nature reviews clinical oncology

https://doi.org/10.1038/s41571-023-00830-6

Review article

Check for updates

# Computational immunogenomic approaches to predict response to cancer immunotherapies

Venkateswar Addala[1,2] ✉, Felicity Newell[1], John V. Pearson[1], Alec Redwood[3,4,5], Bruce W. Robinson[3,4,6,7], Jenette Creaney[3,4,5,6] & Nicola Waddell[1,2] ✉

## Abstract

Cancer immunogenomics is an emerging field that bridges genomics and immunology. The establishment of large-scale genomic collaborative efforts along with the development of new single-cell transcriptomic techniques and multi-omics approaches have enabled characterization of the mutational and transcriptional profiles of many cancer types and helped to identify clinically actionable alterations as well as predictive and prognostic biomarkers. Researchers have developed computational approaches and machine learning algorithms to accurately obtain clinically useful information from genomic and transcriptomic sequencing data from bulk tissue or single cells and explore tumours and their microenvironment. The rapid growth in sequencing and computational approaches has resulted in the unmet need to understand their true potential and limitations in enabling improvements in the management of patients with cancer who are receiving immunotherapies. In this Review, we describe the computational approaches currently available to analyse bulk tissue and single-cell sequencing data from cancer, stromal and immune cells, as well as how best to select the most appropriate tool to address various clinical questions and, ultimately, improve patient outcomes.

## Sections



[1]Cancer Program, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [2]Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia. [3]National Centre for Asbestos Related Diseases, University of Western Australia, Perth, Western Australia, Australia. [4]Institute of Respiratory Health, Perth, Western Australia, Australia. [5]School of Biomedical Science, University of Western Australia, Perth, Western Australia, Australia. [6]Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Perth, Western Australia, Australia. [7]Medical School, University of Western Australia, Perth, Western Australia, Australia. ✉e-mail: venkateswar.addala@qimrberghofer.edu.au; nic.waddell@qimrberghofer.edu.au

## Key points

- Researchers are developing various immunogenomic tools to predict response to treatment in patients with cancer who are receiving immune-checkpoint inhibitors (ICIs), based on cancer-intrinsic and cancer-extrinsic features that can be identified with sequencing, including tumour mutational burden, neoantigens and the presence of immune cells.
- Computational tools for HLA genotyping from whole-genome sequencing, whole-exome sequencing and RNA sequencing have been well established; long-read sequencing is a promising technology that is expected to improve the performance of HLA genotyping.
- Several approaches have been developed to identify immunogenic neoantigens, with a major focus on somatic single-nucleotide variants; however, the identification of neoantigens from non-canonical sources is crucial for a comprehensive understanding of neoantigen load.
- Deconvolution tools provide estimates of the immune cell proportions in the tumour microenvironment but have limitations in identifying low-abundance cell types and subsets; therefore, the use of these tools requires careful consideration of the underlying technical and biological factors.
- Multi-omic machine learning models trained on molecular and clinical features from large cohorts of tumour samples could improve the prediction of patient responses to immunotherapy and reveal key predictive features.
- Functionally verified approaches that integrate genomic intratumour heterogeneity, HLA genotypes and neoantigen trafficking, and expression and immunogenicity, among other features, could improve prediction of response to ICIs.

## Introduction

The identification of somatic alterations that result in pharmacologically targetable modifications in patients with cancer has provided the potential to improve survival in some cancer types. Nevertheless, not all patients have actionable alterations and those who do often have disease relapse and/or develop treatment resistance[1]. Therapies that harness the immune system emerged as a promising new option; one such approach involves the use of immune-checkpoint inhibitors (ICIs), which frequently target the PD-1–PD-L1 axis or CTLA4. The immune checkpoints PD-1 and CTLA4 inhibit T cell activation and thus, the goal of blocking their interactions with ligands is to unleash anti-tumour immunity[2]. ICIs have become a first-line treatment option in several tumour types[3–17]; however, patients can still have relapse and/or develop treatment resistance[2]. Therefore, determining which patients are likely to benefit from ICIs remains one of the major challenges in immunotherapy. In some tumour types, tumour expression of PD-L1 is used as a companion biomarker to select patients to receive ICIs. However, PD-L1 is an imperfect biomarker[18] and poses challenges that need to be addressed, including establishing thresholds for positivity and addressing whether optimum thresholds for ICI response differ between tumour types and/or with different PD-L1 antibodies[19]. These challenges were highlighted in a summary of PD-L1 reporting from 11 phase III trials that illustrated differences in scoring approaches and primary end point cut-offs[20]. Tumour mutational burden (TMB), or the number of somatic mutations per megabase (Mb) in a cancer cell, is another proposed predictive biomarker. Patients with a high TMB (>10 mutations per megabase (mut/Mb)) are more likely to have a response to ICIs[21,22] and thus, TMB was approved by the FDA as a biomarker to select patients eligible to receive pembrolizumab[22,23]. Specific gene expression profile (GEP) signatures linked to response to ICIs have also been described and proposed as biomarkers[24,25]. Nevertheless, an analysis of samples from more than 300 patients enrolled in one of four clinical trials of the anti-PD-1 antibody pembrolizumab revealed that TMB, PD-L1 status or GEP signatures alone have only a modest correlation with response to ICIs in 22 tumour types; these characteristics were defined as independent predictive biomarkers[26]

The lack of robust predictive biomarkers of response to ICIs has sparked the development of computational immunogenomic methods to determine which patients are likely to benefit from these agents and to tailor ICI combinations. Many cancer-intrinsic (within cancer cells) and cancer-extrinsic (related to the tumour microenvironment (TME)) features that affect response to ICIs can be explored using computational immunogenomic approaches to analyse DNA sequencing (DNA-seq) and RNA sequencing (RNA-seq) data from bulk tissues and single cells (Fig. 1).

In this Review, we discuss computational approaches developed to determine, among others, TMB and HLA genotypes, and to estimate mutation-induced neoantigen load, the cellular composition of the TME, and GEP signatures. We highlight the reliability and limitations of these computational approaches in the context of the cancer-intrinsic and cancer-extrinsic features that can be measured in omics data from patients receiving ICIs in clinical trials[27–32].

## Analysis of cancer-intrinsic features

### Tumour purity and genomic ITH

Tumour purity is the proportion or percentage of cancer cells within a tumour tissue sample[33], and thus it affects the performance of computational approaches used to profile cancer-intrinsic and cancer-extrinsic immunogenomic features. Various computational approaches have been developed to measure tumour purity from DNA-seq[34–36] and RNA-seq[37,38], or DNA methylation profiling[33] data, as well as integrated consensus approaches[33] (Fig. 1), and methods generally show good concordance with each other[33].

Genomic intratumour heterogeneity (ITH) occurs in most tumours and refers to the clonal architecture of tumour cells and genomic differences that can exist between individual subclones[39]. ITH is linked to treatment resistance, recurrence and reduced patient survival[40,41]. A large number of computational tools have been developed to estimate genomic ITH from genome sequencing data including, among others, Pyclone[42], SciClone[43], Battenberg[44], HATCHet[45] and FastClone[46]. Several studies have compared the performance of these and other algorithms[47,48], and found that their performance and accuracy can be variable and improve with greater sequencing depth. Sequencing of multiple samples from each tumour to capture spatial and temporal changes enables a better understanding of ITH and tumour evolution[39,41,49]. Some tools, such as HATCHet[45] (which infers subclonal copy number changes and whole-genome duplication), have been developed to analyse multiple samples from the same tumour.

### Tumour mutational burden

Somatic genomic alterations that occur in cancer cells include single-nucleotide variants (SNVs), multinucleotide variants, small

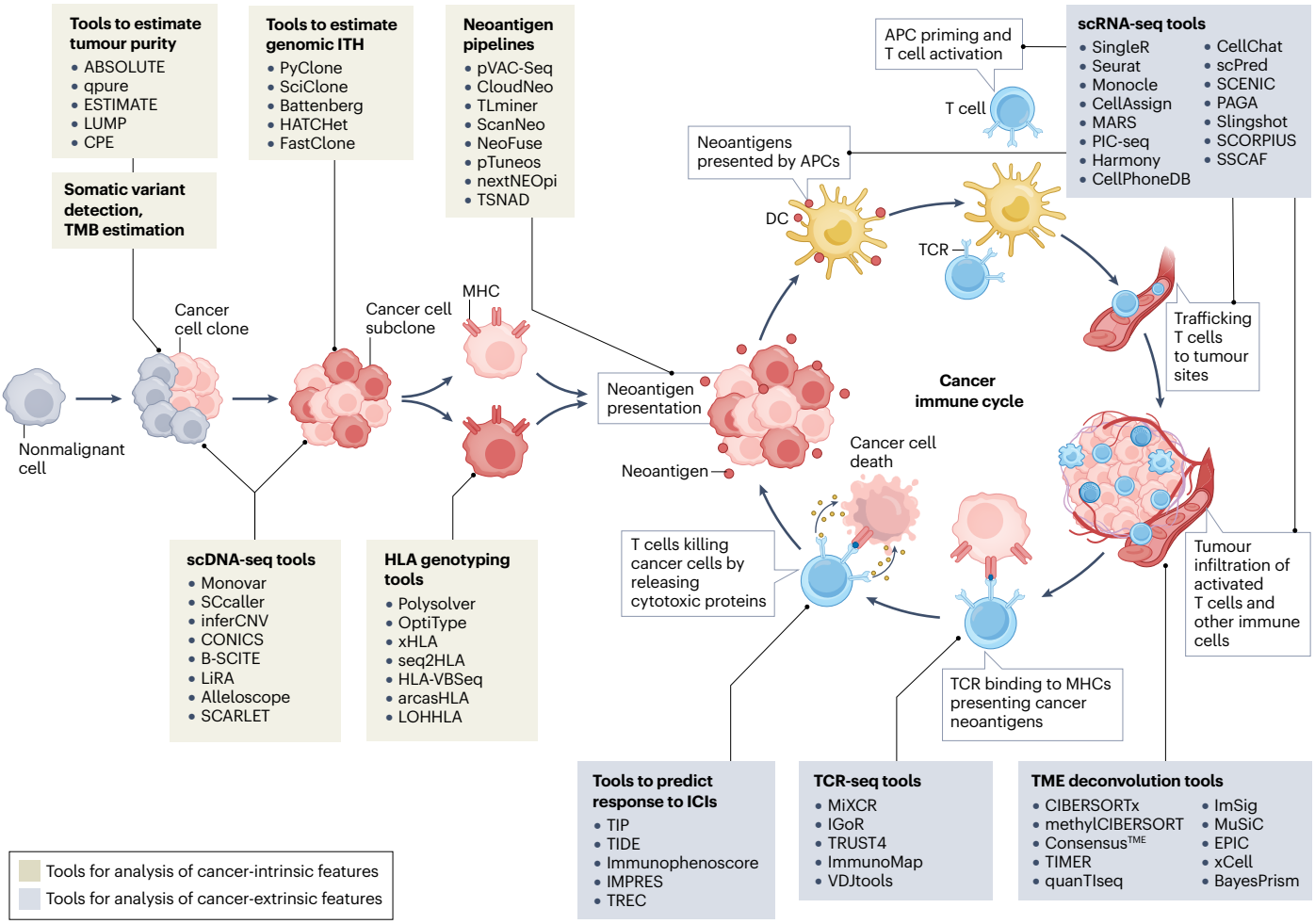



**Fig. 1 | Computational approaches to interrogate cancer-intrinsic and cancer-extrinsic immune phenotypes in bulk tissue samples.** Researchers have developed various computational tools for the analysis of data from tumour DNA and RNA sequencing. Immunogenomic research has used these approaches to characterize cancer immune features and predict patient response to immune-checkpoint inhibitors (ICIs). These computational approaches can be stratified into those designed to study cancer-intrinsic features or cancer-extrinsic features, and single-cell approaches. Some of these tools intersect with steps of the cancer immune cycle (shown here as a seven-step cycle). This figure provides an overview of the tools that have been developed thus far. APC, antigen-presenting cell; DC, dendritic cell; ITH, intratumour heterogeneity; MHC, major histocompatibility complex; scRNA-seq, single-cell RNA sequencing; scDNA-seq, single-cell DNA sequencing; TCR, T cell receptor; TMB, tumour mutational burden; TME, tumour microenvironment.

insertions and deletions (indels), copy number alterations (CNAs) and large chromosome structural variants. TMB does not reflect all somatic alterations: most panels commercially available for its assessment typically consider the number of somatic non-synonymous (that is, affecting protein sequences) with or without synonymous (that is, not affecting protein sequence) SNVs, and in some cases indels, within a tumour sample expressed as mut/Mb[50]. A high TMB has been associated with responsiveness to ICIs[21,22,29,51–54]; for example, cutaneous melanomas tend to have a high TMB owing to the powerful mutagenic effects of UV radiation[55] and patients with this malignancy often have good responses to ICIs[24,56,57]. An association of favourable response with high TMB was reported in the KEYNOTE-158 trial, in which 29% of patients with a high TMB (>10 mut/Mb) had objective responses compared with 6% with a low TMB[22].

Although a simple concept, the estimation of TMB has numerous caveats linked to cancer type, sample preparation, sequencing method, bioinformatic approaches to call mutations and the selection of mutations to calculate TMB. Cancer types can have different ranges of SNV numbers[58], and thus thresholds for a high TMB can differ between cancer types[59]. Samples can be prepared in various states, including fresh frozen or formalin-fixed paraffin-embedded (FFPE) tissues, which will affect the DNA yield and quality, and thus subsequent sequence quality and variant detection. Tumour purity and genomic ITH affect the proportion of DNA that contains one or more particular mutations but are not consistent between samples, thus affecting mutation detection[60]. The depth of sequencing and selection of whole-genome sequencing (WGS), whole-exome sequencing (WES), comprehensive gene panels or targeted panel sequencing approaches can also affect mutation detection[61], as somatic alterations are not distributed evenly within cancer cell genomes[62]. A large interlaboratory comparison of bioinformatic approaches for sequencing and data filtering showed marked differences in variant detection[63]. Subsequently, another group

assessed the effect of experimental and analytical elements on somatic variant detection and formulated best-practice recommendations to improve the reproducibility and accuracy of mutation detection from WGS and WES data[61]. Furthermore, whether the DNA sequenced is from paired cancer and nonmalignant cell samples (which enables bioinformatic subtraction of germline variants) or from cancer cells (with cancer-specific mutations determined as the proportion of sequence reads containing the variant or the absence of detected variants within population databases, such as GnomAD[64]), affects the sensitivity and accuracy of somatic mutation detection. Consequently, when sequencing cancer samples from individuals whose constitutional variation is typically under-represented in population databases, such as African, Asiatic or First Nations individuals[65], researchers should consider sequencing a matched nonmalignant sample.

Multiple approaches are available to sequence samples or select mutations[50]. For example, researchers determined TMB in a clinical dataset from 77 patients with melanoma using WGS to count all somatic mutations (SNVs and indels)[57]. Another group calculated TMB in samples from 312 patients with non-small-cell lung cancer (NSCLC) from somatic missense mutations in the coding region identified by WES[66]. Other researchers calculated TMB from somatic non-synonymous mutations within the coding region identified by sequencing of a gene panel, including in 1,552 patients with NSCLC[67] and 403 patients with advanced-stage cancer ($n = 7$ different cancer types)[68]. Although the TMB estimates calculated using these approaches might correlate, variations in absolute value for a given tumour sample could result in the tumour not being consistently classified as above or below a defined threshold[69,70], which is an important consideration when using TMB to determine eligibility for ICIs. The FDA approval of pembrolizumab for patients with solid tumours with a high TMB includes the use of the FoundationOne CDx assay (which uses FFPE tumour samples) as a companion diagnostic tool[23]. TMB is a continuous variable, and thus whether a hard threshold of 10 mut/Mb should be used for all cancer types remains unclear[50]. Researchers must carefully consider the effects of sample, technical and bioinformatic-related variations on the accuracy of mutation detection, which in turn affects TMB.

The TMB Harmonization Project[71] was established to develop guidelines for the quantification and use of TMB. These guidelines recommend a uniform TMB calculation method for WES analysis, with 300× median coverage, read depth aligned at reference variant position (≥25), variant count ≥3, variant allele frequency cut-off value of 5% and exclusion of synonymous variants[44].

Researchers consider that tumours with a high TMB have an increased potential to encode neoantigens[72], which are tumour-specific 'novel' antigens generated from somatic mutations in cancer cells[29,54,73,74] and constitute potential targets of immune effector cells[75]. Nevertheless, whether a high TMB renders a tumour more immunogenic because it increases the number of neoantigens recognized by the immune system[72,76] or the likelihood of having a few robust neoantigens for which the host has no tolerance[77,78] remains to be elucidated. This consideration is important for the design of future computational approaches and might enable selection of promising neoantigens for inclusion in neoantigen-based vaccines.

## HLA genotype

HLA genes encode major histocompatibility (MHC) molecules, which present antigens, including tumour-specific neoantigens, to T cells. MHC class I (MHC I) molecules present antigens to $CD8^+$ cytotoxic T lymphocytes (CTLs) and MHC II molecules present antigens to $CD4^+$ helper or regulatory T ($T_{reg}$) cells. Therefore, MHC molecules have a key role in infection, inflammatory conditions, autoimmune diseases and antitumour immune responses, and if unable to load neoantigens, are a putative mechanism of immune escape. Indeed, greater HLA diversity, resulting in a more diverse set of MHC molecules that can present a wider range of mutant peptides to immune cells, is associated with a lower risk of developing certain cancers[79] and is associated with a better response to ICIs[80]. By contrast, tumour-associated loss of heterozygosity (LOH) within HLA regions can mimic reduced HLA diversity and potentially facilitate immune evasion by enabling robust neoantigens to 'hide' from host immune responses. An analysis of HLA genotypes from more than 1,500 patients with melanoma or NSCLC receiving ICIs showed that patients with HLA heterozygosity have significantly better overall survival (OS) than those with one or more homozygous HLA class I (HLA-I) alleles ($P \leq 0.003$ in two independent cohorts)[81]. In the TRACERx cohort[82], 40% of patients with early-stage NSCLC had LOH in HLA regions, which was significantly associated with an elevated burden of subclonal neoantigens (or those expressed only by a subset of cancer cells) ($P = 0.008$), potentially contributing to immune escape[83]. Some evidence indicates that specific HLA-I genotypes influence response to ICIs. For example, in patients with melanoma *HLA-B44* (allele frequency: 0.45) supertype alleles (*HLA-B*18:01*, *HLA-B*44:02*, *HLA-B*44:03*, *HLA-B*44:05* and *HLA-B*50:01*) correlate with improved OS (HR 0.61, 95% CI 0.42–0.89; $P = 0.01$) and *HLA-B62* (allele frequency: 0.15) supertype alleles (*HLA-B*15:01*, *HLA-B*07:02* and *HLA-B*53:01*) with unfavourable OS (HR 2.29, 95% CI 1.40–3.74; $P \leq 0.001$)[81]. In light of these results, investigators proposed an HLA-I evolutionary divergence (HED) score to calculate the effect of diversity of HLA-I alleles on the efficacy of ICIs[80], with high HED scores (defined as mean HED greater than or equal to the top quartile) indicating better responses.

Having accurate HLA identification methods is crucial. Conventional HLA genotyping is performed using serology and PCR-based in vitro techniques, enabling up to 2- to 4-digit resolution of HLA genotypes. Researchers have developed methods for HLA genotyping from next-generation sequencing (NGS) data, some of which can determine HLA genotypes with 6- to 8-digit resolution with rapid, high throughput[84] (Supplementary Table 1). Computational tools for HLA genotyping from WGS, WES and RNA-seq data are based on two approaches: alignment, whereby sequence reads are aligned to reference allele sequences and the genotypes are predicted through probabilistic models; and de novo assembly, whereby reads within HLA regions are initially assembled and then aligned to best reference known HLA alleles (Supplementary Table 1). Computational methods used to determine HLA genotype have been described as 99% accurate compared with PCR[84,85]; however, sequence-read depth and sample tumour purity differentially affect each approach, and thus improved methods and benchmarking studies are required. Comparisons of several HLA genotyping algorithms revealed that OptiType, which is alignment based, has the highest level of accuracy for estimating HLA-I genotypes from WES or WGS data[85,86], achieving 100% accuracy HLA-I genotyping using sequencing data with an average read depth of 20× in HLA and flanking intronic regions[85]. Long-read sequencing is an emergent and exciting approach that can resolve highly polymorphic HLA regions by sequencing long reads (>10 kb) to enable full-length sequencing of HLA alleles and genotype prediction[87].

## Neoantigen prediction

Cancer-specific antigen processing and presentation is a complex process that involves transcription of a genomic alteration, translation,

retention of the alteration as a cleaved peptide and MHC binding for presentation on the cell surface. MHC-bound neoantigens are subsequently 'seen' by the immune system[24,56,57]; this signal is the first step in the generation of T cell-mediated antitumour immunity[72].

**Prediction of immunogenic neoantigens.** Analysis of genome sequencing data using advanced bioinformatic algorithms can be used to predict the likelihood of presentation of tumour-specific candidate neoantigens through the characterization of somatic mutations and HLA genotypes. The identification of neoantigens from sequencing data is a multistep process that involves the detection of somatic mutations, prediction of protein cleavage and calculation of peptide–MHC binding affinity and stability (Fig. 2a). The algorithms used to predict proteasomal cleavage, peptide transportation to the endoplasmic reticulum and MHC–peptide binding affinity are limited by the datasets used for their training. Users need to consider these limitations for accurate analysis of large datasets. The 3.1 version of the NetChop tool[88] has been trained with publicly available data from 1,260 HLA alleles (using only C-terminal cleavage data) or 20S cleavage data from in silico datasets. NetCTL[89], a tool that estimates the trafficking of neoantigens by integrating prediction of proteasomal C-terminal cleavage, transporter associated with antigen processing (TAP) complex transport efficiency and MHC–neoantigen binding efficiency, has been trained using a database in which peptides were grouped by affinity with one of only 12 MHC I supertypes. The 4.0 version of the MHC binding-affinity tool NetMHC[90] was trained on a dataset of 81 HLA alleles, and other approaches, such as NetMHCpan[91] and MHCflurry[92], were developed to predict binding affinity to any known MHC. Many peptide–MHC binding prediction algorithms use artificial neural networks (ANNs), a machine learning approach, trained on manually curated binding-affinity data from 1,141,280 peptides catalogued from experimental data on T cell epitopes deposited in the Immune Epitope Database (IEDB)[93], PRIDE[94] and SysteMHC[95]. As such, ANN-based models show a bias towards the selection of viral-like epitopes and are mostly trained on the most frequent HLA alleles (for example, *HLA-A*02:01* and *HLA-B*07:02*)[96,97].

Computational neoantigen prediction pipelines have been developed to integrate a myriad of algorithms specific to each step leading to neoantigen presentation (Fig. 2a and Supplementary Table 2). A major issue with neoantigen prediction is the prioritization of candidate events capable of eliciting an immune response. Only 1–5% of computationally predicted MHC I-bound neoantigens have been shown to induce an immune response, and thus the performance of these approaches is quite poor[98,99]. Researchers developed a modified computational workflow to enhance neoantigen prediction[100] that considers anchor residues for peptide binding to MHC I molecules. To estimate the likelihood and clinical relevance of neoantigen formation and recognition by T cells, Łuksza et al.[101] proposed a neoantigen fitness model based on the neoantigen–MHC binding affinity and nonlinear dependence sequence similarity between candidate neoantigens and known antigens. The model assigned patients with pancreatic cancer[102], lung cancer[29] or melanoma[27,53] who were receiving ICIs to low versus high neoantigen fitness groups, which had statistically significant differences in OS. These results indicate that immune-mediated selection pressures can potentially shape the landscape of neoantigens recognized by T cells.

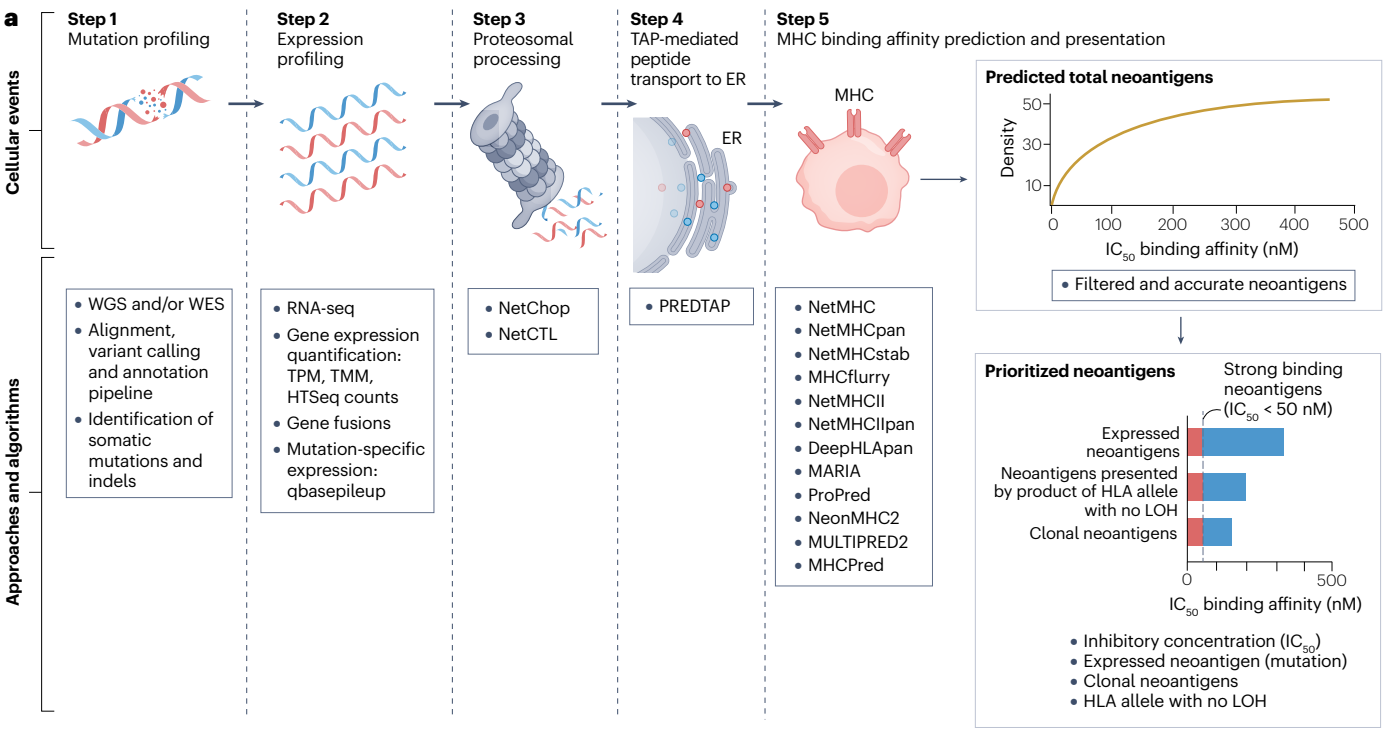
Most neoantigen binding prediction tools have been trained with data on neoantigen binding to MHC I molecules. Prediction of binding to MHC II molecules has proved even less reliable than MHC I–neoantigen binding prediction[103], owing in part to the increased complexity of MHC II presentation whereby longer peptides are presented (16- to 21-mer versus 8- to 11-mer for MHC I), by the 'promiscuity' of MHC II–peptide binding and by the limited number of datasets available for training prediction algorithms. Various computational approaches, including NetMHCII[104], NetMHCIIpan[91,104], MULTIPRED2 (ref. 105) and MHCPred[106], have been developed to predict MHC II-binding epitopes using ANNs. MARIA[107] is a deep learning model trained using mass spectrometry data on naturally MHC II-presented peptides. MARIA outperformed NetMHCIIpan in a melanoma dataset when cross-validated against known MHC II-bound antigens[107], but its performance has yet to be confirmed in independent studies.

To overcome the challenges associated with training datasets, researchers have used mass spectrometry-determined data. However, proteomic approaches have low sensitivity; indeed, a deep-proteome analysis of 29 healthy tissues detected only 2.4% of mutated variants detected at the exome level[108]. Researchers have combined mass spectrometry analysis of MHC-bound peptides with WES data of 16 melanoma samples from seven patients, identifying five neoantigens, three of which elicited immune responses[109]. The integration of the three available databases (IEDB, PRIDE and SysteMHC) and the addition of positive-control data through the efforts of a global consortium[110] to provide a large source of epitope information would help researchers to develop and test new tools and algorithms[93]. The development and sharing of such data will provide valuable information that might improve the accuracy of neoantigen prediction tools and enable the design of neoantigen-based vaccines.

Neoantigen prediction algorithms are popular; however, users should be aware of the strengths and limitations of selected computational approaches before applying them to large datasets (Supplementary Table 3). For example, the prediction of neoantigens is computationally challenging, frequently requiring long computational times. In our experience, for tumours with a high TMB, estimation of the neoantigen load for MHC I molecules can take >400 CPU hours. The best algorithm for a particular application will depend on the specific requirements of the study; each algorithm has specific advantages and disadvantages.

A challenge with current approaches is that they are not designed to take into account genomic ITH, which is a major determinant of resistance to ICIs. Indeed, ITH and the emergence of subclones during tumour development and/or treatment can explain how tumours evade the immune system. Consistent with this principle, researchers reported that patients with NSCLC or melanoma who were receiving anti-PD-1 antibodies and who preferentially harboured clonal neoantigens had better outcomes than those preferentially harbouring subclonal events[31]. Tools that integrate genomic ITH, neoantigen trafficking, clonality, expression and immunogenicity of neoantigens are currently needed.

**Neoantigens from non-canonical sources.** Most approaches designed to identify candidate neoantigens analyse somatic missense SNVs. However, clinical benefit with ICIs has been reported in some patients with tumour types that typically have a low TMB[111,112], such as glioblastoma[113,114] or mesothelioma[115–117], suggesting that SNVs are not the only source of antigens. Potential non-canonical sources of neoantigens include frameshift indel mutations that create novel open reading frames[118], somatic structural variants (including those that disrupt gene loci)[119], gene fusion events[120], aberrant gene splicing[121–123] (including intron retentions[124]), RNA editing[125], long non-coding RNAs[126],



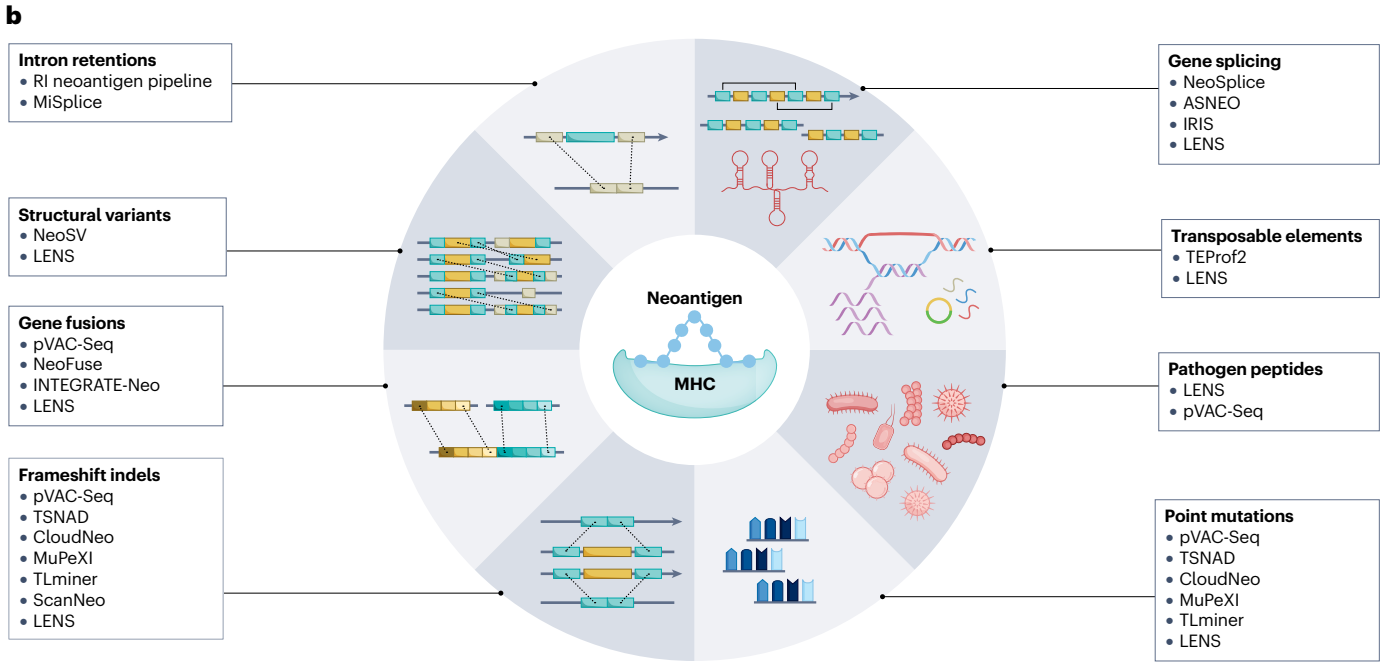




**Fig. 2 | Computational approaches to identify cancer neoantigens from various sources. a**, Researchers have developed approaches to identify cancer-specific neoantigens at various stages of neoantigen processing, from their synthesis to presentation on major histocompatibility complex (MHC) molecules. **b**, Commonly studied sources of neoantigens include canonical sources, such as somatic point mutations, and small insertions and/or deletions (indels), as well as non-canonical sources of somatic neoantigens, including gene fusions, structural variants, transposable elements, intron retentions and aberrant gene splicing. In addition, microbial peptides with some degree of similarity can avert or divert immune responses. The boxes around the diagram include computational approaches to predict neoantigens derived from the sources shown in the inner layer. ER, endoplasmic reticulum; $IC_{50}$, half-maximal inhibitory concentration; LOH, loss of heterozygosity; RNA-seq, RNA sequencing; TAP, transporter associated with antigen processing; TMM, trimmed mean of *M* values; TPM, transcript per million; WES, whole-exome sequencing; WGS, whole-genome sequencing.

other cancer-specific transcription events and endogenous retroelements[127] (Fig. 2b).

Before the 2020s, computational approaches to identify neoantigens from non-canonical sources did not exist. Subsequently, researchers developed NeoSplice[128], IRIS[129] and ASNEO[130] to predict neoantigens generated from splicing events. This source is particularly attractive because a pan-cancer analysis indicates that splicing-derived neoantigens predicted to bind MHC are at least twice as abundant as those generated from SNVs[121]. Nonetheless, whether neoepitopes derived from splicing are more or less likely to trigger an immune response than those derived from other sources is unclear, partly owing to the intricate nature of RNA processing, which poses challenges in accurately identifying candidate splicing-derived neoantigens.

Cancer-specific transposable elements are also receiving attention as a source of neoantigens[131–133]. Peptides derived from non-canonical splicing (using junctions between exons and transposable elements) can produce MHC I-presented cancer-specific antigens for which reactive T cells were isolated from patients with NSCLC[132,133]. These peptides constitute promising targets for cancer vaccines. Another group developed the TEProf2 pipeline[134] using pan-cancer data from The Cancer Genome Atlas (TCGA) to identify more than 2,000 transposable element transcripts, including the recurrent event *L1P2_STK31*, which was found in 15% of tumours analysed and results in a chimeric protein with neoantigen potential.

Researchers developed NeoSV[135] to analyse WGS data and predict MHC I-specific neoantigens from somatic structural variants that occur within gene loci. NeoSV assembles structural variant-derived neotranscripts, computationally translates them into proteins and uses a so-called sliding window approach to determine all potential short peptides. The peptides that contain at least one mutation that confers changes in the amino acid sequence that could potentially generate cancer-specific neopeptides are selected, and the MHC I binding affinities are predicted using NetMHCpan. A challenge when using structural variant events for neoantigen prediction is that most structural variant breakpoints occur in between coding sequences of the genome and thus, predicting the effect of these structural variants on transcription using WGS data alone is often computationally challenging. LENS[136] is a comprehensive approach that uses RNA-seq and DNA-seq data to predict cancer-specific antigens that arise from various genetic alterations including SNVs, indels, self-antigens (such as cancer testis antigen 1; also known as NY-ESO1) and non-canonical sources, such as fusion events, splice variants, viruses and endogenous retroviruses. LENS is a promising approach that warrants further testing to confirm accuracy of detection. Predicting the complete landscape of neoantigens would require comprehensive genome profiling, such as long-read WGS or full-length RNA-seq, coupled with validation of the computational approaches to reliably annotate and retrieve amino acid sequences that are likely to bind to MHC I or II molecules.

**Neoantigen-derived cancer vaccines.** Current computational algorithms can accurately determine patient-specific HLA genotypes but cannot predict all potential immunogenic neoantigens with confidence; therefore, the development of neoantigen-based vaccines has been slow. Nevertheless, candidate neoantigens can be targets of therapeutically effective cancer vaccines[137–139]. Such vaccines will almost always need to be personalized because each patient's tumours express different neoantigens.

Neoantigen-based vaccines have been tested in numerous early-phase clinical studies, either alone or in combination with ICIs; some of them have been proved to be safe and tolerated[140–143]. At the time of writing this Review, ClinicalTrials.gov listed more than 150 clinical trials that included the search terms 'cancer' AND 'neoantigen'. In the past few years, several advances have occurred in the development of vaccines for pancreatic cancer. In a case report, a patient with pancreatic ductal adenocarcinoma (PDAC) who had received *HLA-C*08:02*-restricted T cell receptors (TCRs) targeting a neoantigen derived from the recurrent $KRAS^{G12D}$ mutation had continuous tumour regression after 6 months without any reported toxicity[144]. The engineered T cells remained functional in the bloodstream[144]. In a phase I clinical trial, 16 patients with resected PDAC sequentially received the anti-PD-L1 antibody atezolizumab, an mRNA–lipid-based neoantigen vaccine comprising up to 20, predominantly SNV-derived, neoantigens and then 15 patients received combination chemotherapy (folinic acid, fluorouracil, irinotecan and oxaliplatin (mFOLFIRINOX)). The eight patients with clinical responses had robust T cell responses to at least one neoantigen and, at a median follow-up duration of 18 months showed no disease recurrence, compared with a median recurrence-free survival (RFS) duration of 13.4 months in patients without detectable vaccine-expanded T cells ($P = 0.003$)[145]. Advances have also been reported in patients with melanoma. In a phase II trial, patients with resected melanoma who received a neoantigen vaccine containing up to 34 neoantigens and pembrolizumab had improved 18-month RFS compared with those receiving pembrolizumab alone (78.6% versus 62.2%; HR 0.56, 95% CI 0.31–1.02)[146].

In terms of neoantigens from non-canonical sources, vaccines that target gene fusion products could be used in patients with tumour types in which these alterations are common, including leukaemias[147], breast[148] and colorectal[149] cancer, renal carcinoma[150] and sarcoma[151], or if they are shared between tumour histologies. Even so, collectively shared fusion products can be rare, posing challenges for drug development and requiring a personalized approach. For example, the ongoing response to pembrolizumab and T cell activation of a patient with head and neck squamous cell carcinoma (HNSCC) was increased when they received autologous peripheral blood mononuclear cells pulsed with a peptide derived from the *DEK–AFF2* gene fusion[120], which has been reported in 48% of carcinomas of the sinonasal region and skull base[152]. In a pilot study, patients with Ewing sarcoma or alveolar rhabdomyosarcoma received dendritic cells pulsed with neoantigens derived from translocation breakpoint-encoded fusion peptides combined with a HPV16-E7 peptide. This approach was minimally toxic and 5-year OS was higher in the 30 patients who started therapy (43%) than in all patients who underwent apheresis in the study ($n = 52$; 31%)[153].

Neoantigen approaches have been used in chimeric antigen receptor (CAR) T cell therapies. in which engineered neoantigen-specific TCRs are expressed in T cells[154]. A key example is a first-in-human phase I trial that used a non-viral gene editing method to generate neoantigen-specific TCRs to target patient-specific peptides in individuals with solid tumours[155]. This trial demonstrated the feasibility of such an approach.

Computational approaches to predict neoantigen candidates for cancer vaccines will require analytical validation for different sample types, purity of sample, sequencing methodologies and the quality of sequencing. To further advance the development of these therapies, researchers must thoroughly validate the precision of epitope prediction pipelines via immune monitoring during early clinical investigations, not only in terms of immunogenicity but also to determine the extent of tumour recognition[145]. Neoantigen-based vaccines can be facilitated by early resection of the tumour, which provides researchers

and clinicians with a time window (usually 2–3 months) to design, produce and administer neoantigen-based therapeutic vaccines. Prior cancer treatments (such as chemotherapy, radiotherapy and ICIs) can substantially modify T cell phenotypes, although whether these modifications have a positive or negative effect on the response to neoantigen-based vaccines remains to be determined. Profiling of ITH and monitoring tumour evolution to enable the targeting of clonal neoantigens could enable the design of vaccines against the broadest possible range of tumour neoantigens present in each patient's tumour.

## Analysis of cancer-extrinsic features

Cancer-extrinsic features that can be analysed using sequencing data encompass features of the immune system and the TME (Fig. 1). Tumours orchestrate a unique TME comprising a diverse array of stromal and immune cells (reviewed in detail elsewhere[156–159]); the crosstalk between cancer cells and cells within the TME contributes to tumour evolution and progression through processes such as cancer immunoediting and immune surveillance[160–163]. The TME also contains a dynamic complex network of cytokines, chemokines, growth factors and adhesion molecules that drive interactions between cancer, stromal and immune cells, affecting response to treatment[164–169]. In particular, the presence or absence of activated T cells and CTLs in the TME is characteristic of a T cell-inflamed (also referred to as 'hot') or non-T cell-inflamed (or 'cold') TME, respectively, which might influence response to ICIs[170].

Several algorithms have been developed to interrogate RNA-seq and DNA methylation profiling data to dissect the TME (Fig. 3 and Supplementary Table 4). Overall, these deconvolution algorithms incorporate various statistical and probabilistic tools that compare gene expression or DNA methylation profiling data between an input sample and a defined dataset. A key strength of these methods is the simple output — relative proportion of immune cells, which is easily interpretable — but their predictive performance is influenced by the source of the reference signature matrix, data preprocessing and tumour purity[171,172].

### Approaches using RNA-seq data

CIBERSORT[146] uses the LM22 547-gene expression signature as a reference matrix and excludes genes that are highly expressed in the input tumour gene expression matrix because they would adversely affect the estimation of known or unknown immune cells. The TIMER tool[173] uses reference gene expression signatures for six immune cell types, which are derived from the Human Primary Cell Atlas[174]. ImSig[175] uses gene signatures that represent seven immune cell populations and three biological processes (proliferation, interferon response and translation), defined from a gene correlation network analysis. EPIC[176] incorporates gene expression signatures for six major circulating immune cell types curated from the literature[152–154] and single-cell RNA-seq (scRNA-seq) data for five major melanoma-infiltrating immune cell types[155]. xCell[177] is a webtool that makes a linear comparison between cell types using a curve-fitting method that uses RNA-seq data from 1,822 samples annotated to 64 cell types as a reference matrix and provides diverse, sensitive enrichment of multiple cell types. MCP-counter[178] estimates the absolute abundance of eight tumour-infiltrating immune cell types and two stromal cell types using 1,194 samples representing cells within the TME and 742 cancer samples annotated with 63 cell types. Consensus$^{TME}$ (ref. 179) uses cell-type-specific genes from seven approaches, including CIBERSORT, TIMER, xCELL and MCP-counter to estimate 18 immune cell types in the TME. Other TME deconvolution approaches use RNA-seq data in combination with data from other sources. quanTIseq[180] is a computational pipeline that can deconvolute ten immune cell types from bulk RNA-seq data, with the option to use histology images from whole-tissue sections (immunofluorescence, immunohistochemistry or haematoxylin and eosin-stained tissue slides) to scale the fractions of the deconvoluted immune cell types to cell densities.

Deconvolution approaches have used RNA-seq data to investigate the TME (Fig. 3). Using CIBERSORT, researchers identified six distinct immune subtypes from more than 10,000 samples representing 33 cancer types within TCGA. The inflammatory subtype, marked by high levels of T helper cell markers (specifically $T_H17$ and $T_H1$), had the most favourable prognosis[181]. MCP-counter analysis of the TME from 608 soft-sarcoma samples was used to classify these tumours into three groups: immune low, immune high and highly vascularized[182]. In a phase II clinical trial involving 115 patients with oesophageal adenocarcinoma, researchers used the Consensus$^{TME}$ approach to stratify patients into four immune clusters, including an immune-hot cluster associated with better OS ($q = 0.003$)[183].

### Approaches using DNA methylation profiling data

DNA methylation profiling has emerged as a promising approach to deconvolute immune and stromal cell types. Researchers adapted the CIBERSORT approach, to develop methylCIBERSORT[184], which uses a reference of fibroblasts and seven different immune cell types to estimate the absolute proportion of immune cells and tumour purity. In an analysis of HNSCC, methylCIBERSORT was able to identify 'immune-hot' and 'immune-cold' clusters, with the immune-hot samples displaying increased CTL to $T_{reg}$ cell ratios ($P < 10^{16}$)[184].

An earlier approach that used differentially methylated DNA regions between cell types was developed to analyse methylation array data to assess the abundance of immune cells[185] and has been used by researchers to perform epigenome-wide association studies[186,187]. IDOL[188] is a deconvolution approach that initially compared methylation status for 300 CpG sites to predict six cell types in whole-blood samples. Subsequently, the reference signatures for IDOL were expanded to incorporate 12 immune cell types[189]. EpiDISH[190] uses DNase Hypersensitive Site maps from ENCODE[191] to construct a reference DNA methylation database to quantify cell types in whole-blood samples. Researchers of methylation-based deconvolution approaches should be aware that estimation of DNA methylation in cancer cells is influenced by somatic CNAs. CAMDAC[192] is a copy number-aware deconvolution approach that extracts methylation profiles from bisulfite sequencing data from bulk tumours and matched nonmalignant tissue to estimate nonmalignant cells, cancer cells and intratumour subclonal relationships. CAMDAC was used to reveal copy number-independent allele-specific expression that is linked to epigenomic dysfunction[49].

DNA methylation-based deconvolution methods, which largely focus on array-based methylation profiling, offer advantages over RNA-seq deconvolution methods, as they can work from FFPE samples[184]. The emergence of direct DNA-seq approaches[193,194] will enable the simultaneous assessment of DNA mutations and methylation deconvolution and thus, we expect, will become the preferred approach for methylation-based deconvolution of the TME.

### Incorporation of single-cell information

Various deconvolution tools have been developed to estimate the proportion of immune cells within the TME; however, their use comes with several challenges. Overall, deconvolution tools can provide robust estimates of the most abundant immune cell types (such as

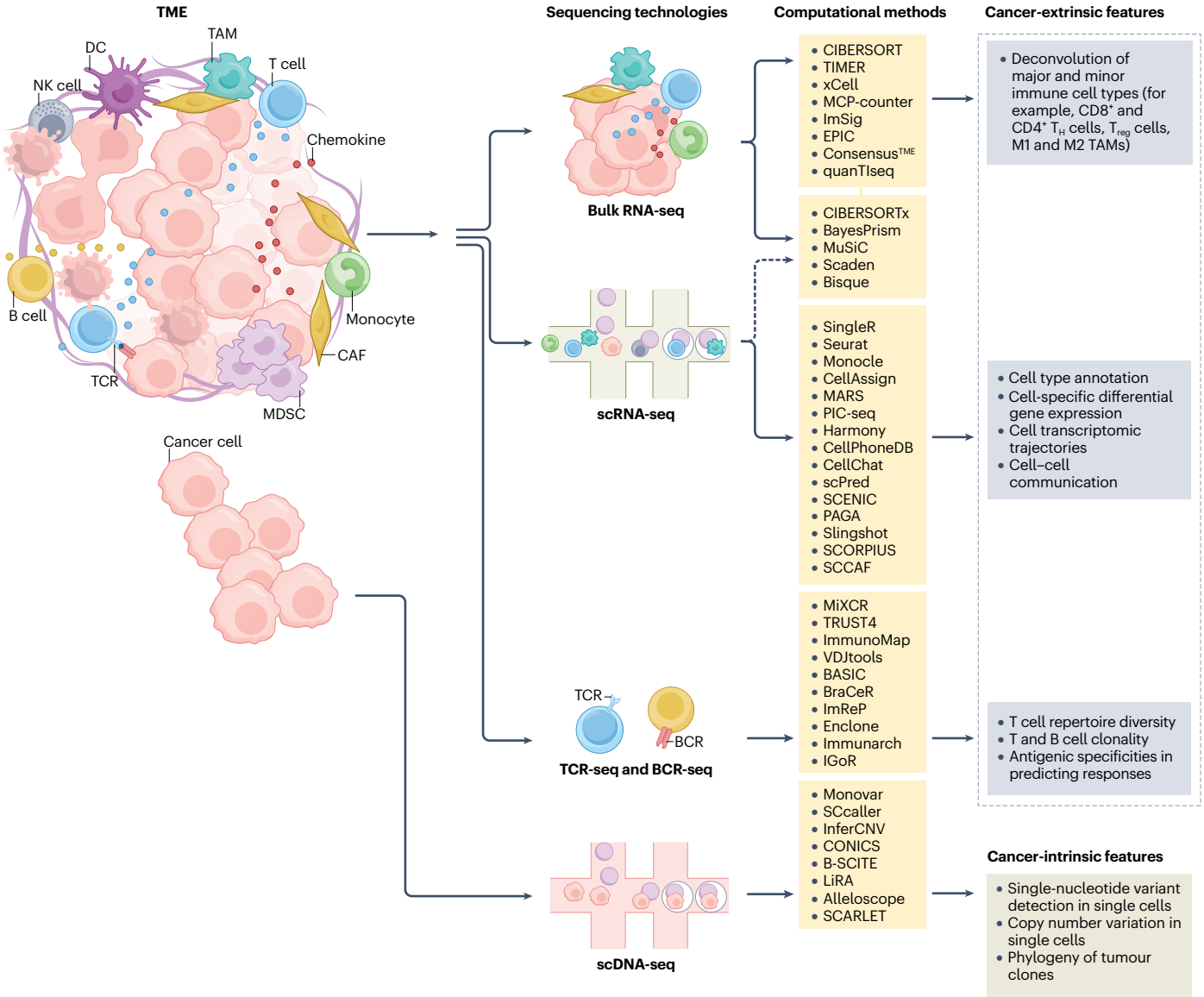



**Fig. 3 | Computational analysis of bulk tissue and single-cell RNA sequencing data to study the tumour microenvironment and associated features.** Multiple computational methods are available to analyse the complexity of, and various cell types that constitute, the tumour microenvironment (TME). Several tools use RNA sequencing (RNA-seq) data from bulk tissue samples to deconvolute the TME, some of which are based on reference signatures derived from single-cell RNA-seq (scRNA-seq) data (dashed arrow). Various scRNA-seq tools have been developed to assign cell labels and determine cell-specific differential gene expression, cell trajectories and cell–cell interactions. The repertoire diversity, clonality and specificity of B and T cells can be identified using computational approaches and sequencing of T cell receptors (TCRs) or B cell receptors (BCRs). Several approaches have been developed to identify somatic mutations and study tumour genomic heterogeneity and evolution using single-cell DNA sequencing (scDNA-seq) data. CAF, cancer-associated fibroblast; DC, dendritic cell; MDSC, myeloid-derived suppressor cell; NK, natural killer; TAM, tumour-associated macrophage; $T_H$ cell, helper T cell; $T_{reg}$ cell, regulatory T cell.

T and B cells), but have limitations in estimating immune cell types with lower abundance or T cell subsets and other admixtures of cell types associated with tumour and stromal cells[172,195]. The emergence of scRNA-seq datasets, leading to the identification of cancer-specific gene signatures, and deep learning models has enabled improvements in deconvolution approaches.

CIBERSORTx[196] is an extended framework of CIBERSORT that requires a signature matrix derived from scRNA-seq data as a reference dataset to infer cell-type-specific gene expression from bulk RNA-seq data without the need for physical cell isolation. MuSiC[197] and Bisque[198] can estimate the proportions of immune cell types from bulk RNA-seq data using cell-type-specific information derived from multiple scRNA-seq datasets. BayesPrism[199] applies a Bayesian statistical method to estimate cell-type fraction and cell-specific gene expression, which uses an scRNA reference dataset. Analysis of samples from pretreated patients with oesophageal adenocarcinoma using BayesPrism revealed

that a high presence of monocytes within the TME after receiving a combination of ICI and chemotherapy was significantly linked to favourable OS (HR 0.38, 95% CI 0.22–0.67; $P = 0.0008$; false discovery rate 0.037)[200]. Scaden[201] is a deep neural network-based deconvolution tool trained on scRNA-seq data that does not rely on GEP signatures, instead inferring the informative features from the cell types that it has never been trained on, and thus it can deconvolve data on those cell types.

### Method selection

Before performing TME deconvolution analysis, researchers should carefully consider which approach is better suited to their needs. When selecting an optimal method, researchers should be cautious about sample and technical factors that can affect estimation accuracy, which include tumour purity, the presence of rare cell types, the reference matrix and data transformation. Using 1,412 bulk RNA-seq and 85 scRNA-seq samples from patients with glioblastoma, HNSCC or cutaneous melanoma, the developers of BayesPrism found that the approach outperformed CIBERSORTx and MuSiC[199]. An independent comparison of 20 deconvolution approaches revealed that data transformation affects three of these methods[171] and suggested that data be kept in linear scale for deconvolution. Subsequently, a comparison of nine approaches to deconvolve breast cancer TME found that BayesPrism, Scaden and MuSiC outperformed other methods in simulated samples of varying tumour purity and that BayesPrism and DWLS have the best performance when deconvolving granular immune lineages[172].

Further benchmarking and validation of the performance of deconvolution tools are required before their implementation as a decision support tool for prognostication or treatment selection in clinical settings. Such benchmarking will inform on whether certain deconvolution tools are better suited for a given cancer type and perform better on certain types of cell. In conclusion, no one of these approaches is generally superior to the others, and the selection of a specific tool should depend on the research question and experimental design.

## Single-cell analysis

In the past few years, single-cell sequencing (scseq) technologies have undergone a transformation that has enabled the creation of datasets and stimulated the development of analytical methods. At the time of writing this Review, 1,518 tools listed in 30 categories (including, among others, mutation, CNA, clonality, clustering, visualization, dimensionality reduction and cell–cell communication) were available at scRNA-tools. Tutorials or best-practice resources have been developed for some of these tools[202,203], and several studies have benchmarked a selection among them[204–207]. However, these initiatives remain ongoing owing to the continuous development of new datasets and tools. As samples are collected prospectively, single-cell analysis from ICI-treated patients will be increasingly performed. Here, we present a succinct overview of computational tools that can be used to assess cancer-intrinsic and cancer-extrinsic features at the single-cell level using sequencing technologies (Fig. 3).

### Cancer-intrinsic features

scDNA-seq and scRNA-seq enable the identification of genetic variations present within each individual cell, providing insights into evolutionary dynamics and ITH. However, identifying somatic mutations through scseq poses key challenges owing to amplification biases, sequencing errors and sample contamination, which might introduce false-positive and false-negative findings, making it challenging to distinguish true mutations from technical artefacts.

Several computational tools have been developed to overcome some of the technical challenges to identify somatic SNVs (Fig. 3). Monovar[208] uses a probabilistic approach to detect mutations in individual cells by comparing their allelic frequencies with those of the bulk reference genome. Researchers used this tool to reveal the cellular-level co-occurrence and mutual exclusivity among driver mutations in acute myeloid leukaemia[209]. LiRA[210] phases sequence reads using germline variants as a reference to differentiate somatic variants from artefacts produced by whole-genome amplification. SCcaller[211] identifies SNVs using a likelihood ratio test to distinguish real SNVs from artefacts. This tool was used to discriminate somatic mutations in human lung epithelial cells as a function of ageing and smoking status[212].

Researchers have also developed multiple approaches to identify CNAs within single cells (Fig. 3). Alleloscope is an approach developed to identify allele-specific CNAs in data obtained using scDNA-seq or so-called scseq assay for transposase-accessible chromatin[213]. Owing to the association of gene expression with CNA status and the higher dynamic range of scRNA-seq over scDNA-seq, researchers have developed approaches that use scRNA-seq data to identify somatic CNAs. InferCNV[214] and CONICS[215] are widely used to estimate the copy number states of genomic regions, providing insight into ITH of pancreatic[216], colorectal[217], ovarian[218] and gastric[219] cancer, NSCLC[220], nasopharyngeal carcinoma[221] and melanoma[222].

The characterization of somatic mutations within single cells facilitates the reconstruction of clonal lineages and can help to infer the order and timing of genomic events[223–226]. Researchers have developed sophisticated methods that integrate various features to track the evolution of cell populations and infer tumour phylogenies. Two such tools are B-SCITE[227], a probabilistic approach that integrates bulk DNA-seq and scDNA-seq data, and SCARLET[228], which identifies and uses somatic SNVs and CNAs from scDNA-seq (Fig. 3). Of note, detection of mutations in single-cell data remains challenging, and the choice of analytical tool will depend on the specific characteristics of the data and the research goals. In the future, the ability of single-cell approaches to characterize cancer-intrinsic features will probably be enhanced through integration with long-read sequencing technologies.

### Cancer-extrinsic features

The introduction of scRNA-seq has transformed immuno-oncology research, with studies from the past decade providing insights into the complex cellular and molecular nature of the TME[229–234]. The application of scRNA-seq to study cancer-extrinsic features is rapidly evolving, with new tools and methods being continuously developed (Fig. 3). A widely used and recommended approach for scRNA-seq analysis is Seurat[235–238], which provides a comprehensive set of tools for quality control, data normalization, dimensionality reduction, cell clustering, differential expression analysis, visualization and multi-omic integration in its latest suite. Harmony[239] seeks to overcome multiple experimental and biological factors to enable the integration of large-scale scRNA-seq data by performing 'batch effect' correction, which has been used with pan-cancer data[240].

A key component of the analysis of single cells within the TME is cell annotation. Tools developed for cell annotation include scPred[241], CellAssign[242] and SCCAF[243], which use machine learning models trained on predefined gene sets related to immune cell types. Similarly, MARS[244] is a deep learning approach that enables annotation of known and unknown cell types. SCENIC[245] is a computational framework for identifying gene regulatory networks and functional cell states. Monocle[246,247] is specifically designed for reconstruction of

developmental trajectories, identification of genes that drive cell fate decisions and characterization of transcriptional dynamics. A study that benchmarked 45 single-cell trajectory inference methods found a large variance in the performance of these approaches; the authors developed guidelines to facilitate the choice of approach[248]. Overall, the graph-based method PAGA[249], the tree method Slingshot[250] and the linear method SCORPIUS[251] generally performed well in all comparisons.

A strength of scRNA-seq is that it can be used to study cell–cell communication networks within the TME. CellChat[252] and CellPhoneDB[253] enable the identification of ligand–receptor interactions between any cell subtype. PIC-seq[254] stains cells in situ and then sorts cells for scRNA-seq with fluorescence-activated cell sorting to identify those that interact, using computational modelling to examine these interactions and their influence on gene expression.

scRNA-seq methods are valuable tools for dissecting the heterogeneity and functional states of immune cells within the TME. Currently, the application of these methods to analyse clinical datasets, such as those derived from patients receiving ICIs, has been limited, but we expect these studies to be performed eventually and to provide insight into response to ICIs.

### T and B cell repertoire diversity

Knowing the proportion of T and B cells within the TME is important; however, measuring the heterogeneity of specific T and B cells is also crucial. T cells can recognize neoantigens and induce antitumour responses through their highly polymorphic TCRs[255]. By profiling the TCR and/or B cell receptor (BCR) sequences of lymphocyte populations including tumour-infiltrating lymphocytes (TILs), researchers can characterize the clonal composition, expansion and heterogeneity of these cells. Various experimental and computational methods have been developed to directly sequence or infer TCR structure from genomic or bulk RNA-seq data (extensively reviewed elsewhere[256]). Single-cell TCR sequencing (scTCR-seq) in combination with scRNA-seq is a promising approach that enables the rapid and precise identification of diverse neoantigen-specific TCRs in lymphocytes from peripheral blood and in TILs[233,240,257] with prognostic implications[258]. This approach has the ability to identify non-genetic heterogeneity between cancer cells, including that related to T cell state transitions and expansions during treatment with ICIs[259,260].

Computational tools designed to analyse T cell and B cell repertoire diversity begin with the recovery of TCR and BCR sequences from raw data, followed by clustering and annotation[261–263]. Subsequent steps involve visualization of the immune repertoire, including clonotype abundance, diversity and V(D)J usage, as well as analysis of individual clonotypes. A wide range of computational tools has been developed to analyse the TCR and BCR repertoires using bulk RNA-seq and scTCR-seq and scBCR-seq data (Box 1 and Fig. 3). Although standards for TCR sequence reconstruction have not been established yet, MiXCR[261] and TRUST4 (ref. 263) are commonly used for this purpose. The continued advances in single-cell technologies present immense possibilities for discovery and innovation; certainly, the methods and tools discussed here will further evolve to enable deeper analyses in the next few years.

## Prediction of response to ICIs

The ability to consistently identify patients who are likely to respond to ICIs would provide enormous benefit to both patients and health-care systems. Although the influence of individual clinical features, such

## Box 1

# Computational approaches to estimate T cell and B cell repertoire diversity

- MiXCR[261] estimates the repertoire of clonotypes from paired-end or single-end reads of DNA or RNA sequences in a reference library of human and mouse germline immunoglobulin V, D, J and C loci sequences from the GenBank[291] database. Paired-end sequence data are more reliable and provide a high level of accuracy when applied to V and J alignment.
- IGoR[262] is a comprehensive framework to estimate B cell receptor (BCR) and T cell receptor (TCR) repertoires from bulk RNA sequencing (RNA-seq) or DNA sequencing (DNA-seq) data.
- TRUST4 (ref. 263) uses bulk RNA-seq or single-cell RNA-seq (scRNA-seq). TRUST4 aligns single-end or paired-end reads of RNA sequences to a reference human genome, followed by read extraction, de novo assembly and annotation.
- ImmunoMap[292] uses a phylogenetic approach to estimate TCR repertoire diversity and clonality from TCR sequencing (TCR-seq) data.
- VDJTools[293] provides a comprehensive platform for TCR and BCR repertoire sequencing analysis. VDJTools performs basic quality control statistics calculation, clustering of repertoires, filtering of unique clonotypes and user-defined annotation to estimate diversity index.
- Immunarch is compatible with Seurat[235–238] and provides support for data loading, analysis and visualization of widely used TCR and BCR computational analysis formats, including single-cell sequencing data.
- Enclone, from 10x genomics was developed to analyse TCR-seq and BCR-seq data. Enclone identifies and groups cells derived from common progenitors, known as clonotypes, and displays them along with notable characteristics, such as mutated amino acids.
- BCR assembly from single cells (BASIC)[294] detects anchor sequences in scRNA-seq data and uses these anchors to assemble BCR sequences.
- BraCeR[295] uses scRNA-seq data to estimate clonality and reconstruct complete BCR sequences.
- ImReP[296] is a computational method to profile the immunoglobulin repertoire using RNA-seq data from bulk tissues.

## Box 2

# Gene signatures and computational tools used to predict cancer immunotherapy responses

- IPRES[24] is a 26-gene expression signature associated with resistance to anti-PD-1 antibodies in patients with melanoma. However, this approach was not validated in other independent cohorts of patients with melanoma treated with immune-checkpoint inhibitors (ICIs)[56], suggesting unique features of RNA-seq analyses in the initial cohort.
- IMPRES[24] is a signature predictive of response to ICIs in melanoma encompassing 15 pairwise transcriptomic relationships between genes encoding immune checkpoints. However, other researchers raised concerns that IMPRES does not reproducibly predict response to ICIs in patients with metastatic melanoma[297]. Subsequently, the creators of IMPRES have validated the score using an independent melanoma dataset[56] and suggested that differences resulted from processing of the RNA-seq data[298].
- Inflammatory and epithelial-to-mesenchymal transition (EMT) signatures[299]: a 27-inflammatory-gene signature was significantly higher in ICI-treated patients with lung cancer with a clinical response versus those without a response ($P$=0.014). In the same cohort, a 12-gene EMT signature score was significantly lower (indicating predominance of epithelial genes) in responders versus non-responders ($P$=0.016). Similar trends were observed with the EMT signature in patients with malignant peritoneal mesotheliomas treated with atezolizumab plus bevacizumab[300].
- T cell gene expression signature score[301]: a 13-gene expression signature derived from an ICI-treated melanoma mouse model was proposed to activate the cancer-intrinsic β-catenin signalling pathway resulting in T cell exclusion, thus leading to ICI resistance. This signature was applied in a phase II prostate cancer trial[302]; however, the status of β-catenin signalling was not reported in this study and no association with T cell gene expression signature score and response rate was observed.
- IFNγ[25] T cell-inflammed gene expression profiles, a signature encompassing six genes related to IFNγ and expanded to 18 genes, was developed to predict clinical responses to ICI in melanoma. This approach was validated in 20 cancer types, as high scores have predictive value to ICI responses[303].
- TGFβ response signature[304] is a 19-gene expression signature that identifies immune evasion status and predicts response to ICIs in metastatic urothelial carcinoma. This signature is associated with resistance to ICIs via active T cell exclusion. This signature was validated in a phase II trial with patients with inoperable urothelial carcinoma treated with atezolizumab and neoadjuvant atezolizumab[305].
- Immunophenoscore[306] encompasses 20 single factors and six immune cell types and uses a random forest approach to predict responses to ICIs in 20 tumour types. This approach is not reliable and failed to predict responses in clinical datasets[307,308].
- TIDE[308] is a computational method to predict response to ICIs and tumour immune evasion in patients with melanoma or non-small-cell lung cancer. This approach was validated in a dataset of patients with melanoma[24,27,56,309] and non-small-cell lung cancer treated with ICIs[307,310].
- TREC[311] is a computational tool that predicts response to ICIs by estimation of T cell expansion from DNA sequencing data. This approach was validated in the TRACERx study[41].
- Netie[312] is a hierarchical Bayesian model that interrogates antitumour T cell interactions to infer evolution of neoantigens generated by each single-nucleotide variant over time and also estimates immune selective pressure during tumour progression.

as age[264], sex[265] and ethnicity[266], on the outcomes of patients receiving ICIs has been somewhat studied, analysis of cancer-intrinsic and cancer-extrinsic features has identified several features associated with response.

In addition to PD-L1 expression, several genomic features are associated with response to ICIs[24,27,57], highlighting the importance of comprehensively identification of tumour-specific genomic features. The presence of DNA mismatch repair deficiency[51,52], a high TMB or neoantigen load[24,57,73] have all been independently linked with a favourable response, whereas other genomic features, such as the presence of tumour aneuploidy[267] and a high structural variant mutational load[57], are independently correlated with poor responses.

Cancer-extrinsic features within the TME or peripheral blood, including the presence of TILs, have demonstrated value for predicting response to ICIs[27,166]. Immune cell populations are involved in tumour immune escape and are thus associated with a worse prognosis in many cancer types, for example, myeloid-derived suppressor cells (MDSCs)[268] in hepatocellular carcinoma[269] and bladder cancer[270], $T_{reg}$ cells[271] in colorectal[272] and lung[273,274] cancer and tumour-associated macrophages (TAMs)[275] in breast[276] and gastric[277] cancer. TCRs and BCRs are reported to be predictive biomarkers of response to ICIs[143–145]. For example, in patients with metastatic melanoma receiving ICIs, a high diversity in the T cell repertoire in peripheral blood samples (assembled and quantified using MiXCR) was correlated with the likelihood of having severe immune-related adverse events[278]. In another study involving patients with metastatic melanomas, those with a high level of TCR clonality in TILs before treatment (assessed with ImRep) had improved responses to anti-PD-1 antibodies[279].

Researchers have identified a wide variety of gene signatures, many of which are associated with TILs, reported to predict a response to ICIs (Box 2 and Supplementary Table 5). These signatures and methods are being adopted in ongoing clinical trials[280]; however, they need to be validated in experiments comparing samples obtained before and after receiving ICIs.

Prediction of whether a patient will respond to ICIs requires an approach that integrates clinical information with cancer-intrinsic

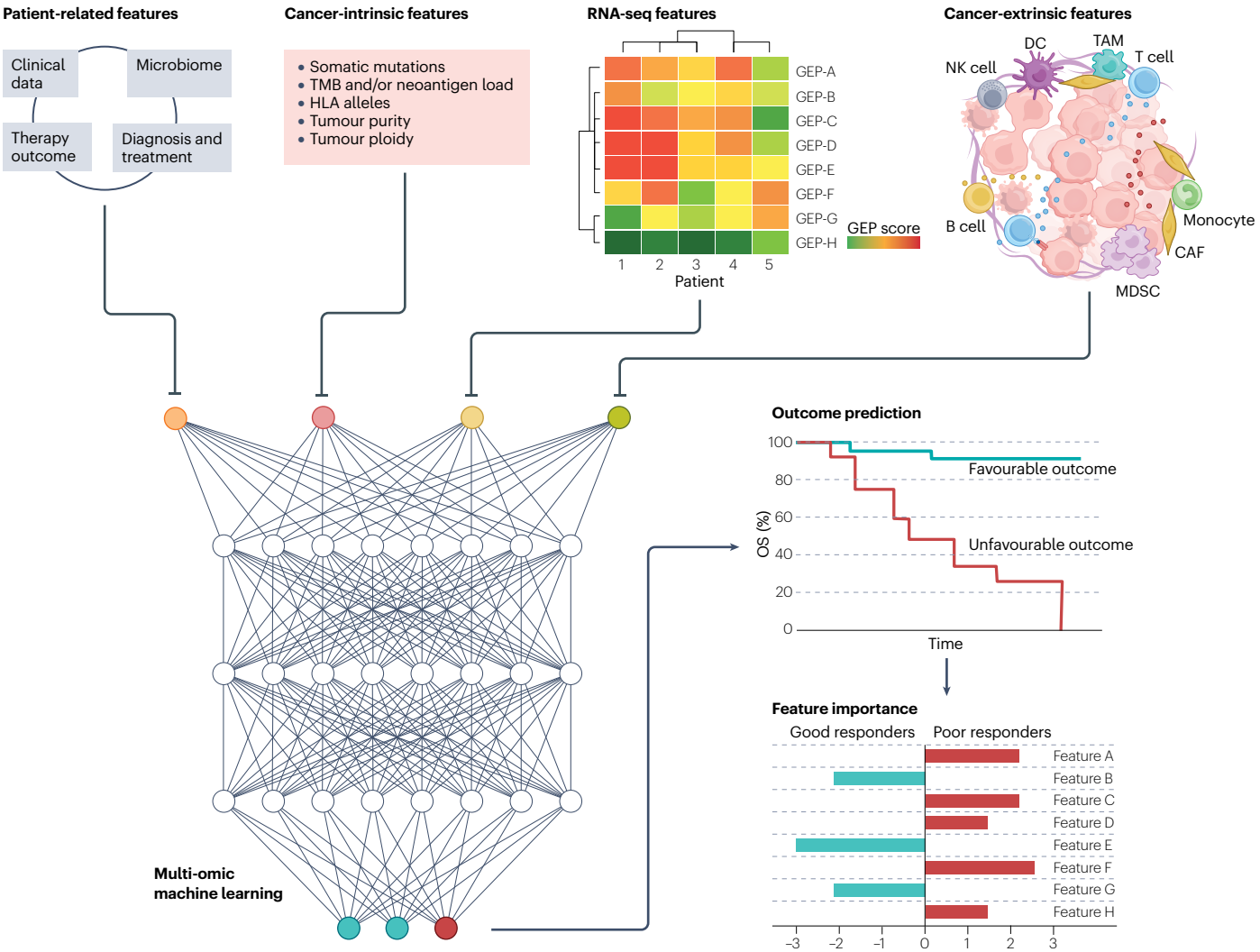



**Fig. 4 | Multi-omic machine learning to predict responses to immune-checkpoint inhibitors.** Multiple features from various sources such as clinical features, tumour somatic mutations, gene expression profile (GEP) or the proportion of deconvoluted immune and stromal cells can be used to develop and train machine learning models to predict response to immune-checkpoint inhibitors. The key discriminatory features linked to outcomes can be determined. CAF, cancer-associated fibroblast; DC, dendritic cell; MDSC, myeloid-derived suppressor cell; NK, natural killer; OS, overall survival; RNA-seq, RNA sequencing; TAM, tumour-associated macrophage; TMB, tumour mutational burden.

and cancer-extrinsic features (Fig. 4). Researchers have developed approaches to predict response to ICIs that incorporate multiple features. For example, a model using WES data that incorporates multiple DNA features (aneuploidy, CNA clonality, mutational signatures, TMB, presence of mutations in selected driver genes, intratumoural TCR clones, HLA alleles and neoantigen load) enabled the distinction of patients with NSCLC receiving ICIs with favourable or unfavourable OS[281]. Researchers performed a meta-analysis of genomic and transcriptomic biomarkers associated with response to ICIs in more than 3,600 patients across 12 tumour types and proposed a model, PredictIO, to identify potential targets, including *F2RL1* and *RBFOX2*, to overcome resistance to ICIs[282]. Litchfield et al.[166] compiled data from more than 1,000 ICI-treated patients with matched genomic and transcriptomic profiles across seven tumour types and constructed a multivariable model to predict response to ICIs for each cancer type using 11 features from this dataset to train and test a machine learning cancer-specific classifier, XGBoost. In this model, clonal TMB was the strongest predictor of response to ICIs, followed by total TMB and *CXCL9* expression. Furthermore, other factors, such as the patient's virome and gut microbiome, can influence response to ICIs[283,284].

Integrated multi-omic approaches that incorporate all relevant clinical and molecular features to predict response to ICIs require complex machine learning approaches. To date, the application of machine learning to analyse response to ICIs has been limited owing to insufficient availability of datasets; most approaches have focused on data from patients with melanoma or NSCLC. Liu et al.[28] applied a logistic regression-based classifier trained on clinical features plus treatment-naive genomic and transcriptomic profiles to predict

resistance (area under the curve (AUC) 0.73–0.83) to anti-PD-1 antibodies in patients with advanced-stage melanoma. Johannet et al.[285] adapted an advanced approach using a convolutional neural network trained and tested on clinical characteristics and histopathology slides from patients with advanced-stage melanoma and predicted responses to ICIs (AUC 0.80). The development of more-advanced machine learning models that use neural network approaches could provide a more complete overview of cancer–immune cell interactions when larger, multimodal datasets become available.

## Conclusions

Tumours and the immune system are in a constant delicate balance that affects host responses to ICIs. In the past few years, developments in computational immunogenomics have provided an unprecedented opportunity to study this complexity with the ultimate aim of improving patient outcomes. These improvements include, but are not limited to, enabling the selection of patients who are more likely to benefit from ICIs, developing highly effective neoantigen vaccines and improving disease monitoring by tracking immune responses in patients. Researchers can now use several sequencing techniques to profile the genome, transcriptome and epigenome and ultimately dissect the host tumour immune landscape. A high TMB and the abundance of immune cells (such as CTLs) are independently associated with response to ICIs; however, similar to the assessment of PD-L1 expression, each marker alone is imperfect. Researchers are developing various new computational approaches to study different cancer-intrinsic and cancer-extrinsic aspects. Although promising, many of these approaches will benefit from the availability of more patient data to train models and/or to validate their predictive performance.

Advances in molecular biology, nanotechnology, microfluidics and machine learning are enabling the deeper dissection of the TME. Researchers are using single-cell technologies to study the heterogeneity of the TME and the immune system to understand response or resistance to immunotherapies[220,286,287]. Continuous advances in single-cell analysis technologies to enable genome sequencing or full-length transcript sequencing[288] are expected to provide further insights into genomic ITH, and neoantigen formation and presentation at single-cell level. This approach has already generated accurate full-length antigen–receptor sequences from T and B cells[289].

The extent of immune activity can vary across the tumour site or across multiple lesions within the same patient. None of the existing computational tools or methods considers the site of biopsy or resection as a major factor in deconvoluting the TME or in understanding the mechanisms of immunoediting. Furthermore, with the promising efficacy of neoadjuvant ICIs[290], approaches to predict response to ICIs using biopsy samples from pretreated patients are essential. A more sophisticated model that integrates various factors including tumour purity, TME composition, genomic ITH, tumour evolution, genetic make-up and immunogenic neoantigen load is required. The development of more-holistic approaches that capture the entire tumour and integrate molecular and clinical data could reveal complete interactions between cancer cells and immune cells. Such integrated approaches could help to solve the challenge of predicting response to ICIs; however, models will need to be trained on large tumour-specific datasets and, given that these therapies tend to be combined with other agents, predictive models will probably require a more personalized approach.

Published online: 31 October 2023

## References


1. Sun, C. et al. Reversible and adaptive resistance to BRAF(V600E) inhibition in melanoma. *Nature* **508**, 118–122 (2014).
2. Buchbinder, E. I. & Desai, A. CTLA-4 and PD-1 pathways: similarities, differences, and implications of their inhibition. *Am. J. Clin. Oncol.* **39**, 98–106 (2016).
3. McDermott, D., Haanen, J., Chen, T. T., Lorigan, P. & O'Day, S. Efficacy and safety of ipilimumab in metastatic melanoma patients surviving more than 2 years following treatment in a phase III trial (MDX010-20). *Ann. Oncol.* **24**, 2694–2698 (2013).
4. Eggermont, A. M. M. et al. Prolonged survival in stage III melanoma with ipilimumab adjuvant therapy. *N. Engl. J. Med.* **375**, 1845–1855 (2016).
5. Hammers, H. J. et al. Safety and efficacy of nivolumab in combination with ipilimumab in metastatic renal cell carcinoma: the CheckMate 016 study. *J. Clin. Oncol.* **35**, 3851–3858 (2017).
6. Motzer, R. J. et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N. Engl. J. Med.* **373**, 1803–1813 (2015).
7. Overman, M. J. et al. Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. *Lancet Oncol.* **18**, 1182–1191 (2017).
8. Brahmer, J. et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 123–135 (2015).
9. Ansell, S. M. et al. PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N. Engl. J. Med.* **372**, 311–319 (2014).
10. Sharma, P. et al. Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. *Lancet Oncol.* **18**, 312–322 (2017).
11. El-Khoueiry, A. B. et al. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *Lancet* **389**, 2492–2502 (2017).
12. Chen, R. et al. Phase II study of the efficacy and safety of pembrolizumab for relapsed/refractory classic Hodgkin lymphoma. *J. Clin. Oncol.* **35**, 2125–2132 (2017).
13. Diaz, L. et al. 386P—efficacy of pembrolizumab in phase 2 KEYNOTE-164 and KEYNOTE-158 studies of microsatellite instability high cancers. *Ann. Oncol.* **28**, v128–v129 (2017).
14. Chung, H. C. et al. Efficacy and safety of pembrolizumab in previously treated advanced cervical cancer: results from the phase II KEYNOTE-158 study. *J. Clin. Oncol.* **37**, 1470–1478 (2019).
15. Zhu, A. X. et al. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. *Lancet Oncol.* **19**, 940–952 (2018).
16. Nghiem, P. et al. Durable tumor regression and overall survival in patients with advanced Merkel cell carcinoma receiving pembrolizumab as first-line therapy. *J. Clin. Oncol.* **37**, 693–702 (2019).
17. Alley, E. W. et al. Clinical safety and activity of pembrolizumab in patients with malignant pleural mesothelioma (KEYNOTE-028): preliminary results from a non-randomised, open-label, phase 1b trial. *Lancet Oncol.* **18**, 623–630 (2017).
18. Patel, S. P. & Kurzrock, R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Mol. Cancer Ther.* **14**, 847 (2015).
19. Duffy, M. J. & Crown, J. Biomarkers for predicting response to immunotherapy with immune checkpoint inhibitors in cancer patients. *Clin. Chem.* **65**, 1228–1238 (2019).
20. Fundytus, A., Booth, C. M. & Tannock, I. F. How low can you go? PD-L1 expression as a biomarker in trials of cancer immunotherapy. *Ann. Oncol.* **32**, 833–836 (2021).
21. Hellmann, M. D. et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N. Engl. J. Med.* **378**, 2093–2104 (2018).
22. Marabelle, A. et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol.* **21**, 1353–1365 (2020).
23. FDA. *FDA Approves Pembrolizumab for Adults and Children with TMB-H Solid Tumors*, https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-pembrolizumab-adults-and-children-tmb-h-solid-tumors (2020).
24. Hugo, W. et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* **165**, 35–44 (2016).
25. Ayers, M. et al. IFN-γ-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Investig.* **127**, 2930–2940 (2017).
26. Cristescu, R. et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* **362**, eaar3593 (2018).
27. Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207 (2015).
28. Liu, D. et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat. Med.* **25**, 1916–1927 (2019).
29. Rizvi, N. A. et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124 (2015).
30. Anagnostou, V. et al. Multimodal genomic features predict outcome of immune checkpoint blockade in non-small-cell lung cancer. *Nat. Cancer* **1**, 99–111 (2020).
31. McGranahan, N. et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463 (2016).
32. Kim, S. T. et al. Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer. *Nat. Med.* **24**, 1449–1458 (2018).
33. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).

34. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
35. Raine, K. M. et al. ascatNgs: identifying somatically acquired copy-number alterations from whole-genome sequencing data. *Curr. Protoc. Bioinformatics* **56**, 15.19.11–15.19.17 (2016).
36. Song, S. et al. qpure: a tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS ONE* **7**, e45835 (2012).
37. Yoshihara, K. et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612 (2013).
38. Revkov, E., Kulshrestha, T., Sung, K. W. & Skanderup, A. J. PUREE: accurate pan-cancer tumor purity estimation from gene expression data. *Commun. Biol.* **6**, 394 (2023).
39. Dentro, S. C. et al. Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. *Cell* **184**, 2239–2254.e39 (2021).
40. Frankell, A. M. et al. The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature* **616**, 525–533 (2023).
41. Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
42. Roth, A. et al. PyClone: statistical inference of clonal population structure in cancer. *Nat. Methods* **11**, 396–398 (2014).
43. Miller, C. A. et al. SciClone: inferring clonal architecture and tracking the spatial and temporal patterns of tumor evolution. *PLoS Comput. Biol.* **10**, e1003665 (2014).
44. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
45. Zaccaria, S. & Raphael, B. J. Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. *Nat. Commun.* **11**, 4301 (2020).
46. Xiao, Y. et al. FastClone is a probabilistic tool for deconvoluting tumor heterogeneity in bulk-sequencing samples. *Nat. Commun.* **11**, 4469 (2020).
47. Salcedo, A. et al. Crowd-sourced benchmarking of single-sample tumour subclonal reconstruction. Preprint at https://doi.org/10.1101/2022.06.14.495937 (2022).
48. Tanner, G., Westhead, D. R., Droop, A. & Stead, L. F. Benchmarking pipelines for subclonal deconvolution of bulk tumour sequencing data. *Nat. Commun.* **12**, 6396 (2021).
49. Martinez-Ruiz, C. et al. Genomic-transcriptomic evolution in lung cancer and metastasis. *Nature* **616**, 543–552 (2023).
50. Sha, D. et al. Tumor mutational burden as a predictive biomarker in solid tumors. *Cancer Discov.* **10**, 1808–1825 (2020).
51. Le, D. T. et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
52. Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
53. Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
54. Yarchoan, M., Hopkins, A. & Jaffee, E. M. Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* **377**, 2500–2501 (2017).
55. Newell, F. et al. Comparative genomics provides etiologic and biological insight into melanoma subtypes. *Cancer Discov.* **12**, 2856–2879 (2022).
56. Riaz, N. et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell* **171**, 934–949.e16 (2017).
57. Newell, F. et al. Multiomic profiling of checkpoint inhibitor-treated melanoma: identifying predictors of response and resistance, and markers of biological discordance. *Cancer Cell* **40**, 88–102.e107 (2022).
58. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
59. Strickler, J. H., Hanks, B. A. & Khasraw, M. Tumor mutational burden as a predictor of immunotherapy response: is more always better? *Clin. Cancer Res.* **27**, 1236–1241 (2021).
60. Kassahn, K. S. et al. Somatic point mutation calling in low cellularity tumors. *PLoS ONE* **8**, e74380 (2013).
61. Xiao, W. et al. Toward best practice in cancer mutation detection with whole-genome and whole-exome sequencing. *Nat. Biotechnol.* **39**, 1141–1150 (2021).
62. Akdemir, K. C. et al. Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure. *Nat. Genet.* **52**, 1178–1188 (2020).
63. Alioto, T. S. et al. A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. *Nat. Commun.* **6**, 10001 (2015).
64. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
65. Caron, N. R. et al. Indigenous genomic databases: pragmatic considerations and cultural contexts. *Front. Public Health* **8**, 111 (2020).
66. Carbone, D. P. et al. First-line nivolumab in stage IV or recurrent non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2415–2426 (2017).
67. Ricciuti, B. et al. Association of high tumor mutation burden in non-small cell lung cancers with increased immune infiltration and improved clinical outcomes of PD-L1 blockade across PD-L1 expression levels. *JAMA Oncol.* **8**, 1160–1168 (2022).
68. Aggarwal, C. et al. Assessment of tumor mutational burden and outcomes in patients with diverse advanced cancers treated with immunotherapy. *JAMA Netw. Open* **6**, e2311181 (2023).
69. Buchhalter, I. et al. Size matters: dissecting key parameters for panel-based tumor mutational burden analysis. *Int. J. Cancer* **144**, 848–858 (2019).
70. Ramarao-Milne, P. et al. Comparison of actionable events detected in cancer genomes by whole-genome sequencing, in silico whole-exome and mutation panels. *ESMO Open* **7**, 100540 (2022).
71. Merino, D. M. et al. Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project. *J. Immunother. Cancer* **8**, e000147 (2020).
72. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69 (2015).
73. Goodman, A. M. et al. Tumor mutational burden as an independent predictor of response to immunotherapy in diverse cancers. *Mol. Cancer Ther.* **16**, 2598 (2017).
74. Chan, T. A. et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Ann. Oncol.* **30**, 44–56 (2019).
75. Yamamoto, T. N., Kishton, R. J. & Restifo, N. P. Developing neoantigen-targeted T cell-based treatments for solid tumors. *Nat. Med.* **25**, 1488–1499 (2019).
76. Lee, C.-H., Yelensky, R., Jooss, K. & Chan, T. A. Update on tumor neoantigens and their utility: why it is good to be different. *Trends Immunol.* **39**, 536–548 (2018).
77. Chandran, S. S. et al. Immunogenicity and therapeutic targeting of a public neoantigen derived from mutated PIK3CA. *Nat. Med.* **28**, 946–957 (2022).
78. Luksza, M. et al. Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. *Nature* **606**, 389–395 (2022).
79. Wang, Q. L. et al. Association of HLA diversity with the risk of 25 cancers in the UK Biobank. *EBioMedicine* **92**, 104588 (2023).
80. Chowell, D. et al. Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy. *Nat. Med.* **25**, 1715–1720 (2019).
81. Chowell, D. et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582 (2018).
82. Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell lung cancer. *New Engl. J. Med.* **376**, 2109–2121 (2017).
83. McGranahan, N. et al. Allele-specific HLA loss and immune escape in lung cancer evolution. *Cell* **171**, 1259–1271.e11 (2017).
84. Bauer, D. C. & Thorne, N. P. et al. Evaluation of computational programs to predict HLA genotypes from genomic sequencing data. *Brief. Bioinform.* **19**, 179–187 (2016).
85. Yi, J., Chen, L., Xiao, Y., Zhao, Z. & Su, X. Investigations of sequencing data and sample type on HLA class Ia typing with different computational tools. *Brief. Bioinform.* **22**, bbaa143 (2021).
86. Kiyotani, K., Mai, T. H. & Nakamura, Y. Comparison of exome-based HLA class I genotyping tools: identification of platform-specific genotyping errors. *J. Hum. Genet.* **62**, 397–405 (2017).
87. Lang, K. et al. Full-length HLA class I genotyping with the MinION nanopore sequencer. *Methods Mol. Biol.* **1802**, 155–162 (2018).
88. Nielsen, M., Lundegaard, C., Lund, O. & Keşmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57**, 33–41 (2005).
89. Larsen, M. V. et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* **8**, 424 (2007).
90. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**, 511–517 (2016).
91. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
92. O'Donnell, T. J., Rubinsteyn, A. & Laserson, U. MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Syst.* **11**, 42–48 e47 (2020).
93. Vita, R. et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* **47**, D339–D343 (2018).
94. Vizcaíno, J. A. et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, 11033 (2016).
95. Shao, W. et al. The SysteMHC atlas project. *Nucleic Acids Res.* **46**, D1237–D1247 (2018).
96. Frahm, N. et al. Extensive HLA class I allele promiscuity among viral CTL epitopes. *Eur. J. Immunol.* **37**, 2419–2433 (2007).
97. Gfeller, D. & Bassani-Sternberg, M. Predicting antigen presentation—what could we learn from a million peptides? *Front. Immunol.* **9**, 1716 (2018).
98. Singh-Jasuja, H., Emmerich, N. P. & Rammensee, H. G. The Tubingen approach: identification, selection, and validation of tumor-associated HLA peptides for cancer therapy. *Cancer Immunol. Immunother.* **53**, 187–195 (2004).
99. Yadav, M. et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**, 572–576 (2014).
100. Xia, H. et al. Computational prediction of MHC anchor locations guides neoantigen identification and prioritization. *Sci. Immunol.* **8**, eabg2200 (2023).
101. Łuksza, M. et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* **551**, 517–520 (2017).
102. Balachandran, V. P. et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).
103. No authors listed. The problem with neoantigen prediction. *Nat. Biotechnol.* **35**, 97 (2017).
104. Jensen, K. K. et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **154**, 394–406 (2018).
105. Zhang, G. L. et al. MULTIPRED2: a computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles. *J. Immunol. Methods* **374**, 53–61 (2011).

106. Guan, P., Hattotuwagama, C. K., Doytchinova, I. A. & Flower, D. R. MHCPred 2.0: an updated quantitative T-cell epitope prediction server. *Appl. Bioinformatics* **5**, 55–61 (2006).
107. Chen, B. et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat. Biotechnol.* **37**, 1332–1343 (2019).
108. Wang, D. et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **15**, e8503 (2019).
109. Kalaora, S. et al. Combined analysis of antigen presentation and T-cell recognition reveals restricted immune responses in melanoma. *Cancer Discov.* **8**, 1366–1375 (2018).
110. Wells, D. K. et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell* **183**, 818 (2020).
111. Nowak, A. K. et al. Durvalumab with first-line chemotherapy in previously untreated malignant pleural mesothelioma (DREAM): a multicentre, single-arm, phase 2 trial with a safety run-in. *Lancet Oncol.* **21**, 1213–1223 (2020).
112. Zhao, J. et al. Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nat. Med.* **25**, 462 (2019).
113. Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
114. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
115. Bueno, R. et al. Comprehensive genomic analysis of malignant pleural mesothelioma identifies recurrent mutations, gene fusions and splicing alterations. *Nat. Genet.* **48**, 407–416 (2016).
116. Creaney, J. et al. Comprehensive genomic and tumour immune profiling reveals potential therapeutic targets in malignant pleural mesothelioma. *Genome Med.* **14**, 58 (2022).
117. Hmeljak, J. et al. Integrative molecular characterization of malignant pleural mesothelioma. *Cancer Discov.* **8**, 1548–1565 (2018).
118. Turajlic, S. et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* **18**, 1009–1021 (2017).
119. Mansfield, A. S. et al. Neoantigenic potential of complex chromosomal rearrangements in mesothelioma. *J. Thorac. Oncol.* **14**, 276–287 (2019).
120. Yang, W. et al. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat. Med.* **25**, 767–775 (2019).
121. Kahles, A. et al. Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* **34**, 211–224.e6 (2018).
122. Park, J. & Chung, Y. J. Identification of neoantigens derived from alternative splicing and RNA modification. *Genomics Inform.* **17**, e23 (2019).
123. Shen, L., Zhang, J., Lee, H., Batista, M. T. & Johnston, S. A. RNA transcription and splicing errors as a source of cancer frameshift neoantigens for vaccines. *Sci. Rep.* **9**, 14184 (2019).
124. Smart, A. C. et al. Intron retention is a source of neoepitopes in cancer. *Nat. Biotechnol.* **36**, 1056–1058 (2018).
125. Zhang, M. et al. RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nat. Commun.* **9**, 3919 (2018).
126. Barczak, W. et al. Long non-coding RNA-derived peptides are immunogenic and drive a potent anti-tumour response. *Nat. Commun.* **14**, 1078 (2023).
127. Smith, C. C. et al. Alternative tumour-specific antigens. *Nat. Rev. Cancer* **19**, 465–478 (2019).
128. Chai, S. et al. NeoSplice: a bioinformatics method for prediction of splice variant neoantigens. *Bioinform. Adv.* **2**, vbac032 (2022).
129. Pan, Y. et al. IRIS: discovery of cancer immunotherapy targets arising from pre-mRNA alternative splicing. *Proc. Natl Acad. Sci. USA* **120**, e2221116120 (2023).
130. Zhang, Z. et al. ASNEO: identification of personalized alternative splicing based neoantigens with RNA-seq. *Aging* **12**, 14633–14648 (2020).
131. Attig, J. et al. LTR retroelement expansion of the human cancer transcriptome and immunopeptidome revealed by de novo transcript assembly. *Genome Res.* **29**, 1578–1590 (2019).
132. Burbage, M. et al. Epigenetically controlled tumor antigens derived from splice junctions between exons and transposable elements. *Sci. Immunol.* **8**, eabm6360 (2023).
133. Merlotti, A. et al. Noncanonical splicing junctions between exons and transposable elements represent a source of immunogenic recurrent neo-antigens in patients with lung cancer. *Sci. Immunol.* **8**, eabm6359 (2023).
134. Shah, N. M. et al. Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. *Nat. Genet.* **55**, 631–639 (2023).
135. Shi, Y., Jing, B. & Xi, R. Comprehensive analysis of neoantigens derived from structural variation across whole genomes from 2528 tumors. *Genome Biol.* **24**, 169 (2023).
136. Vensko, S. P. et al. LENS: landscape of effective neoantigens software. *Bioinformatics* **39**, btad322 (2023).
137. Hu, Z., Ott, P. A. & Wu, C. J. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat. Rev. Immunol.* **18**, 168–182 (2018).
138. Lang, F., Schrors, B., Lower, M., Tureci, O. & Sahin, U. Identification of neoantigens for individualized therapeutic cancer vaccines. *Nat. Rev. Drug Discov.* **21**, 261–282 (2022).
139. Xie, N. et al. Neoantigens: promising targets for cancer therapy. *Signal Transduct. Target Ther.* **8**, 9 (2023).
140. Ott, P. A. et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217 (2017).
141. Sahin, U. et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).
142. Hilf, N. et al. Actively personalized vaccination trial for newly diagnosed glioblastoma. *Nature* **565**, 240–245 (2019).
143. Keskin, D. B. et al. Neoantigen vaccine generates intratumoral T cell responses in phase Ib glioblastoma trial. *Nature* **565**, 234–239 (2019).
144. Leidner, R. et al. Neoantigen T-cell receptor gene therapy in pancreatic cancer. *N. Engl. J. Med.* **386**, 2112–2119 (2022).
145. Rojas, L. A. et al. Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer. *Nature* **618**, 144–150 (2023).
146. Khattak, A. et al. Distant metastasis-free survival results from the randomized, phase 2 mRNA-4157-P201/KEYNOTE-942 trial. *J. Clin. Oncol.* **41**, LBA9503 (2023).
147. Rowley, J. D. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–293 (1973).
148. Tognon, C. et al. Expression of the ETV6–NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell* **2**, 367–376 (2002).
149. Seshagiri, S. et al. Recurrent R-spondin fusions in colon cancer. *Nature* **488**, 660–664 (2012).
150. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
151. Xiao, X., Garbutt, C. C., Hornicek, F., Guo, Z. & Duan, Z. Advances in chromosomal translocations and fusion genes in sarcomas and potential therapeutic applications. *Cancer Treat. Rev.* **63**, 61–70 (2018).
152. Rooper, L. M. et al. DEK-AFF2 carcinoma of the sinonasal region and skull base: detailed clinicopathologic characterization of a distinctive entity. *Am. J. Surg. Pathol.* **45**, 1682–1693 (2021).
153. Mackall, C. L. et al. A pilot study of consolidative immunotherapy in patients with high-risk pediatric sarcomas. *Clin. Cancer Res.* **14**, 4850–4858 (2008).
154. Poorebrahim, M. et al. TCR-like CARs and TCR-CARs targeting neoepitopes: an emerging potential. *Cancer Gene Ther.* **28**, 581–589 (2021).
155. Foy, S. P. et al. Non-viral precision T cell receptor replacement for personalized cell therapy. *Nature* **615**, 687–696 (2023).
156. Anderson, N. M. & Simon, M. C. The tumor microenvironment. *Curr. Biol.* **30**, R921–R925 (2020).
157. Grisaru-Tal, S., Rothenberg, M. E. & Munitz, A. Eosinophil–lymphocyte interactions in the tumor microenvironment and cancer immunotherapy. *Nat. Immunol.* **23**, 1309–1316 (2022).
158. Mensurado, S., Blanco-Dominguez, R. & Silva-Santos, B. The emerging roles of gammadelta T cells in cancer immunotherapy. *Nat. Rev. Clin. Oncol.* **20**, 178–191 (2023).
159. Pittet, M. J., Michielin, O. & Migliorini, D. Clinical relevance of tumour-associated macrophages. *Nat. Rev. Clin. Oncol.* **19**, 402–421 (2022).
160. Whiteside, T. L. The tumor microenvironment and its role in promoting tumor growth. *Oncogene* **27**, 5904 (2008).
161. Mihm, M. C., Clemente, C. G. & Cascinelli, N. Tumor infiltrating lymphocytes in lymph node melanoma metastases: a histopathologic prognostic indicator and an expression of local immune response. *Lab. Invest.* **74**, 43–47 (1996).
162. Balkwill, F. & Mantovani, A. Inflammation and cancer: back to Virchow? *Lancet* **357**, 539–545 (2001).
163. Kim, R., Emi, M. & Tanabe, K. Cancer immunoediting from immune surveillance to immune escape. *Immunology* **121**, 1–14 (2007).
164. TextXHarjunpää, H., Llort Asens, M., Guenther, C. & Fagerholm, S. C. Cell adhesion molecules and their roles and regulation in the immune and tumor microenvironment. *Front. Immunol.* **10**, 1078 (2019).
165. Limagne, E. et al. MEK inhibition overcomes chemoimmunotherapy resistance by inducing CXCL10 in cancer cells. *Cancer Cell* **40**, 136–152.e12 (2022).
166. Litchfield, K. et al. Meta-analysis of tumor- and T cell-intrinsic mechanisms of sensitization to checkpoint inhibition. *Cell* **184**, 596–614.e14 (2021).
167. Nagarsheth, N., Wicha, M. S. & Zou, W. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nat. Rev. Immunol.* **17**, 559–572 (2017).
168. Valkenburg, K. C., de Groot, A. E. & Pienta, K. J. Targeting the tumour stroma to improve cancer therapy. *Nat. Rev. Clin. Oncol.* **15**, 366–381 (2018).
169. Waldmann, T. A. Cytokines in cancer immunotherapy. *Cold Spring Harb. Perspect. Biol.* **10**, a028472 (2018).
170. Maleki Vareki, S. High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. *J. Immunother. Cancer* **6**, 157 (2018).
171. Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdagh, P. & De Preter, K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat. Commun.* **11**, 5650 (2020).
172. Tran, K. A. et al. Performance of tumour microenvironment deconvolution methods in breast cancer using single-cell simulated bulk mixtures. *Nat. Commun.* **14**, 5758 (2023).
173. Li, T. et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.* **77**, e108 (2017).
174. Mabbott, N. A., Baillie, J. K., Brown, H., Freeman, T. C. & Hume, D. A. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics* **14**, 632 (2013).
175. Nirmal, A. J. et al. Immune cell gene signatures for profiling the microenvironment of solid tumors. *Cancer Immunol. Res.* **6**, 1388–1400 (2018).

176. Racle, J. & Gfeller, D. EPIC: a tool to estimate the proportions of different cell types from bulk gene expression data. *Methods Mol. Biol.* **2120**, 233–248 (2020).
177. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
178. Becht, E. et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* **17**, 218 (2016).
179. Jimenez-Sanchez, A., Cast, O. & Miller, M. L. Comprehensive benchmarking and integration of tumor microenvironment cell estimation methods. *Cancer Res.* **79**, 6238–6246 (2019).
180. Finotello, F. et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* **11**, 34 (2019).
181. Thorsson, V. et al. The immune landscape of cancer. *Immunity* **48**, 812–830.e14 (2018).
182. Petitprez, F. et al. B cells are associated with survival and immunotherapy response in sarcoma. *Nature* **577**, 556–560 (2020).
183. M, M. N. et al. Multi-omic features of oesophageal adenocarcinoma in patients treated with preoperative neoadjuvant therapy. *Nat. Commun.* **14**, 3155 (2023).
184. Chakravarthy, A. et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat. Commun.* **9**, 3220 (2018).
185. Houseman, E. A. et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).
186. Demerath, E. W. et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum. Mol. Genet.* **24**, 4464–4479 (2015).
187. Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **12**, 529–541 (2011).
188. Koestler, D. C. et al. Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). *BMC Bioinformatics* **17**, 120 (2016).
189. Salas, L. A. et al. Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nat. Commun.* **13**, 761 (2022).
190. Teschendorff, A. E., Breeze, C. E., Zheng, S. C. & Beck, S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. *BMC Bioinforma.* **18**, 105 (2017).
191. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
192. Elizabeth Larose, C. et al. Copy number-aware deconvolution of tumor-normal DNA methylation profiles. Preprint at *bioRxiv*, https://doi.org/10.1101/2020.11.03.366252 (2020).
193. Flusberg, B. A. et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).
194. Simpson, J. T. et al. Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* **14**, 407–410 (2017).
195. Sturm, G. et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* **35**, i436–i445 (2019).
196. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).
197. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).
198. Jew, B. et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).
199. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat. Cancer* **3**, 505–517 (2022).
200. Carroll, T. M. et al. Tumor monocyte content predicts immunochemotherapy outcomes in esophageal adenocarcinoma. *Cancer Cell* **41**, 1222–1241.e27 (2023).
201. Menden, K. et al. Deep learning-based cell composition analysis from tissue expression profiles. *Sci. Adv.* **6**, eaba2619 (2020).
202. Clarke, Z. A. et al. Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nat. Protoc.* **16**, 2749–2764 (2021).
203. Heumos, L. et al. Best practices for single-cell analysis across modalities. *Nat. Rev. Genet.* **24**, 550–572 (2023).
204. Ahlmann-Eltze, C. & Huber, W. Comparison of transformations for single-cell RNA-seq data. *Nat. Methods* **20**, 665–672 (2023).
205. Chen, W. et al. A multicenter study benchmarking single-cell RNA sequencing technologies using reference samples. *Nat. Biotechnol.* **39**, 1103–1114 (2021).
206. Mallory, X. F., Edrisi, M., Navin, N. & Nakhleh, L. Assessing the performance of methods for copy number aberration detection from single-cell DNA sequencing data. *PLoS Comput. Biol.* **16**, e1008012 (2020).
207. Nguyen, H. C. T., Baik, B., Yoon, S., Park, T. & Nam, D. Benchmarking integration of single-cell differential expression. *Nat. Commun.* **14**, 1570 (2023).
208. Zafar, H., Wang, Y., Nakhleh, L., Navin, N. & Chen, K. Monovar: single-nucleotide variant detection in single cells. *Nat. Methods* **13**, 505–507 (2016).
209. Morita, K. et al. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat. Commun.* **11**, 5327 (2020).
210. Bohrson, C. L. et al. Linked-read analysis identifies mutations in single-cell DNA-sequencing data. *Nat. Genet.* **51**, 749–754 (2019).
211. Dong, X. et al. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat. Methods* **14**, 491–493 (2017).
212. Huang, Z. et al. Single-cell analysis of somatic mutations in human bronchial epithelial cells in relation to aging and smoking. *Nat. Genet.* **54**, 492–498 (2022).
213. Wu, C. Y. et al. Integrative single-cell analysis of allele-specific copy number alterations and chromatin accessibility in cancer. *Nat. Biotechnol.* **39**, 1259–1269 (2021).
214. Broad Institute. *inferCNV of the Trinity CTAT Project*, https://github.com/broadinstitute/inferCNV (2020).
215. Muller, S., Cho, A., Liu, S. J., Lim, D. A. & Diaz, A. CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. *Bioinformatics* **34**, 3217–3219 (2018).
216. Dominguez, C. X. et al. Single-cell RNA sequencing reveals stromal evolution into LRRC15$^{+}$ myofibroblasts as a determinant of patient response to cancer immunotherapy. *Cancer Discov.* **10**, 232–253 (2020).
217. Joanito, I. et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. *Nat. Genet.* **54**, 963–975 (2022).
218. Kumar, M., Bowers, R. R. & Delaney, J. R. Single-cell analysis of copy-number alterations in serous ovarian cancer reveals substantial heterogeneity in both low- and high-grade tumors. *Cell Cycle* **19**, 3154–3166 (2020).
219. Kumar, V. et al. Single-cell atlas of lineage states, tumor microenvironment, and subtype-specific expression programs in gastric cancer. *Cancer Discov.* **12**, 670–691 (2022).
220. Wu, F. et al. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **12**, 2540 (2021).
221. Liu, Y. et al. Tumour heterogeneity and intercellular networks of nasopharyngeal carcinoma at single cell resolution. *Nat. Commun.* **12**, 741 (2021).
222. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
223. Demeulemeester, J. et al. Tracing the origin of disseminated tumor cells in breast cancer using single-cell sequencing. *Genome Biol.* **17**, 250 (2016).
224. Gawad, C., Koh, W. & Quake, S. R. Dissecting the clonal origins of childhood acute lymphoblastic leukemia by single-cell genomics. *Proc. Natl Acad. Sci. USA* **111**, 17947–17952 (2014).
225. Kester, L. et al. Integration of multiple lineage measurements from the same cell reconstructs parallel tumor evolution. *Cell Genom.* **2**, 100096 (2022).
226. Navin, N. et al. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
227. Malikic, S., Jahn, K., Kuipers, J., Sahinalp, S. C. & Beerenwinkel, N. Integrative inference of subclonal tumour evolution from single-cell and bulk sequencing data. *Nat. Commun.* **10**, 2750 (2019).
228. Satas, G., Zaccaria, S., Mon, G. & Raphael, B. J. SCARLET: single-cell tumor phylogeny inference with copy-number constrained mutation losses. *Cell Syst.* **10**, 323–332.e8 (2020).
229. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308.e36 (2018).
230. Kieffer, Y. et al. Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. *Cancer Discov.* **10**, 1330–1351 (2020).
231. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
232. Schelker, M. et al. Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. *Nat. Commun.* **8**, 2032 (2017).
233. Wu, T. D. et al. Peripheral T cell expansion predicts tumour infiltration and clinical response. *Nature* **579**, 274–278 (2020).
234. Yost, K. E. et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. *Nat. Med.* **25**, 1251–1259 (2019).
235. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
236. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
237. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
238. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
239. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
240. Zheng, L. et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* **374**, abe6474 (2021).
241. Alquicira-Hernandez, J., Sathe, A., Ji, H. P., Nguyen, Q. & Powell, J. E. scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* **20**, 264 (2019).
242. Zhang, A. W. et al. Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nat. Methods* **16**, 1007–1015 (2019).
243. Miao, Z. et al. Putative cell type discovery from single-cell gene expression data. *Nat. Methods* **17**, 621–628 (2020).
244. Brbic, M. et al. MARS: discovering novel cell types across heterogeneous single-cell experiments. *Nat. Methods* **17**, 1200–1206 (2020).
245. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
246. Qiu, X. et al. Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315 (2017).
247. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

248. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).
249. Wolf, F. A. et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019).
250. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
251. Cannoodt, R. et al. SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. Preprint at *bioRxiv*, https://doi.org/10.1101/079509 (2016).
252. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
253. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
254. Giladi, A. et al. Dissecting cellular crosstalk by sequencing physically interacting cells. *Nat. Biotechnol.* **38**, 629–637 (2020).
255. Chandran, S. S. & Klebanoff, C. A. T cell receptor-based cancer immunotherapy: emerging efficacy and pathways of resistance. *Immunol. Rev.* **290**, 127–147 (2019).
256. Pai, J. A. & Satpathy, A. T. High-throughput and single-cell T cell receptor sequencing technologies. *Nat. Methods* **18**, 881–892 (2021).
257. Jin, X. et al. Identification of shared characteristics in tumor-infiltrating T cells across 15 cancers. *Mol. Ther. Nucleic Acids* **32**, 189–202 (2023).
258. Song, L. et al. Comprehensive characterizations of immune receptor repertoire in tumors and cancer immunotherapy studies. *Cancer Immunol. Res.* **10**, 788–799 (2022).
259. Oliveira, G. & Wu, C. J. Dynamics and specificities of T cells in cancer immunotherapy. *Nat. Rev. Cancer* **23**, 295–316 (2023).
260. Pai, J. A. et al. Lineage tracing reveals clonal progenitors and long-term persistence of tumor-specific T cells during immune checkpoint blockade. *Cancer Cell* **41**, 776–790.e77 (2023).
261. Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
262. Marcou, Q., Mora, T. & Walczak, A. M. High-throughput immune repertoire analysis with IGoR. *Nat. Commun.* **9**, 561 (2018).
263. Song, L. et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat. Methods* **18**, 627–630 (2021).
264. Kugel, C. H. III et al. Age correlates with response to anti-PD1, reflecting age-related differences in intratumoral effector and regulatory T-cell populations. *Clin. Cancer Res.* **24**, 5347–5356 (2018).
265. Jang, S. R. et al. Association between sex and immune checkpoint inhibitor outcomes for patients with melanoma. *JAMA Netw. Open* **4**, e2136823 (2021).
266. Olateju, O. A. et al. Investigation of racial differences in survival from non-small cell lung cancer with immunotherapy use: a Texas study. *Front. Oncol.* **12**, 1092355 (2022).
267. Davoli, T., Uno, H., Wooten, E. C. & Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355**, eaaf8399 (2017).
268. Gabrilovich, D. I. & Nagaraj, S. Myeloid-derived suppressor cells as regulators of the immune system. *Nat. Rev. Immunol.* **9**, 162–174 (2009).
269. Arihara, F. et al. Increase in $CD14^{+}HLA\text{-}DR^{-/low}$ myeloid-derived suppressor cells in hepatocellular carcinoma patients and its impact on prognosis. *Cancer Immunol. Immunother.* **62**, 1421–1430 (2013).
270. Yang, G. et al. Accumulation of myeloid-derived suppressor cells (MDSCs) induced by low levels of IL-6 correlates with poor prognosis in bladder cancer. *Oncotarget* **8**, 38378–38388 (2017).
271. Facciabene, A., Motz, G. T. & Coukos, G. T-regulatory cells: key players in tumor immune escape and angiogenesis. *Cancer Res.* **72**, 2162–2171 (2012).
272. Saito, T. et al. Two $FOXP3^{+}CD4^{+}$ T cell subpopulations distinctly control the prognosis of colorectal cancers. *Nat. Med.* **22**, 679–684 (2016).
273. Petersen, R. P. et al. Tumor infiltrating $Foxp3^{+}$ regulatory T-cells are associated with recurrence in pathologic stage I NSCLC patients. *Cancer* **107**, 2866–2872 (2006).
274. Shimizu, K. et al. Tumor-infiltrating $Foxp3^{+}$ regulatory T cells are correlated with cyclooxygenase-2 expression and are associated with recurrence in resected non-small cell lung cancer. *J. Thorac. Oncol.* **5**, 585–590 (2010).
275. Pathria, P., Louis, T. L. & Varner, J. A. Targeting tumor-associated macrophages in cancer. *Trends Immunol.* **40**, 310–327 (2019).
276. Tiainen, S. et al. High numbers of macrophages, especially M2-like (CD163-positive), correlate with hyaluronan accumulation and poor outcome in breast cancer. *Histopathology* **66**, 873–883 (2015).
277. Zhang, H. et al. Infiltration of diametrically polarized macrophages predicts overall survival of patients with gastric cancer after surgical resection. *Gastric Cancer* **18**, 740–750 (2015).
278. Lozano, A. X. et al. T cell characteristics associated with toxicity to immune checkpoint blockade in patients with melanoma. *Nat. Med.* **28**, 353–362 (2022).
279. Valpione, S. et al. The T cell receptor repertoire of tumor infiltrating T cells is predictive and prognostic for cancer survival. *Nat. Commun.* **12**, 4098 (2021).
280. NCI. *Cancer Immunotherapy Clinical Trials*, https://www.cancer.gov/about-cancer/treatment/clinical-trials/intervention/pembrolizumab (2022).
281. Richard, C. et al. Exome analysis reveals genomic markers associated with better efficacy of nivolumab in lung cancer patients. *Clin. Cancer Res.* **25**, 957–966 (2019).
282. Bareche, Y. et al. Leveraging big data of immune checkpoint blockade response identifies novel potential targets. *Ann. Oncol.* **33**, 1304–1317 (2022).
283. Khan, M. A. W., Ologun, G., Arora, R., McQuade, J. L. & Wargo, J. A. Gut microbiome modulates response to cancer immunotherapy. *Dig. Dis. Sci.* **65**, 885–896 (2020).
284. Lecuelle, J. et al. MER4 endogenous retrovirus correlated with better efficacy of anti-PD1/PD-L1 therapy in non-small cell lung cancer. *J. Immunother. Cancer* **10**, e004241 (2022).
285. Johannet, P. et al. Using machine learning algorithms to predict immunotherapy response in patients with advanced melanoma. *Clin. Cancer Res.* **27**, 131–140 (2021).
286. Gohil, S. H., Iorgulescu, J. B., Braun, D. A., Keskin, D. B. & Livak, K. J. Applying high-dimensional single-cell technologies to the analysis of cancer immunotherapy. *Nat. Rev. Clin. Oncol.* **18**, 244–256 (2021).
287. Gonzalez-Silva, L., Quevedo, L. & Varela, I. Tumor functional heterogeneity unraveled by scRNA-seq technologies. *Trends Cancer* **7**, 265 (2021).
288. Grimes, S. M. Single-cell multi-gene identification of somatic mutations and gene rearrangements in cancer. *NAR Cancer* **5**, zcad034 (2023).
289. Singh, M. et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. *Nat. Commun.* **10**, 3120 (2019).
290. Patel, S. P. et al. Neoadjuvant-adjuvant or adjuvant-only pembrolizumab in advanced melanoma. *N. Engl. J. Med.* **388**, 813–823 (2023).
291. Benson, D. A. et al. GenBank. *Nucleic Acids Res.* **41**, D36–D42 (2013).
292. Sidhom, J.-W. et al. ImmunoMap: a bioinformatics tool for T-cell repertoire analysis. *Cancer Immunol. Res.* **6**, 151–162 (2018).
293. Shugay, M. et al. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput. Biol.* **11**, e1004503 (2015).
294. Canzar, S., Neu, K. E., Tang, Q., Wilson, P. C. & Khan, A. A. BASIC: BCR assembly from single cells. *Bioinformatics* **33**, 425–427 (2017).
295. Lindeman, I. et al. BraCeR: B-cell-receptor reconstruction and clonality inference from single-cell RNA-seq. *Nat. Methods* **15**, 563–565 (2018).
296. Mandric, I. et al. Profiling immunoglobulin repertoires across multiple human tissues using RNA sequencing. *Nat. Commun.* **11**, 3126 (2020).
297. Carter, J. A., Gilbo, P. & Atwal, G. S. IMPRES does not reproducibly predict response to immune checkpoint blockade therapy in metastatic melanoma. *Nat. Med.* **25**, 1833–1835 (2019).
298. Auslander, N., Lee, J. S. & Ruppin, E. Reply to: 'IMPRES does not reproducibly predict response to immune checkpoint blockade therapy in metastatic melanoma'. *Nat. Med.* **25**, 1836–1838 (2019).
299. Thompson, J. C. et al. Gene signatures of tumor inflammation and epithelial-to-mesenchymal transition (EMT) predict responses to immune checkpoint blockade in lung cancer with high accuracy. *Lung Cancer* **139**, 1–8 (2020).
300. Raghav, K. et al. Efficacy, safety, and biomarker analysis of combined PD-L1 (atezolizumab) and VEGF (bevacizumab) blockade in advanced malignant peritoneal mesothelioma. *Cancer Discov.* **11**, 2738–2747 (2021).
301. Spranger, S., Bao, R. & Gajewski, T. F. Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. *Nature* **523**, 231–235 (2015).
302. Sharma, P. et al. Nivolumab plus ipilimumab for metastatic castration-resistant prostate cancer: preliminary analysis of patients in the CheckMate 650 trial. *Cancer Cell* **38**, 489–499.e3 (2020).
303. Ott, P. A. et al. T-cell-inflamed gene-expression profile, programmed death ligand 1 expression, and tumor mutational burden predict efficacy in patients treated with pembrolizumab across 20 cancers: KEYNOTE-028. *J. Clin. Oncol.* **37**, 318–327 (2019).
304. Mariathasan, S. et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**, 544–548 (2018).
305. Powles, T. et al. Clinical efficacy and biomarker analysis of neoadjuvant atezolizumab in operable urothelial carcinoma in the ABACUS trial. *Nat. Med.* **25**, 1706–1714 (2019).
306. Charoentong, P. et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* **18**, 248–262 (2017).
307. Rosenthal, R. et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).
308. Jiang, P. et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat. Med.* **24**, 1550–1558 (2018).
309. Campbell, K. M. et al. Prior anti-CTLA-4 therapy impacts molecular characteristics associated with anti-PD-1 response in advanced melanoma. *Cancer Cell* **41**, 791–806.e4 (2023).
310. Prat, A. et al. Immune-related gene expression profiling after PD-1 blockade in non-small cell lung carcinoma, head and neck squamous cell carcinoma, and melanoma. *Cancer Res.* **77**, 3540–3550 (2017).
311. Bentham, R. et al. Using DNA sequencing data to quantify T cell fraction and therapy response. *Nature* **597**, 555–560 (2021).
312. Lu, T. et al. Netie: inferring the evolution of neoantigen-T cell interactions in tumors. *Nat. Methods* **19**, 1480–1489 (2022).



## Acknowledgements

N.W. is supported by the Australian National Health and Medical Research Council Research Fellowship. The authors thank K. Tran, R. L. Johnston and M. W. L. Teng (all at QIMR Berghofer Medical Research Institute) for intellectual input and helpful discussions.


## Author contributions

V.A. and N.W. wrote the manuscript. All authors researched data for the manuscript, made substantial contributions to discussion of the content and edited and/or reviewed the manuscript before submission.

## Competing interests
J.V.P. and N.W. are co-founders of genomiQa. The other authors declare no competing interests.

## Additional information
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41571-023-00830-6.

**Peer review information** *Nature Reviews Clinical Oncology* thanks L. Penter and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



## Related links
**10x genomics**: https://10xgenomics.github.io/enclone/
**Immunarch**: https://immunarch.com/
**NetChop 3.1**: https://services.healthtech.dtu.dk/services/NetChop-3.1/
**NetMHC 4.0**: https://services.healthtech.dtu.dk/services/NetMHC-4.0/
**scRNA-tools**: https://www.scrna-tools.org/