer Diagnosis Program, National Cancer Institute Division of Cancer Treatment and Diagnosis, Bethesda, Maryland, USA

<sup>15</sup>Dana-Farber Cancer Institute, Boston, Massachusetts, USA

<sup>16</sup>Center for Biomedical Informatics and Information Technology, National Cancer Institute, Rockville, Maryland, USA



<sup>17</sup>Department of Public Health Sciences, University of Chicago Division of the Biological Sciences, Chicago, Illinois, USA

<sup>18</sup>Department of Pathology, Stanford University, Stanford, California, USA

<sup>19</sup>Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA

<sup>20</sup>Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA

<sup>21</sup>Department of Immunology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

<sup>22</sup>Informatics and Data Science Program, Center for Biomedical Informatics and Information Technology, National Cancer Institute, Bethesda, Maryland, USA

<sup>23</sup>Cancer Data Science Laboratory, National Cancer Institute, Bethesda, Maryland, USA

<sup>24</sup>Division of Computing and Mathematical Science, Caltech, Pasadena, California, USA

<sup>25</sup>Howard Hughes Medical Institute, Chevy Chase, Maryland, USA

<sup>26</sup>Department of Thoracic and Head and Neck Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>27</sup>Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>28</sup>Harvard Medical School, Boston, Massachusetts, USA

<sup>29</sup>Division of Population Sciences, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

<sup>30</sup>Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA

<sup>31</sup>Cancer Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

<sup>32</sup>Center for Computational and Genomic Medicine, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

<sup>33</sup>Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

<sup>34</sup>Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

**X** Riyue Bao @RiyueSunnyBao, Alan Hutson @Amherst73, Anant Madabhushi @anantm, Edgar Gonzalez-Kozlova @EdgarEGK and Eliezer M Van Allen @vanallenlab

**Acknowledgements** The authors thank Aviv Regev, PhD, from Genentech for contributing to the National Cancer Institute computational immuno-oncology workshop and the preparation of the manuscript. Additionally, we thank Jason J Luke, PhD, from the University of Pittsburgh for reading the manuscript and providing feedback.

**Contributors** RB, AH, AM, DM, SL, EMVA and YX conceived the study. RB, AH, AM, VDJ, SRR, JSB-S, EJF, DM, SL, EMVA and YX substantively drafted and revised the paper. HM, LH, JA, QC, JD, AJG, EG-K, SG, EK, ML, MM, ER, DVV, HZ and NV drafted, reviewed or revised the paper.

**Funding** Scientific and financial support for the Immuno-Oncology Translational Network (IOTN) are provided through the National Cancer Institute (NCI) Cooperative Agreements, U01CA233097, U01CA233056, U01DE029255, UG3CA244687, UG3CA244697, U01CA239258, U01CA232758, U01CA233078, U01AA027681, U01CA233084, U01CA233102, U01DE028227, U01DE028233, U01CA233100, U01CA233074, U01CA233085, U01CA233096, U01CA244291, U01CA244452, U01CA244314, U01DE029188, U01CA247573, U01AI156189, U01DK124165, U01AR077511, U54CA244711, U54CA244438, U54CA244726, U54CA244719, U24CA233032 (co-funded with NIHMD), and U24CA232979. Scientific and financial support for the Cancer Immune Monitoring and Analysis Centers-Cancer Immunologic Data Center (CIMAC-CIDC) Network are provided through the NCI Cooperative Agreements, U24CA224319 (to the Icahn School of Medicine at Mount Sinai CIMAC), U24CA224331 (to the Dana-Farber Cancer Institute CIMAC), U24CA224285 (to the MD Anderson Cancer Center CIMAC), U24CA224309 (to the Stanford University CIMAC), and U24CA224316 (to the CIDC at Dana-Farber Cancer Institute), and through NCI contract 140D0421D0007 to the CIDC operated by NCI. RB acknowledged National Institutes of Health (NIH) grants R01DE031729, P50CA097190, P50CA254865, P50CA272218, P30CA047904, and American Lung Association (ALA) Innovation Award IA-1273956. AH, MM and SL acknowledged NIH grant U24CA274159. NV acknowledged NIH grant R01CA276178. YX acknowledged NIH grants R01CA287673 and P30CA016520-47S3.

**Competing interests** RB declares PCT/US15/612657 (Cancer Immunotherapy), PCT/US18/36052 (Microbiome Biomarkers for Anti-PD-1/PD-L1 Responsiveness: Diagnostic, Prognostic and Therapeutic Uses Thereof), PCT/US63/055227 (Methods

and Compositions for Treating Autoimmune and Allergic Disorders). AM is an equity holder in Picture Health, Elucid Bioimaging, and Inspirata. Currently, he serves on the advisory board of Picture Health, and SimBioSys. He currently consults for Takeda; He also has sponsored research agreements with AstraZeneca and Bristol Myers-Squibb; his technology has been licensed to Picture Health and Elucid Bioimaging; he is also involved in two different R01 grants with Inspirata; he also serves as a member for the Frederick National Laboratory Advisory Committee. VDJ has performed consulting for VincerX Pharmaceuticals, Providence St. John, Los Angeles; she is a founder of Bioinformatica. JSB-S, EK, DM are full time, paid employees of NCI/NIH. EJF is on the Scientific Advisory Board of Resistance Bio, a consultant for Mestag and Merck, and receives research funding from AbbVie. SG reports current and past research funding from Regeneron Pharmaceuticals, Boehringer Ingelheim, BMS (Celgene), Genentech, EMD Serono, Pfizer, and Takeda. ER is a co-founder of MedAware, Metabomed and Pangea Biomed (divested), and an unpaid member of Pangea Biomed's and GSK Oncology scientific advisory boards. DVV is a co-founder and chief scientist of Barrier Biosciences and holds equity in the company. NV is on the Consulting/Advisory Boards of Sanofi/Genzyme (2022), Oncocyte (2021), Eli Lilly (2021), Regeneron (2022), Amgen (2023), Xencor (2023), AstraZeneca (2023), Tempus (2023), Pfizer (2024), Summit Therapeutics (2024), OncoHost (2024); she reports travel reimbursement from Regeneron; research grants from Oncocyte (2022–), Circulogene and Mirati, funding from Circulogene, and honoraria from Nebraska Oncology Society, Scienomics Group, Grace, OnLive and OMNI-Oncology. EMVA reports personal fees from Tango Therapeutics, Genome Medical, Genomic Life, Monte Rosa Therapeutics, Manifold Bio, Enara Bio, Riva Therapeutics, Foaley & Hoag, and Serinus Bio, grants and personal fees from Novartis Institute for Biomedical Research, and grants from BMS, Janssen, and Sanofi outside the submitted work; in addition, EMVA has a patent for institutional patents filed on chromatin mutations and immunotherapy response and methods for clinical interpretation pending.

**Patient consent for publication** Not applicable.

**Ethics approval** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Riyue Bao <http://orcid.org/0000-0002-6105-1704>

Anant Madabhushi <http://orcid.org/0000-0002-4141-8548>

Mark Long <http://orcid.org/0000-0003-1120-8176>

Eytan Ruppim <http://orcid.org/0000-0003-4299-7657>

Natalie Vokes <http://orcid.org/0000-0002-3766-5335>

Eliezer M Van Allen <http://orcid.org/0000-0002-0201-4444>

## REFERENCES

- Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol* 2020;20:651–68.
- Tang J, Shalabi A, Hubbard-Lucey VM. Comprehensive analysis of the clinical immuno-oncology landscape. *Ann Oncol* 2018;29:84–91.
- Shiravand Y, Khodadadi F, Kashani SMA, et al. Immune checkpoint inhibitors in cancer therapy. *Curr Oncol* 2022;29:3044–60.
- Sterner RC, Sterner RM. CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer J* 2021;11:69.
- van den Berg JH, Heemskerk B, van Rooij N, et al. Tumor infiltrating lymphocytes (TIL) therapy in metastatic melanoma: boosting of neoantigen-specific T cell reactivity and long-term follow-up. *J Immunother Cancer* 2020;8:e000848.
- Lin MJ, Svensson-Arvelund J, Lubitz GS, et al. Cancer vaccines: the next immunotherapy frontier. *Nat Cancer* 2022;3:911–26.
- Ghaffari Laleh N, Liger M, Perez-Lopez R, et al. Facts and hopes on the use of artificial intelligence for predictive immunotherapy biomarkers in cancer. *Clin Cancer Res* 2023;29:316–23.
- Augustin RC, Cai WL, Luke JJ, et al. Facts and hopes in using omics to advance combined immunotherapy strategies. *Clin Cancer Res* 2024;30:1724–32.

- 9 Prelaj A, Miskovic V, Zanitti M, *et al.* Artificial intelligence for predictive biomarker discovery in immuno-oncology: a systematic review. *Ann Oncol* 2024;35:29–65.
- 10 Computational Immuno-oncology Webinar Series, Available: <https://www.sitcancer.org/education/webinars-online/computational>
- 11 Valentiniuzzi D, Jeraj R. Computational modelling of modern cancer immunotherapy. *Phys Med Biol* 2020;65:24TR01.
- 12 Ruiz-Martinez A, Gong C, Wang H, *et al.* Simulations of tumor growth and response to immunotherapy by coupling a spatial agent-based model with a whole-patient quantitative systems pharmacology model. *PLoS Comput Biol* 2022;18:e1010254.
- 13 Bai JPF, Earp JC, Florian J, *et al.* Quantitative systems pharmacology: Landscape analysis of regulatory submissions to the US Food and Drug Administration. *CPT Pharmacometrics Syst Pharmacol* 2021;10:1479–84.
- 14 Chelliah V, Lazarou G, Bhatnagar S, *et al.* Quantitative systems pharmacology approaches for immuno-oncology: adding virtual patients to the development paradigm. *Clin Pharmacol Ther* 2021;109:605–18.
- 15 Hosseini I, Gadkar K, Stefanich E, *et al.* Mitigating the risk of cytokine release syndrome in a Phase I trial of CD20/CD3 bispecific antibody mosunetuzumab in NHL: impact of translational system modeling. *NPJ Syst Biol Appl* 2020;6:28.
- 16 Kosinsky Y, Dovedi SJ, Peskov K, *et al.* Radiation and PD-(L)1 treatment combinations: immune response and dose optimization via a predictive systems model. *J Immunother Cancer* 2018;6:17.
- 17 Sung W, Hong TS, Poznansky MC, *et al.* Mathematical modeling to simulate the effect of adding radiation therapy to immunotherapy and application to hepatocellular carcinoma. *Int J Radiat Oncol Biol Phys* 2022;112:1055–62.
- 18 Zhou J, Wang G, Chen Y, *et al.* Immunogenic cell death in cancer therapy: present and emerging inducers. *J Cell Mol Med* 2019;23:4854–65.
- 19 Yoshida K, Chan P, Marchand M, *et al.* Tumor Growth Inhibition-Overall Survival (TGI-OS) model for subgroup analysis based on post-randomization factors: application for Anti-drug Antibody (ADA) subgroup analysis of atezolizumab in the IMPower150 study. *AAPS J* 2022;24:58.
- 20 Wages NA, Chiuzan C, Panageas KS. Design considerations for early-phase clinical trials of immune-oncology agents. *J Immunother Cancer* 2018;6:81.
- 21 Goldberg KB, Blumenthal GM, McKee AE, *et al.* The FDA oncology center of excellence and precision medicine. *Exp Biol Med (Maywood)* 2018;243:308–12.
- 22 Lam C, Meinert E, Yang A, *et al.* Impact of fast-track regulatory designations on strategic commercialization decisions for autologous cell therapies. *Regen Med* 2022;17:155–74.
- 23 Jardim DL, de Melo Gagliato D, Giles FJ, *et al.* Analysis of drug development paradigms for immune checkpoint inhibitors. *Clin Cancer Res* 2018;24:1785–94.
- 24 Chaudhari M, Crowley J, Hoering A. A comparative study on the SWOG two-stage design extension to stop early for efficacy in single arm Phase II trials. *Stat Biopharm Res* 2022;14:542–51.
- 25 Han Y, Liu H, Cao S, *et al.* TSNP: a two-stage nonparametric phase I/II clinical trial design for immunotherapy. *Pharm Stat* 2021;20:282–96.
- 26 Guo B, Garrett-Mayer E, Liu S. A Bayesian Phase I/II design for cancer clinical trials combining an immunotherapeutic agent with a chemotherapeutic agent. *J R Stat Soc Ser C Appl Stat* 2021;70:1210–29.
- 27 Bruno R, Bottino D, de Alwis DP, *et al.* Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models. *Clin Cancer Res* 2020;26:1787–95.
- 28 Al-Huniti N, Feng Y, Yu JJ, *et al.* Tumor growth dynamic modeling in oncology drug development and regulatory approval: past, present, and future opportunities. *CPT Pharmacometrics Syst Pharmacol* 2020;9:419–27.
- 29 Hoering A, Durie B, Wang H, *et al.* End points and statistical considerations in immuno-oncology trials: impact on multiple myeloma. *Future Oncol* 2017;13:1181–93.
- 30 Twomey JD, Zhang B. Cancer immunotherapy update: FDA-approved checkpoint inhibitors and companion diagnostics. *AAPS J* 2021;23:39.
- 31 Kareff SA, Samtani S, Burotto M, *et al.* Current landscape of immunotherapy trials involving the programmed cell death protein 1/programmed death-ligand 1 axis in intrathoracic tumors. *JTO Clin Res Rep* 2021;2:100149.
- 32 SWOG. SWOG Cancer Research Network. 2024. Available: [www.swog.org](http://www.swog.org)
- 33 Friedman CF, Hainsworth JD, Kurzrock R, *et al.* Atezolizumab treatment of tumors with high tumor mutational burden from my pathway, a multicenter, open-label, phase IIA multiple basket study. *Cancer Discov* 2022;12:654–69.
- 34 Darragh LB, Knitz MM, Hu J, *et al.* A phase I/II trial and biological correlate analysis of neoadjuvant SBRT with single-dose durvalumab in HPV-unrelated locally advanced HNSCC. *Nat Cancer* 2022;3:1300–17.
- 35 Oldenhuis CNAM, Oosting SF, Gietema JA, *et al.* Prognostic versus predictive value of biomarkers in oncology. *Eur J Cancer* 2008;44:946–53.
- 36 Jensen MA, Ferretti V, Grossman RL, *et al.* The NCI genomic data commons as an engine for precision medicine. *Blood* 2017;130:453–9.
- 37 Fedorov A, Longabaugh WJR, Pot D, *et al.* NCI imaging data commons. *Cancer Res* 2021;81:4188–93.
- 38 NCTN/NCORP Data Archive, 2024. Available: <https://nctn-data-archive.nci.nih.gov/>
- 39 Keenan TE, Burke KP, Van Allen EM. Genomic correlates of response to immune checkpoint blockade. *Nat Med* 2019;25:389–402.
- 40 National-Cancer-Institute. National Clinical Trials Network (NCTN) Navigator, 2024. Available: <https://navigator.ctsu.org/navigator/>
- 41 National-Cancer-Institute. Informatics Technology for Cancer Research, 2024. Available: <https://www.cancer.gov/about-nci/organization/ctss/research/itcr>
- 42 Sadybekov AV, Katritch V. Computational approaches streamlining drug discovery. *Nature New Biol* 2023;616:673–85.
- 43 Cai Z, Poulos RC, Liu J, *et al.* Machine learning for multi-omics data integration in cancer. *i Sci* 2022;25:103798.
- 44 Bender BJ, Gahbauer S, Luttens A, *et al.* A practical guide to large-scale docking. *Nat Protoc* 2021;16:4799–832.
- 45 Parvathaneni V, Kulkarni NS, Muth A, *et al.* Drug repurposing: a promising tool to accelerate the drug discovery process. *Drug Discov Today* 2019;24:2076–85.
- 46 You Y, Lai X, Pan Y, *et al.* Artificial intelligence in cancer target identification and drug discovery. *Signal Transduct Target Ther* 2022;7:156.
- 47 Schneider BJ, Naidoo J, Santomaso BD, *et al.* Management of immune-related adverse events in patients treated with immune checkpoint inhibitor therapy: ASCO guideline update. *J Clin Oncol* 2021;39:4073–126.
- 48 Yuan S, Wen C, *et al.* Immune checkpoint inhibitor-related chronic pneumonitis: a case report and literature review. *Immunother (Los Angel)* 2023;15:1117–23.
- 49 Adashek JJ, Subbiah IM, Matos I, *et al.* Hyperprogression and immunotherapy: fact, fiction, or alternative fact? *Trends Cancer* 2020;6:181–91.
- 50 Colen RR, Fujii T, Bilen MA, *et al.* Radiomics to predict immunotherapy-induced pneumonitis: proof of concept. *Invest New Drugs* 2018;36:601–7.
- 51 Huang J, Chen X, Xia B, *et al.* Advances in CT features and radiomics of checkpoint inhibitor-related pneumonitis: a short review. *Front Immunol* 2023;14:1082980.
- 52 Chen M, Lu H, Copley SJ, *et al.* A novel radiogenomics biomarker for predicting treatment response and pneumotoxicity from programmed cell death protein or ligand-1 inhibition immunotherapy in NSCLC. *J Thorac Oncol* 2023;18:718–30.
- 53 Qiu Q, Xing L, Wang Y, *et al.* Development and validation of a radiomics nomogram using computed tomography for differentiating immune checkpoint inhibitor-related pneumonitis from radiation pneumonitis for patients with non-small cell lung cancer. *Front Immunol* 2022;13:870842.
- 54 Cheng J, Pan Y, Huang W, *et al.* Differentiation between immune checkpoint inhibitor-related and radiation pneumonitis in lung cancer by CT radiomics and machine learning. *Med Phys Mex Symp Med Phys* 2022;49:1547–58.
- 55 Vaidya P, Bera K, Patil PD, *et al.* Novel, non-invasive imaging approach to identify patients with advanced non-small cell lung cancer at risk of hyperprogressive disease with immune checkpoint blockade. *J Immunother Cancer* 2020;8:e001343.
- 56 Tunali I, Gray JE, Qi J, *et al.* Novel clinical and radiomic predictors of rapid disease progression phenotypes among lung cancer patients treated with immunotherapy: an early report. *Lung Cancer (Auckl)* 2019;129:75–9.
- 57 Fadlullah MZH, Lin C-N, Coleman S, *et al.* Exploring the landscape of immune checkpoint inhibitor-induced adverse events through big data mining of pan-cancer clinical trials. *Oncologist* 2024;29:415–21.
- 58 Kerepesi C, Abushukair HM, Ricciuti B, *et al.* Association of baseline tumor-specific neoantigens and CD8<sup>+</sup> T-cell infiltration with immune-related adverse events secondary to immune checkpoint inhibitors. *JCO Precis Oncol* 2024;8:e2300439.



- 59 Patel MI, Lopez AM, Blackstock W, *et al.* Cancer disparities and health equity: a policy statement from the American society of clinical oncology. *J Clin Oncol* 2020;38:3439–48.
- 60 Hoskins KF, Danciu OC, Ko NY, *et al.* Association of race/ethnicity and the 21-gene recurrence score with breast cancer-specific mortality among US women. *JAMA Oncol* 2021;7:370–8.
- 61 Vaidya A, Chen RJ, Williamson DFK, *et al.* Demographic bias in misdiagnosis by computational pathology models. *Nat Med* 2024;30:1174–90.
- 62 Bhargava HK, Leo P, Elliott R, *et al.* Computationally derived image signature of stromal morphology is prognostic of prostate cancer recurrence following prostatectomy in african american patients. *Clin Cancer Res* 2020;26:1915–23.
- 63 Chang FU, Chien CR. The importance of out-of-pocket cost information in the era of financial toxicity of cancer care. *Support Care Cancer* 2021;29:1149.
- 64 McLouth LE, Nightingale CL, Levine BJ, *et al.* Unmet care needs and financial hardship in patients with metastatic non-small-cell lung cancer on immunotherapy or chemoimmunotherapy in clinical practice. *JCO Oncol Pract* 2021;17:e1110–9.
- 65 Alilou M, Khorrami M, Prasanna P, *et al.* A tumor vasculature-based imaging biomarker for predicting response and survival in patients with lung cancer treated with checkpoint inhibitors. *Sci Adv* 2022;8:eabq4609.
- 66 Wang X, Barrera C, Bera K, *et al.* Spatial interplay patterns of cancer nuclei and tumor-infiltrating lymphocytes (TILs) predict clinical benefit for immune checkpoint inhibitors. *Sci Adv* 2022;8.
- 67 Hiremath A, Corredor G, Li L, *et al.* An integrated radiology-pathology machine learning classifier for outcome prediction following radical prostatectomy: Preliminary findings. *Heliyon* 2024;10:e29602.
- 68 National Academies of Sciences. Changing the Culture of Data Management and Sharing: A Workshop, 2021. Available: <https://www.nationalacademies.org/our-work/changing-the-culture-of-data-management-and-sharing-a-workshop>
- 69 Litchfield K, Reading JL, Puttick C, *et al.* Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* 2021;184:596–614.
- 70 National-Institute-of-Health. CDM and FHIR Harmonization Resources, 2024. Available: <https://datascience.nih.gov/fhir-initiatives/resources/cdmcatalog>
- 71 Post AR, Burningham Z, Halwani AS. Electronic health record data in cancer learning health systems: challenges and opportunities. *JCO Clin Cancer Inform* 2022;6:e2100158.
- 72 Rajkomar A, Oren E, Chen K, *et al.* Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018;1:18.
- 73 Yang X, Chen A, PourNejati N, *et al.* A large language model for electronic health records. *NPJ Digit Med* 2022;5:194.
- 74 Keloth VK, Selek S, Chen Q, *et al.* Large language models for social determinants of health information extraction from clinical notes - a generalizable approach across institutions. *medRxiv* 2024;2024:2024.
- 75 Miao D, Margolis CA, Vokes NI, *et al.* Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nat Genet* 2018;50:1271–81.
- 76 Banchereau R, Leng N, Zill O, *et al.* Molecular determinants of response to PD-L1 blockade across tumor types. *Nat Commun* 2021;12:3969.
- 77 Wells DK, van Buuren MM, Dang KK, *et al.* Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell* 2020;183:818–34.
- 78 Pelka K, Hofree M, Chen JH, *et al.* Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* 2021;184:4734–52.
- 79 Andreatta M, Corria-Osorio J, Müller S, *et al.* Interpretation of T cell states from single-cell transcriptomics data using reference atlases. *Nat Commun* 2021;12:2965.
- 80 Sidiropoulos DN, Stein-O'Brien GL, Danilova L, *et al.* Integrated T cell cytometry metrics for immune-monitoring applications in immunotherapy clinical trials. *JCI Insight* 2022;7:e160398.
- 81 Chen RJ, Lu MY, Wang J, *et al.* Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. *IEEE Trans Med Imaging* 2022;41:757–70.
- 82 Chen RJ, Lu MY, Williamson DFK, *et al.* Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell* 2022;40:865–78.
- 83 Bottomly D, McWeeney S. Just how transformative will AI/ML be for immuno-oncology? *J Immunother Cancer* 2024;12:e007841.
- 84 Yamaguchi H, Hsu J-M, Sun L, *et al.* Advances and prospects of biomarkers for immune checkpoint inhibitors. *Cell Rep Med* 2024;5:101621.
- 85 Hu J, Han C, Zhong J, *et al.* Dynamic Network biomarker of pre-exhausted CD8+ T cells contributed to T cell exhaustion in colorectal cancer. *Front Immunol* 2021;12:691142.
- 86 Liu X, Chang X, Leng S, *et al.* Detection for disease tipping points by landscape dynamic network biomarkers. *Natl Sci Rev* 2019;6:775–85.
- 87 Creemers JHA, Lesterhuis WJ, Mehra N, *et al.* A tipping point in cancer-immune dynamics leads to divergent immunotherapy responses and hampers biomarker discovery. *J Immunother Cancer* 2021;9:e002032.
- 88 Jazieh K, Khorrami M, Saad A, *et al.* Novel imaging biomarkers predict outcomes in stage III unresectable non-small cell lung cancer treated with chemoradiation and durvalumab. *J Immunother Cancer* 2022;10:e003778.
- 89 Acs B, Ahmed FS, Gupta S, *et al.* An open source automated tumor infiltrating lymphocyte algorithm for prognosis in melanoma. *Nat Commun* 2019;10:5440.
- 90 Fang C, Xu D, Su J, *et al.* DeePaN: deep patient graph convolutional network integrating clinico-genomic evidence to stratify lung cancers for immunotherapy. *NPJ Digit Med* 2021;4:14.
- 91 Truhn D, Eckardt JN, Ferber D, *et al.* Large language models and multimodal foundation models for precision oncology. *NPJ Precis Oncol* 2024;8:72.
- 92 Cui H, Wang C, Maan H, *et al.* scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nat Methods* 2024;21:1470–80.
- 93 Mirza FN, Tang OY, Connolly ID, *et al.* Using ChatGPT to facilitate truly informed medical consent. *NEJM AI* 2024;1:Alcs2300145.
- 94 Lu MY, Chen B, Williamson DFK, *et al.* A visual-language foundation model for computational pathology. *Nat Med* 2024;30:863–74.
- 95 Chen RJ, Ding T, Lu MY, *et al.* Towards a general-purpose foundation model for computational pathology. *Nat Med* 2024;30:850–62.
- 96 Black S, Phillips D, Hickey JW, *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat Protoc* 2021;16:3802–35.
- 97 Lin J-R, Izar B, Wang S, *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *Elife* 2018;7.
- 98 Moldoveanu D, Ramsay L, Lajoie M, *et al.* Spatially mapping the immune landscape of melanoma using imaging mass cytometry. *Sci Immunol* 2022;7:eabi5072.
- 99 Schürch CM, Bhat SS, Barlow GL, *et al.* Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* 2020;182:1341–59.
- 100 Nerretter T, Letschert S, Götz R, *et al.* Super-resolution microscopy reveals ultra-low CD19 expression on myeloma cells that triggers elimination by CD19 CAR-T. *Nat Commun* 2019;10:3137.
- 101 Kiemen AL, Braxton AM, Grahn MP, *et al.* CODA: quantitative 3D reconstruction of large tissues at cellular resolution. *Nat Methods* 2022;19:1490–9.
- 102 He S, Bhatt R, Brown C, *et al.* High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nat Biotechnol* 2022;40:1794–806.
- 103 Quadrato G, Nguyen T, Macosko EZ, *et al.* Cell diversity and network dynamics in photosensitive human brain organoids. *Nature New Biol* 2017;545:48–53.
- 104 Medina J, Jax SA, Coslett HB. Impairments in action and perception after right intraparietal damage. *Cor* 2020;122:288–99.
- 105 Zhang D, Schroeder A, Yan H, *et al.* Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nat Biotechnol* 2024;42:1372–7.
- 106 Waylen LN, Nim HT, Martelotto LG, *et al.* From whole-mount to single-cell spatial assessment of gene expression in 3D. *Commun Biol* 2020;3:602.
- 107 Schott M, León-Perián D, Splendiani E, *et al.* n.d. Open-ST: High-resolution spatial transcriptomics in 3D. *bioRxiv* 2023:2023.
- 108 Helmink BA, Reddy SM, Gao J, *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature New Biol* 2020;577:549–55.
- 109 Cabrita R, Lauss M, Sanna A, *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature New Biol* 2020;577:561–5.
- 110 Erickson A, He M, Berglund E, *et al.* Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature New Biol* 2022;608:360–7.
- 111 Zhang S, Yuan L, Danilova L, *et al.* Spatial transcriptomics analysis of neoadjuvant cabozantinib and nivolumab in advanced hepatocellular carcinoma identifies independent mechanisms of resistance and recurrence. *Genome Med* 2023;15:72.



- 112 Erbe R, Gore J, Gemmill K, *et al.* The use of machine learning to discover regulatory networks controlling biological systems. *Mol Cell* 2022;82:260–73.
- 113 Boschman J, Farahani H, Darbandsari A, *et al.* The utility of color normalization for AI-based diagnosis of hematoxylin and eosin-stained pathology images. *J Pathol* 2022;256:15–24.
- 114 Ding R, Prasanna P, Corredor G, *et al.* Image analysis reveals molecularly distinct patterns of TILs in NSCLC associated with treatment outcome. *NPJ Precis Oncol* 2022;6:33.
- 115 Pan Y, Kadash-Edmondson KE, Wang R, *et al.* RNA dysregulation: an expanding source of cancer immunotherapy targets. *Trends Pharmacol Sci* 2021;42:268–82.
- 116 Park E, Pan Z, Zhang Z, *et al.* The expanding landscape of alternative splicing variation in human populations. *Am J Hum Genet* 2018;102:11–26.
- 117 Byrne A, Cole C, Volden R, *et al.* Realizing the potential of full-length transcriptome sequencing. *Philos Trans R Soc Lond B Biol Sci* 2019;374:20190097.
- 118 Chan TA, Wolchok JD, Snyder A. Genetic basis for clinical response to CTLA-4 blockade in Melanoma. *N Engl J Med* 2015;373:1984.
- 119 Joglekar AV, Leonard MT, Jeppson JD, *et al.* T cell antigen discovery via signaling and antigen-presenting bifunctional receptors. *Nat Methods* 2019;16:191–8.
- 120 Li G, Bethune MT, Wong S, *et al.* T cell antigen discovery via trogocytosis. *Nat Methods* 2019;16:183–90.
- 121 Danilova L, Anagnostou V, Caushi JX, *et al.* The mutation-associated neoantigen functional expansion of specific T cells (manafest) assay: a sensitive platform for monitoring antitumor immunity. *Cancer Immunol Res* 2018;6:888–99.
- 122 Caushi JX, Zhang J, Ji Z, *et al.* Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. *Nature New Biol* 2021;596:126–32.
- 123 Oliveira G, Stromhaug K, Klaeger S, *et al.* Phenotype, specificity and avidity of antitumour CD8<sup>+</sup> T cells in melanoma. *Nature New Biol* 2021;596:119–25.
- 124 Oliveira G, Stromhaug K, Cieri N, *et al.* Landscape of helper and regulatory antitumour CD4<sup>+</sup> T cells in melanoma. *Nature New Biol* 2022;605:532–8.
- 125 Lowery FJ, Krishna S, Yossef R, *et al.* Molecular signatures of antitumor neoantigen-reactive T cells from metastatic human cancers. *Science* 2022;375:877–84.
- 126 Simoni Y, Becht E, Fehlings M, *et al.* Bystander CD8<sup>+</sup> T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature New Biol* 2018;557:575–9.
- 127 Wu TD, Madireddi S, de Almeida PE, *et al.* Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature New Biol* 2020;579:274–8.
- 128 Drost F, An Y, Bonafonte-Pardàs I, *et al.* Multi-modal generative modeling for joint analysis of single-cell T cell receptor and gene expression data. *Nat Commun* 2024;15:5577.
- 129 Schattgen SA, Guion K, Crawford JC, *et al.* Integrating T cell receptor sequences and transcriptional profiles by clonotype neighbor graph analysis (CoNGA). *Nat Biotechnol* 2022;40:54–63.
- 130 Slyper M, Porter CBM, Ashenberg O, *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* 2020;26:792–802.
- 131 Jumper J, Evans R, Pritzel A, *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature New Biol* 2021;596:583–9.
- 132 Arunkumar M, Zielinski CE. T-cell receptor repertoire analysis with computational tools—an immunologist's perspective. *Cells* 2021;10:3582.
- 133 Dash P, Fiore-Gartland AJ, Hertz T, *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature New Biol* 2017;547:89–93.
- 134 Chiou S-H, Tseng D, Reuben A, *et al.* Global analysis of shared T cell specificities in human non-small cell lung cancer enables HLA inference and antigen discovery. *Immunity* 2021;54:586–602.
- 135 Lu T, Zhang Z, Zhu J, *et al.* Deep learning-based prediction of the T cell receptor-antigen binding specificity. *Nat Mach Intell* 2021;3:864–75.
- 136 Zheng L, Qin S, Si W, *et al.* Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* 2021;374.
- 137 Ma K-Y, Schonnesen AA, He C, *et al.* High-throughput and high-dimensional single-cell analysis of antigen-specific CD8<sup>+</sup> T cells. *Nat Immunol* 2021;22:1590–8.
- 138 Gastman B, Agarwal PK, Berger A, *et al.* Defining best practices for tissue procurement in immuno-oncology clinical trials: consensus statement from the Society for immunotherapy of cancer surgery committee. *J Immunother Cancer* 2020;8:e001583.
- 139 Lee JC, Green MD, Huppert LA, *et al.* The liver–immunity nexus and cancer immunotherapy. *Clin Cancer Res* 2022;28:5–12.
- 140 Chen HX, Song M, Maecker HT, *et al.* Network for biomarker immunoprofiling for cancer immunotherapy: Cancer Immune Monitoring and Analysis Centers and Cancer Immunologic Data Commons (CIMAC-CIDC). *Clin Cancer Res* 2021;27:5038–48.
- 141 National-Cancer-Institute. Cancer Research Data Commons (CRDC), 2024. Available: <https://datascience.cancer.gov/data-commons>
- 142 National-Cancer-Institute. NIH Data Sharing Policy and Implementation Guidance, 2003. Available: <https://sharing.nih.gov/data-management-and-sharing-policy>
- 143 National-Cancer-Institute. NIH Data Sharing Policies, 2019. Available: <https://sharing.nih.gov/other-sharing-policies/nih-institute-and-center-data-sharing-policies>
- 144 National-Cancer-Institute. Final NIH Policy for Data Management and Sharing, 2020. Available: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>
- 145 National-Cancer-Institute. HHS Regulations for the Protection of Human Subjects, 2022. Available: <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html#46.102>
- 146 Li H, Bera K, Toro P, *et al.* Collagen fiber orientation disorder from H&E images is prognostic for early stage breast cancer: clinical trial validation. *NPJ Breast Cancer* 2021;7:104.
- 147 Esteva A, Feng J, van der Wal D, *et al.* Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *NPJ Digit Med* 2022;5:71.
- 148 Di Tommaso P, Chatzou M, Floden EW, *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol* 2017;35:316–9.
- 149 Mölder F, Jablonski KP, Letcher B, *et al.* Sustainable data analysis with snakemake. *F1000Res* 2021;10:33.
- 150 Google. Google Colaboratory, 2024. Available: <https://colab.research.google.com>
- 151 Beaulieu-Jones BK, Greene CS. Reproducibility of computational workflows is automated using continuous analysis. *Nat Biotechnol* 2017;35:342–6.
- 152 Huber W, Carey VJ, Gentleman R, *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 2015;12:115–21.
- 153 Jiang P, Sinha S, Aldape K, *et al.* Big data in basic and translational cancer research. *Nat Rev Cancer* 2022;22:625–39.
- 154 Mulder N, Schwartz R, Brazas MD, *et al.* The development and application of bioinformatics core competencies to improve bioinformatics training and education. *PLoS Comput Biol* 2018;14:e1005772.
- 155 Ranganathan S. Bioinformatics education—perspectives and challenges. *PLoS Comput Biol* 2005;1:e52.
- 156 Hack C, Kendall G. Bioinformatics: Current practice and future challenges for life science education. *Biochem Mol Biol Educ* 2005;33:82–5.
- 157 SITC-Big-Data-and-Data-Sharing-Committee. AI in IO: SITC-NCI Computational Immuno-Oncology Webinar Series. Society for Immunotherapy of Cancer, National Cancer Institute, and Roswell Park Comprehensive Cancer Center; 2024. Available: <https://www.sitcancer.org/education/webinars-online/computational>
- 158 Vandereyken K, Sifrim A, Thienpont B, *et al.* Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* 2023;24:494–515.
- 159 Elmarakeby HA, Hwang J, Arafeh R, *et al.* Biologically informed deep neural network for prostate cancer discovery. *Nat New Biol* 2021;598:348–52.
- 160 Monastersky R, Van Noorden R. 150 years of nature: a data graphic charts our evolution. *Nat New Biol* 2019;575:22–3.
- 161 Annapragada A, Sikora A, Bollard C, *et al.* Cancer moonshot Immuno-Oncology Translational Network (IOTN): accelerating the clinical translation of basic discoveries for improving immunotherapy and immunoprevention of cancer. *J Immunother Cancer* 2020;8:e000796.
- 162 National-Institute-of-Health. NIH Data Management and Sharing Policy, 2023. Available: <https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/2023-nih-data-management-sharing-policy>
- 163 Mutha P, Duan F, Viswanathan VS, *et al.* 175P radiomic biomarker of vessel tortuosity for monitoring treatment change: preliminary findings in prospective evaluation of ECOG-ACRIN EA5163. *Ann Oncol* 2023;34:S251.