Synthetic biomarkers: a twenty-first century path to early cancer detection

Gabriel A. Kwong[1,2,3,4,5,✉], **Sharmistha Ghosh**[6,✉], **Lena Gamboa**[1], **Christos Patriotis**[6], **Sudhir Srivastava**[6,✉], **Sangeeta N. Bhatia**[7,8,9,10,11,✉]

[1]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory School of Medicine, Atlanta, GA, USA.

[2]Parker H. Petit Institute of Bioengineering and Bioscience, Atlanta, GA, USA.

[3]Institute for Electronics and Nanotechnology, Georgia Tech, Atlanta, GA, USA.

[4]The Georgia Immunoengineering Consortium, Emory University and Georgia Tech, Atlanta, GA, USA.

[5]Winship Cancer Institute, Emory University, Atlanta, GA, USA.

[6]Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

[7]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA.

[8]Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA.

[9]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA.

[10]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.

[11]Howard Hughes Medical Institute, Chevy Chase, MD, USA.

Abstract

Detection of cancer at an early stage when it is still localized improves patient response to medical interventions for most cancer types. The success of screening tools such as cervical cytology to reduce mortality has spurred significant interest in new methods for early detection (for example, using non-invasive blood-based or biofluid-based biomarkers). Yet biomarkers shed from early lesions are limited by fundamental biological and mass transport barriers — such as short circulation times and blood dilution — that limit early detection. To address this issue, synthetic

✉: gkwong@gatech.edu; ghoshjanjigias@mail.nih.gov; srivasts@mail.nih.gov; sbhatia@mit.edu.

biomarkers are being developed. These represent an emerging class of diagnostics that deploy bioengineered sensors inside the body to query early-stage tumours and amplify disease signals to levels that could potentially exceed those of shed biomarkers. These strategies leverage design principles and advances from chemistry, synthetic biology and cell engineering. In this Review, we discuss the rationale for development of biofluid-based synthetic biomarkers. We examine how these strategies harness dysregulated features of tumours to amplify detection signals, use tumour-selective activation to increase specificity and leverage natural processing of bodily fluids (for example, blood, urine and proximal fluids) for easy detection. Finally, we highlight the challenges that exist for preclinical development and clinical translation of synthetic biomarker diagnostics.

The earliest stages of cancer detection are when our existing clinical interventions can be more successful. Detecting pre-invasive tumours before clinical symptoms appear is likely to enhance the effect of medical interventions such as surgical resection, which can be curative for most types of localized cancers that have not metastasized[1]. When accurate tests are available, risk-based cancer screening of populations is recommended by regulatory agencies, and contributes to lowering cancer deaths. Examples include mammography for breast cancer, colonoscopy for colorectal cancer, Papanicolaou test (Pap smear) for cervical cancer and low-dose chest computed tomography for those at high risk of lung cancer[2–7]. However, accurate tests based on imaging and/or non-invasive analysis of patient fluids such as blood are not available for the vast majority of cancer types, and the diagnostic specificity of current tests is insufficient to allow routine screening of asymptomatic segments of the population where the cancer prevalence is low. A test with low positive predictive value would lead to an unacceptably high percentage of false positives and unnecessary medical interventions, precluding broad deployment. The continuing debate over whether the only widely used blood biomarker test, the prostate-specific antigen (PSA) test, is useful for reducing prostate cancer mortality despite its drawbacks (such as unnecessary treatments, patient morbidity and costs) serves as an important lesson for future tests[8].

There are several ongoing efforts towards detecting other endogenous biomarkers (for example, cell-free nucleic acids, proteins, lipids and metabolites) via analysis of blood and other biofluids[9–17]. Significant strides have been made with sequencing of cancer genes from circulating tumour DNA (ctDNA), as evidenced by the recent success of a multianalyte, multicancer test in a prospective study of women without a history of cancer in which the feasibility of using a blood test to detect multiple cancers was established[9,14]. However, biological and technical challenges remain obstacles to the early detection of cancer, especially before symptoms are apparent; a test with high sensitivity would be required to detect very low signal levels, but such a test must not contribute substantially to the overdiagnosis of inconsequential cancers. The expression or release of biomarkers is variable and compounded by interpatient variation, tumour heterogeneity, comorbidities and background secretion by healthy cells. Moreover, individual biomarkers often lack specificity because their levels can be elevated in non-cancerous conditions, as in the case of DNA mutations from non-malignant clonal haematopoiesis of indeterminate potential (CHIP)[18] or PSA level increase from benign prostatic hyperplasia[19], or they are shed across many types of cancer, as is the case for carcinoembryonic antigen (CEA), the level of

which is elevated in cancers of the colon, breast, lung and other organs[20]. This necessitates identifying multianalyte panels that combine different classes of biomarkers into a single predictive score to assess the presence of disease and localize the cancer to anatomical sites[9,14,21].

These lessons learned are informing the design of an emerging class of diagnostics based on the design of bioengineered sensors — such as molecular probes or genetically encoded vectors — that exploit dysregulated features of early-stage tumours or their precursors, which have the potential to become lethal, to produce an amplified signal that cancer cells would otherwise not produce or would produce at undetectable levels. These exogenously administered sensors harness tumour-dependent activation mechanisms such as enzymatic amplification to drive the production and amplification of synthetic biomarkers. Cancers can also be detected by imaging systems that may share essential features of a synthetic biomarker approach, such as reporter gene imaging, whereby an exogenous molecular tracer (for example, a positron-emitting probe) is systemically infused[22,23]. Imaging will also play an essential role in detecting the location of the tumour following confirmation of a detectable synthetic biomarker signal. However, as advances in cancer imaging have been extensively reviewed elsewhere[22,24], this Review focuses on synthetic biomarkers detectable from biofluids such as blood and urine. First, we highlight the challenges associated with early cancer detection that have motivated research into synthetic biomarkers. We then review advances in activity-based and genetically encoded sensors, which are the two major strategies being used for synthetic biomarker development. Finally, we discuss the challenges that exist for this growing field in the setting of preclinical studies and strategies for clinical translation.

The challenge of early detection

The rationale for synthetic biomarker development comes from the biological, physiological and mathematical limitations of endogenous biomarkers (FIG. 1). For continuously shed biomarkers such as proteins, patient tumours are not universally biomarker positive, and secretion rates can vary by as much as four orders of magnitude, even for cells of the same tumour type[25]. Moreover, biomarkers that are released only by dead or dying cells are shed just once, and their detection is confounded by background shedding from healthy tissues. Cell-free DNA (cfDNA), for example, is released from non-cancerous cells throughout the body, which makes the proportion of somatic mutations in malignant cells versus normal cells, or the variant allele frequency (VAF), increasingly difficult to detect at low tumour burdens. Analysis of the Tracking Non-Small-Cell Lung Cancer Evolution Through Therapy (TRACERx) study predicted that primary tumour burdens of 1 cm^3, 10 cm^3 or 100 cm^3 would result in average clonal plasma VAFs of 0.006%, 0.1% or 1.3%, respectively[26]. For a typical 4 ml of plasma from a 10-ml blood draw and a VAF of 0.1%, it has been estimated that there would be an average of just six molecules per tube carrying the respective somatic mutation[27]. Further compounding the technical challenge, shed biomarkers are diluted by a large pool of blood (~5 l) and circulate for short periods owing to degradation or clearance; ctDNA, for example, has a circulation half-life of less than 1.5 h in blood[28,29]. By comparison, the resolution of clinical positron emission tomography (PET)-based molecular

imaging (specifically using Jaszczak phantoms and fluorine-18) has been reported[30] as ~200 mm[3], which is equivalent to a tumour diameter of ~7 mm.

Despite the technical challenges associated with detecting shed biomarkers, mathematical model predictions and genomic timeline studies consistently estimate a window of opportunity for early cancer detection that may span at least a decade. Work on multicompartment models[31–34] to understand the relationship between tumour volumes and shed biomarker levels has resulted in predictions that tumours could remain undetectable for more than 10 years following initiation of tumorigenesis[32]. Genomic studies on cancer progression timelines have estimated periods of ~7 years or more from the birth of a founder carcinoma cell to macrometastatic tumours, given the inherent inefficiency of individual tumour cells to seed and survive in distant organs[35–38]. It is important to keep in mind, however, that cancers that are present for a decade or more are likely to be of indolent nature and eventually detectable at some point by existing screening modalities that are not based on biofluids. By contrast, fast-growing and highly aggressive cancers — including interval cancers that are diagnosed during the time between a regular screening that appears normal and the next screening — may rapidly progress within a relatively narrow window of months to years and are associated with poor clinical outcomes. Examples include triple-negative breast cancer and high-grade serous ovarian carcinoma (HGSOC) in women whose tumours have *BRCA1* or *BRCA2* mutations, or homologous recombination deficiency[39,40]. Detecting such aggressive cancers at an early stage would likely require identification of cancer precursors (such as serous tubal intraepithelial carcinoma for HGSOC) and the development of new ultrasensitive approaches that permit increased frequency of testing. Advances that are occurring in the field of synthetic biomarker research aim to address these challenges, with the main approaches being those that leverage activity-based or genetically encoded mechanisms for early detection.

Multicompartment models

A mathematical modelling technique whereby distinct compartments are used to represent organs, tissues, blood or lymph to predict how an administered drug is absorbed, distributed, metabolized or excreted.

Activity-based synthetic biomarkers

The systemic administration of exogenous agents to assess biological function in vivo has a long clinical history. Examples include infusion of patients with inulin, which is an inert polysaccharide that is not digestible or absorbed, to measure kidney function[41] and indocyanine green, a fluorescent dye, to quantify liver dysfunction[42]. These biomarkers and other similar tests target known features of human physiology (for example, plasma clearance via hepatocytes) or established disease mechanisms with a biologically inert probe to produce a readout that is not normally found in the body, thereby maximizing the signal-to-noise ratio. Activity-based synthetic biomarkers are based on this paradigm but include sensor components that are activated by enzymes in the tumour or its microenvironment to provide a mechanism for molecular amplification of tumour biomarkers. Here we discuss the

key design considerations for activity-based synthetic biomarkers, with a particular focus on protease-activated sensors and small-molecule probes.

Protease-activated synthetic biomarkers.

The human genome encodes more than 550 proteases, and their dysregulation has broad implications at the molecular level (for example, protein activation and matrix degradation), cellular level (for example, immune cell cytotoxicity and apoptosis) and systems level (for example, cancer-induced hypercoagulable state) in cancer[43]. For example, matrix metalloproteinases (MMPs) are overexpressed across the vast majority of cancer types[44] as one of their key functions is to regulate the bioavailability of vascular endothelial growth factor (VEGF) during the angiogenic 'switch', a process that occurs when nascent tumours reach 1–2 mm in diameter and require increased access to blood nutrients to overcome diffusion-limited growth[45,46]. Recent studies have also highlighted that dysregulated protease expression can be used for predictive cancer classification, for example, separating prostate cancer into aggressive and indolent phenotypes[47] using machine learning algorithms. Another study showed that protease transcript signatures can differentiate lung adenocarcinoma from interstitial lung disease or chronic obstructive pulmonary disease[48], demonstrating the potential of protease-based classifiers for differential diagnosis.

Protease-activated synthetic biomarkers comprise peptide substrates conjugated to the surface of an inert carrier[25,47–51] that upon enzymatic cleavage by tumour proteases release reporters into the blood or urine for detection (FIG. 2). Proteases are particularly potent molecular amplifiers because hydrolysis of peptide bonds is irreversible and proteases are not consumed during peptidolysis, thereby allowing a single copy to turn over thousands of substrates[52]. In addition to molecular amplification, another key strategy to attain the limit of detection (LOD) required for early detection involves harnessing features of human physiology to increase synthetic biomarker concentration in biofluids. One approach is to take advantage of size filtration by the kidneys by selecting a carrier with a hydrodynamic radius larger than the ~5-nm size cut-off of the glomerular filtration barrier[53] to prevent surface-conjugated peptides from being cleared into urine. Production of detection signals occurs after intravenous administration when the peptides are cleaved from the surface of the carrier by tumour proteases, releasing synthetic biomarkers into the circulation that are then rapidly cleared into urine for detection based on their reduced hydrodynamic diameters[25,47–49,51,54]. Although the use of a nanoparticle carrier increases the circulation time of surface-conjugated peptides, one limitation is the reliance on passive delivery to tumour sites. Approaches that use carriers with smaller hydrodynamic diameters, such as polyethylene glycol (PEG) polymers, which are characterized by higher passive diffusion rates than larger, iron oxide nanoparticles (IONPs)[55], could increase delivery to tumours[51]. Another approach is to functionalize sensors with tumour-penetrating ligands that engage active trafficking pathways to the tumour microenvironment[44].

Hydrodynamic radius

For a macromolecule in solution, the radius of an equivalent hard sphere diffusing at the same rate as the macromolecule.

Proteases are promiscuous enzymes capable of cleaving a variety of substrate sequences, which limits the detection specificity of a single sensor. Therefore, another key design principle is to design a multiplexed library of sensors to detect cancer by signature analysis. This approach requires each synthetic biomarker within a cocktail to be labelled with a unique molecular barcode. Various strategies have been developed, including mass barcodes[25,47,48,56], whereby reporters are differentially enriched with stable isotopes such as ^{13}C to generate a unique mass detectable by tandem mass spectrometry; DNA barcodes[57], whereby each reporter is labelled with a unique DNA sequence for detection by sequencing, PCR or CRISPR–Cas; ligand-encoded reporters[50,58,59], which are labelled with small molecules for detection by antibodies; volatile organic compounds[60] that are emitted as gases after cleavage; and ultrasmall gold nanoclusters[61] to catalyse a colorimetric readout. The most densely multiplexed cocktails of sensors reported to date are mass-barcoded 14-plex[48,56] or 19-plex[47] systems. These densely multiplexed approaches allow classifiers to be trained on the basis of multivariate machine learning algorithms that have the potential to indicate disease with increased diagnostic sensitivity and specificity compared with univariate classifiers trained on a single biomarker. Indeed, development and utilization of machine learning approaches have great potential for expediting the development of synthetic biomarkers for a variety of clinical applications. Machine learning or deep learning methods (BOX 1), when powered correctly along with considerations for comorbid conditions and other confounders, can help to increase the signal-to-noise ratio. In addition, deconvolution of complex signatures also has the potential to reveal new biological insights.

Small-molecule probes.

In light of the increasing number of tumour-specific antigens, cell surface markers and metabolic pathways that are targetable with small molecules, a number of studies are emerging that focus on engineered molecular probes to generate synthetic biomarkers for cancer detection (FIG. 3). Nishihara and colleagues[62] reported a strategy to generate synthetic biomarkers by targeting cancer cell-surface lectins using a two-step strategy. First, they labelled LoVo human colorectal carcinoma cells with a protein conjugate composed of the enzyme β-galactosidase conjugated to avidin. Avidin is a positively charged protein that contains terminal *N*-acetylglucosamine and mannose residues that bind to lectins overexpressed by tumour cells. In a second step, they administered a substrate for β-galactosidase called 'β-galactosidase-responsive acetaminophen' that is converted into acetaminophen (also known as paracetamol) by exogenous β-galactosidase on the surface of tumour cells. They found that acetaminophen plasma levels generated in this two-step process were elevated within 60 minutes in tumour-bearing mice. A similar approach was reported to quantify H_2O_2 activity using H_2O_2-responsive acetaminophen.[63]

Small molecules labelled with stable isotopes have been widely used as diagnostic probes in research laboratories for more than 30 years[64]. The advantages of stable isotope labelling include the lack of radiation risk to patients, indistinguishable metabolism compared with their unmodified counterparts and high signal-to-noise ratio owing to the lack of background signal. Among the first clinically approved tests was the ^{13}C-urea test for *Helicobacter pylori* that detects urease activity central to *H. pylori* metabolism and virulence based on the level of $^{13}CO_2$ released in the breath[65]. Several isotope-labelled probes, including ^{13}C-

methacetin[66] and ^{13}C-cholate[67], which measure hepatic cytochrome P450 activity and liver shunting, respectively, are approved by the FDA for measurement of liver dysfunction in the context of liver fibrosis, which is an important risk factor for hepatocellular carcinoma.

Natural volatile organic compounds (VOCs) present in patient breath samples have also been investigated for cancer diagnosis[68,69]. However, identifying a VOC signature for cancer is not trivial because of the high variability and low concentration of natural VOCs in breath. Lange and colleagues[70] used an isotope-labelled synthetic VOC called 'D$_5$-ethyl-β-D-glucuronide' (EtGlu), which is a deuterated metabolite of ethanol. Following intravenous administration, EtGlu is enzymatically converted by β-glucuronidase, an extracellular enzyme secreted by solid tumours, into D$_5$-ethanol, which is then detected from the breath by gas chromatography coupled with high-resolution mass spectrometry. In various tumour models, including a transgenic mouse model of mammary tumours, Lange and colleagues found significantly increased D$_5$-ethanol levels in breath following a single injection of EtGlu, and used D$_5$-ethanol levels to monitor the response to chemotherapy. In the future, isotope-labelled probes have the potential to be expanded into synthetic agents that sense different classes of tumour enzymes, including proteases[60]; they could also be densely multiplexed by mass to allow rapid analysis of breath samples for synthetic biomarkers.

Deuterated metabolite

A compound in which one or more hydrogen atoms have been replaced by the stable isotope deuterium to distinguish it from its unmodified counterpart.

Genetically encoded synthetic biomarkers

Design-driven advances in mammalian synthetic biology are pushing the boundaries for biological sensing. In addition to activity-based probes, genetically encoded constructs form the other major group of strategies that use engineered components or cells to amplify the release of synthetic biomarkers. These methods focus on strategies that drive resident cells or infiltrated cells within the tumour microenvironment to produce or secrete bio-orthogonal reporters[71–77]. The main advantage of these approaches is the ability to transcriptionally target synthetic biomarker production to cells of a particular phenotype, thereby potentially reducing the number of false positives caused by background production in healthy tissues. Here, we review advances in three main classes of genetically encoded systems for producing synthetic biomarkers, namely vector-based, mammalian cell-based and bacterial cell-based systems (FIG. 4).

Bio-orthogonal reporters

Non-native reporters that do not interfere with biological functions.

Vector-based synthetic biomarkers.

Transcriptional targeting with gene vectors is a powerful method to restrict transgene expression in target tissues and has been extensively explored for cancer imaging and

therapy[71,78]. Building upon this foundation, vector-based systems rely on two key design components: a tissue-selective or cancer-selective promoter to drive transcription and a synthetic biomarker designed to be secreted into blood or urine for detection[79,80]. Tissue-selective promoters provide the first level of specificity — for example, the promoter for the gene pulmonary surfactant-associated protein B (*SFTPB*) restricts transgene expression to alveolar type II cells and Clara cells of the lung[81], and similarly, use of the promoter for the gene glial fibrillary acidic protein (*GFAP*) restricts expression almost exclusively to astrocytes[82]. However, with this approach, systemic delivery will result in transgene production by normal as well as tumour tissue derived from the same cell type, thereby increasing the background signal. By comparison, cancer-selective promoters increase the precision of transcriptional targeting as these are driven primarily by cancer cells but have limited activity in normal cells. One example is the promoter for the normally silent human telomerase reverse transcriptase (*TERT*), which encodes telomerase, which is frequently activated in cancer cells to achieve proliferative immortality[83], one of the hallmarks of cancer. As *TERT* is expressed at high levels in ~90% of human cancers but silenced in almost all somatic cells, the *TERT* promoter has been used to drive expression of genes in a wide variety of tumour cells[84–86].

The second component of vector-based strategies is a secreted reporter that acts as synthetic biomarker and can be detected in blood or urine. Secreted embryonic alkaline phosphatase (SEAP) was among the first reporters to be engineered for applications in vivo. SEAP is an engineered form of human placental alkaline phosphatase that contains a termination codon at the membrane-anchoring domain to convert it into a truncated but fully active secreted reporter[87,88]. In xenograft tumour models, production of SEAP by cancer cells allowed early and long-term measurement of tumour growth and response to drug treatment as SEAP levels directly correlated with tumour size and cell numbers[89–91]. The limitations associated with SEAP include its high molecular mass (64 kDa), which limits its use to a synthetic blood biomarker as it is not normally excreted in urine[92]. Moreover, alkaline phosphatases are naturally expressed by major organs and may leak into the bloodstream through tissue injury and interfere with SEAP measurements. Another commonly used reporter is a luciferase cloned from the marine copepod *Gaussia princeps* (Gluc)[93–98]. Unlike earlier luciferases such as *Photinus pyralis* luciferase (Fluc) and *Renilla reniformis* luciferase (Rluc), Gluc is naturally secreted and is among the smallest luciferases at 19.9 kDa, and its initial activity per mole is about 100–1,000 times higher than that of Rluc or Fluc and it is more than 20,000-fold more sensitive than SEAP[93]. On the basis of these favourable properties, Wurdinger and colleagues[94] demonstrated that Gluc could detect as few as 1,000 Gli36 human glioma cells in vivo compared with a LOD of ~500,000 cells with use of SEAP.

One limitation of vector-based strategies is the requirement for efficient tumour delivery without the use of viral vectors, given concerns regarding immunogenicity and insertion mutagenesis, particularly for early detection applications that will require longitudinal assessment and repeated administrations. Fang and colleagues[99] designed plasmid vectors that were charge-complexed with cationic polyethylenimine for detection of bladder cancer. These constructs used cancer-selective promoters from cyclooxygenase 2 (*Cox2*) and osteopontin (*Opn*) to drive production of Gluc for detection from urine samples

in mice. Although plasmids have a superior safety profile compared with viral vectors, they are limited by low gene transfer rates and transient expression profiles. By contrast, DNA 'minicircles'[100], which are minimal vectors free of prokaryotic components that conform to regulatory principles for plasmids free of antibiotic resistance genes (pFAR)[101], have increased delivery efficiency, enhanced expression and reduced transcriptional silencing compared with plasmids. For cancer detection, Ronald and colleagues[77] designed minicircles that encoded SEAP driven by the cancer-selective survivin promoter. Systemic delivery of these in a melanoma lung metastasis mouse model led to detectable elevations of SEAP levels in plasma that correlated with tumour burden. With this approach, it would be possible to create bespoke vectors for particular cancer types by designing minicircles with alternative promoters, such as the mucin 1 promoter for breast cancer[102]. In addition, their application could be extended beyond early cancer detection, for example, to assess the aggressiveness of prostate cancer[103].

Mammalian cell-based synthetic biomarkers.

The recent clinical successes of adoptive cell therapies have inspired the idea of engineered mammalian cells as living biosensors (FIG. 4). A clear advantage of cells as diagnostic vehicles is that some are capable of homing to and infiltrating cancer sites, in contrast to molecular probes, which are limited by their reliance on passive diffusion from the vasculature to accumulate in tumours. Mesenchymal stem cells (MSCs) are adult multipotent stem cells that possess regenerative and immunomodulatory properties, and systemically infused MSCs selectively home to primary and metastatic tumours[104]. Liu and colleagues[105] used a mouse model to demonstrate the use of engineered MSCs for detection of cancer metastasis from blood. First, MSCs were engineered to secrete humanized Gluc; upon intravenous administration, engineered MSCs persisted longer in mice with MDA-MB-231 breast cancer lung metastases than in tumour-free mice, resulting in higher blood levels of humanized Gluc. However, as MSCs exhibit tropism to sites of inflammation and injury[106], or may themselves participate in cancer progression[107], additional studies are needed to understand these potential limitations.

Aalipour and colleagues[76] further developed the concept of cell-based diagnostics using engineered macrophages as living cellular sensors. Within the tumour microenvironment, a subset of macrophages is polarized to an M2 tumour-associated metabolic profile that promotes an immunosuppressive microenvironment. Aalipour and colleagues found that M2 reprogramming led to striking changes in the levels of arginase 1 (encoded by *ARG1*), which was upregulated by as much as 200-fold by adoptively transferred macrophages in solid tumours. On the basis of this finding, they used the *ARG1* promoter to drive production of Gluc upon macrophage M2 polarization. This study laid the foundation for the concept of cellular immunodiagnostics, and considering that a number of other immune cells likewise modulate expression of metabolic genes in the context of the tumour microenvironment, this approach could also be extended to T cells[108], B cells[109] and natural killer cells[110]. Several limitations are worth noting, including observations that macrophage sensors did not detect visibly necrotic tumours in high tumour burden settings, which could be attributed to poor infiltration. Another limitation is the high cost of adoptive cell transfer and the complex pipeline for good manufacturing practice (GMP) cell manufacture that

would prevent this approach from being a routine screening tool. However, numerous efforts are under way to reduce the time and cost, including in situ reprogramming of circulating cells[111], which circumvents the need for ex vivo cell isolation, and allogeneic 'off-the-shelf' immune cells[112].

Bacterial cell-based synthetic biomarkers.

Certain types of bacteria infiltrate and selectively grow in tumours, which has been attributed to suppressed immunosurveillance and increased levels of nutrients released by necrotic cells within the core of solid tumours[113–116]. This has prompted the use of engineered tumour-targeting bacteria as programmable vehicles for cancer detection. Panteli and colleagues[117,118] genetically modified an attenuated strain of *Salmonella enterica* that is 10,000-fold less toxic than its wild-type counterpart to release ZsGreen as a fluorescent biomarker, or 'fluoromarker'. Following intravenous administration in tumour-bearing mice, fluoromarker levels in serum were dependent on tumour mass and were predicted by mathematical modelling to have the capacity to detect tumours as small as 120 mg. Danino and colleagues[119] showed that the nonvirulent probiotic bacterium *Escherichia coli* Nissle 1917, genetically engineered with a *lacZ* reporter, selectively colonizes colorectal cancer liver metastases following oral delivery in recipient mice. One limitation of gene circuits constructed on intracellular plasmids is that they lose stability and function over time and under environmental conditions that disrupt cellular homeostasis. Therefore, the team engineered a dual-maintenance vector including a toxin–antitoxin system that simultaneously produces a long-lived host-killing (hok) toxin and a short-lived suppression of killing (sok) antitoxin, such that in the event of plasmid loss, the cell will be killed by the long-lived toxin. To detect the presence of liver metastases, the team showed that a LacZ substrate could be administered to produce a colorimetric reporter in urine. A demonstrated advantage of this approach is the ability of tumour-targeting bacteria to expand by more than 10^6-fold after colonization, providing yet another mechanism to amplify detection signals beyond enzymatic turnover and urinary enrichment.

Moving forward, several challenges need to be addressed for bacteria to be used for early cancer detection. Although engineered strains, including *Clostridium*, *E. coli* Nissle and *Salmonella*, have been shown to be non-pathogenic in animals and humans[113], the inherent toxicity of bacterial components and the potential to revert to virulence pose safety concerns. It is also not clear whether all tumour types and nascent lesions that lack a necrotic core can be colonized by systemically delivered bacteria. Advances in synthetic biology could offer solutions to mitigate these challenges as well as providing the opportunity to engineer 'smart' micro-organisms with specified and controlled behaviour[120]. For instance, bacteria engineered with quorum-sensing biocircuits can be used for bacterial communication to synchronize activity[121,122] and produce emergent behaviour such as timed release of therapeutic cargo after a threshold population density has been reached to either kill tumours[123] or promote systemic anti-tumour immunity[124,125]. Applied to the field of early cancer detection, these biocircuits have the potential to increase specificity by reducing background activity from healthy tissues, since off-target bacteria would not reach a quorum and therefore not falsely produce a reporter. In the future, these genetically programmable vehicles may have the potential to be developed into safe and regularly

ingested food products (for example, yogurt) to allow routine cancer screening or cancer chemoprevention[126].

Preclinical studies

A number of preclinical studies have been reported that demonstrate the potential of activity-based synthetic biomarkers to achieve the LOD required for earlier detection (FIG. 5). In a xenograft mouse model, an activity-based sensor composed of IONPs conjugated with mass-barcoded peptide substrates was able to discriminate LS174T colorectal tumours that were 60% smaller in volume than those detected by the shed serum biomarker CEA (130 mm^3 versus 330 mm^3 on average, respectively) with an area under the receiver operating characteristic curve (AUROC) of 0.94 (REF.[25]). By contrast, a separate study by Aalipour et al.[76] that also used the LS174T colorectal cancer model found that ctDNA was detectable from blood only when tumour volumes reached ~1,000 mm^3. Kwon and colleagues[44] reported a formulation of activity-based sensors that incorporated tumour-penetrating peptides to target and increase their delivery to metastatic nodules in an orthotopic ovarian cancer model to further lower the LOD. By quantifying cleaved synthetic biomarkers enriched in urine, they reported the ability to detect disseminated disease with near-perfect accuracy (AUROC of 0.99) when the median nodule diameter was less than 2 mm and the average total tumour burden was 36 mm^3. By comparison, the shed human epididymis protein 4 (HE4) serum biomarker was able to indicate disease only when the average tumour burden reached 88 mm^3. This 59% reduction in tumour burden LOD was an important demonstration considering that current transvaginal ultrasonography can reliably resolve individual tumour nodules only when they are larger than 5 mm in diameter (equivalent to 65 mm^3 per nodule), and estimates indicate that decreasing serous ovarian cancer mortality by 50% would require a test capable of detecting nodules smaller than 5 mm in diameter[127].

Depending on the cancer type, use of different delivery routes also provides another approach to reduce the LOD by reducing sensor activation by off-target organs. Kirkpatrick and colleagues[48] showed that intrapulmonary delivery of a 14-plex cocktail of sensors, composed of mass-barcoded peptides conjugated to an eight-arm PEG carrier, could be used to query the lungs for early tumours by producing cleaved synthetic biomarkers detectable in urine. In the *Kras*- and *Trp53*-mutant genetically engineered mouse model of lung adenocarcinoma, they reported the ability to detect a total average tumour burden of 2.8 mm^3. This LOD compared favourably with that in an independent publication by Rakhit and colleagues[128], where they showed that ctDNA from an autochthonous $Kras^{G12D}$-mutant lung cancer model was detectable only when average tumour volumes were 7.1 mm^3. Kirkpatrick and colleagues further showed that a random forest machine learning classifier (BOX 1) trained on the 14-plex synthetic biomarker signature predicted lung cancer progression with high accuracy (AUROC greater than 0.90) and distinguished lung cancer-bearing mice from mice with benign lung inflammation (AUROC greater than 0.97)[48].

In vivo LOD studies have also been reported for genetically encoded synthetic biomarkers. Aalipour and colleagues[76] showed that the adoptive transfer of engineered macrophage sensors detected moderately sized CT26 colorectal tumours (volume 50–250 mm^3) with 100% sensitivity and specificity, while small tumours (25–50 mm^3) were also discriminated,

with an AUROC of ~0.85, compared with healthy animals. The team further demonstrated the potential for translation by using primary bone marrow-derived macrophages in addition to the RAW264.7 macrophage cell line. They found that engineered bone marrow-derived macrophage sensors detected CT26 tumours with a volume of 60–75 mm^3 with an AUROC of 0.81. In benchmarking studies comparing the performance of their macrophage sensors with either plasma CEA secreted by LS174T tumours or cfDNA released by CT26 tumours, they reported a lower LOD; tumours ~136 mm^3 in volume were detectable by CEA, while tumours larger than 1,500 mm^3 were detectable by cfDNA.

Challenges for clinical translation

Preclinical limitations.

Significant sources of noise for synthetic biomarker strategies include off-target and on-target, off-tumour activation. For activity-based synthetic biomarkers, most published protease substrates were identified through in vitro selection; therefore, these substrates were not selected against background activity arising from circulating blood (for example, coagulation and complement proteases) or organ-associated proteases in vivo. Therefore, it will be important to develop screening strategies that permit substrate discovery by negative selection under healthy as well as comorbid conditions. Ideally, the substrate development pipeline would include steps conducted in vivo or, at a minimum, with appropriate control plasma samples in vitro that account for the anticoagulant used during sample collection and the classes of proteases it inhibits. Developing peptide display technologies that permit sequence selection based on on-target and off-target protease activity in vivo would significantly advance the design of peptide-based protease sensors[129,130].

Standardized and better preclinical models are needed to accurately recapitulate pre-invasive and early-stage cancer. The vast majority of immortalized cancer cell lines were derived from patients with advanced metastatic disease, which do not fully reflect early or pre-cancerous conditions. Moreover, the rate at which endogenous biomarkers are produced by these cancer cell lines can vary by as much as four orders of magnitude[25], which makes benchmarking studies difficult to compare across laboratories unless the same cell lines are used. Additional methods need to be developed to increase the information we are able to collect from patients, including the 'age' of a tumour[131–134], the relationship between tumour sizes and secreted biomarker levels, and the permeability of tumours. This increased understanding will provide important clinical data to support the development and validation of predictive mathematical models and to optimize formulations of synthetic biomarkers[51]. Genetically engineered animal models that recapitulate pre-invasive conditions, such as prostatic intraepithelial neoplasia (PIN)[135], would provide a rich test bed for future synthetic biomarker studies geared towards early detection of cancers.

Allometric scaling.

It is likely that several key system parameters will be linearly proportional between preclinical rodent models and humans. For example, for protease-activated synthetic biomarkers, it is estimated that more than 500 of 628 mouse proteases are considered true orthologues[136] of the ~550 proteases encoded by the human genome. Thus, the efficiency

with which a protease cleaves a substrate sequence (that is, the catalytic efficiency k_{cat}/K_m, where k_{cat} is the catalytic rate constant and K_m is the Michaelis–Menten constant[137]) would likely be similar between rodent and human orthologues, especially for proteases that perform conserved functions. For others that are substantially different, substrate screening technologies could be used to identify target substrates with similar Michaelis–Menten constants between species. Therefore, it is likely that the kinetics of protease cleavage and signal amplification observed in mouse models would be reflected in humans. Similar assumptions could be drawn for other parameters, such as tumour transfection efficiencies, biomarker secretion and degradation rates, and safety and clearance from the body.

However, there are significant physiological differences between mice and humans, including blood pool volume (2 ml versus 5 l), urine volume (500 μl versus 500 ml) and glomerular filtration rates, such that allometric scaling across species would likely be non-linear. For example, Kwong and co-authors[51] developed a physiologically based pharmacokinetic model to understand how probe and physiological parameters affect the performance of activity-based synthetic biomarkers. Their model revealed a number of intuitive relationships (for example, signal is proportional to sensor delivery) but also predicted relationships that were non-linear and non-intuitive. Several of these non-linear relationships (for example, signal-to-noise ratios are largely independent of the dose of the administered sensor) have been experimentally validated in mice[44,51,58] but have yet to be shown in humans. Moreover, synthetic biomarkers shed by genetically encoded vectors into the circulation will be diluted by ~3,500-fold in humans over mice if scaling is calculated linearly on the basis of only blood volume. Yet this does not imply that a synthetic biomarker that can discriminate ~5-mm^3 tumours in mice can discriminate only tumours that are ~3,500-fold larger in volume in humans, as both clinical data and mathematical modelling support that smaller tumour sizes are detectable even by shed endogenous blood biomarkers[31,32]. Clearly, biological factors other than tumour burden affect detection sensitivity.

Tumour localization.

For cancer screening applications, a blood or urine synthetic biomarker has limited utility unless it also reports on which organ should be followed up for tumour localization. One approach could involve signal normalizers such as a probe that reports on organ-specific proteases (for example, liver hepsin) or, alternatively, normalization against a synthetic biomarker released by a tissue-specific promoter. Another potential approach is to combine synthetic biomarkers with different classes of endogenous analytes and clinical variables such that a multianalyte classifier can be trained to predict potential tumour sites. This strategy was recently demonstrated by CancerSEEK[9], a blood test designed to detect eight common cancer types by ctDNA sequencing. The test included 31 proteins and the patient's sex to generate a score that correctly localized the tumour to one of the two top predicted anatomical sites in 83% of patients. A similar approach was reported for a stool-based test for colon cancer screening that included a haemoglobin immunoassay and was able to detect significantly more cancers than a faecal immunochemical test alone[21].

Imaging will play a critical role in determining the location of the tumour following confirmation of a synthetic biomarker signal. Molecular imaging with reporter genes is a rich area of research for platforms such as single photon emission computed tomography (SPECT) and PET that are not limited by depth or tumour site within the body compared with optical modalities[138]. Genetically encoded synthetic biomarkers are most amenable to these approaches, which essentially involve exchanging the secreted synthetic biomarker for a reporter gene. For example, macrophage sensors could be engineered to express the herpes simplex virus 1 thymidine kinase (HSV1-TK) reporter gene to allow tumour site-induced M2 polarization of macrophages to be detected by PET[22]. Similar approaches have also been demonstrated with vector-based strategies; for example, the tumour-specific progression elevated gene 3 (*Peg3*) promoter (PEG-Prom) has been used to drive HSV1-TK expression, enabling tumour-specific imaging of lung metastasis[71]. This approach showed the ability to detect small lesions that were missed by fluorodeoxyglucose PET in preclinical studies. Imaging strategies that integrate concepts from synthetic biology have the potential to further increase the sensitivity and specificity of cancer imaging, as illustrated by Widen and colleagues[139], who described an AND gate optical probe requiring two cleavage events by multiple tumour proteases to produce a signal.

Herpes simplex virus 1 thymidine kinase

(HSV1-TK). The enzyme expressed by the reporter gene phosphorylates radiolabelled purine and pyrimidine nucleoside analogues to trap the probe within cells and thereby allow visualization by positron emission tomography (PET).

AND gate

A Boolean logic gate operation that outputs a value of 1 if and only if both inputs are 1; otherwise it outputs 0.

Strategies for clinical testing.

A densely multiplexed cocktail of synthetic probes would likely be necessary to achieve the selectivity required to handle the expected tumour heterogeneity, interpatient variations and comorbidities in the human population. Importantly, in the reported preclinical studies with activity-based synthetic biomarkers[25,47,48], the study authors showed that the diagnostic performance of two or three probes was sufficient to attain the sensitivity and specificity of the entire panel (more than ten probes). Moreover, when the same multiplexed probe set was used, different subsets of probes distinguished different disease states, including liver fibrosis progression from regression[25], lung cancer from benign lung inflammation[48] and response from resistance to checkpoint blockade immunotherapy[56]. While the ability to distinguish disease states in mice with low-dimensional data can be attributed partially to the lack of variation in isogenic tumour models, these observations also highlight a potential strategy for clinical trial design that makes use of a 'superset' of probes. This approach may provide the ability to capture high-dimensional data for classifier training (BOX 1), while allowing the possibility of down-selection after classifier validation. In addition, once

the safety and immunogenicity of a superset of probes have been demonstrated, it could potentially be applied to different clinical use cases without changing its composition, which could reduce the amount of resources required and regulatory burden. Similar specificity challenges lie ahead for genetically encoded sensors, which thus far have been designed to produce a single synthetic biomarker. One approach to increase the specificity of tumour detection would be to multiplex several metabolic gene reporters. This could be accomplished with different promoters paired with different secreted reporters, such as artificial microRNAs[140]. Cell-based sensors could also be engineered with synthetic circuitry to endow them with the capacity to perform logic-based computations. This could take the form of biocircuits that require the presence of multiple environmental inputs before a single output reporter is produced, for example, using AND-gated sensing[141–143] to increase tumour-selective activation or analogue-to-digital conversion[143] to reduce background noise.

As the field is in its infancy, human testing of synthetic biomarkers has yet to proceed to pivotal trials to determine their use for early cancer detection. To the best of the authors' knowledge, the synthetic biomarker formulation that has advanced the furthest in clinical testing is mass-barcoded PEGylated peptides[48], which were found to be well tolerated and safe in healthy human volunteers on the basis of preliminary data from a recent phase I study[144,145]. The first several clinical use cases for synthetic biomarkers need to be carefully considered, as an early failure may set the field back. Screening asymptomatic patients for early cancer is highly challenging and may present ethical challenges in clinical trial validation studies; for example, patients with a positive synthetic biomarker test result may need to wait for confirmation by imaging (that is, allow tumours to grow) before therapeutic intervention. Potential clinical entry points such as pharmacodynamic assessment of treatment response[56] or monitoring for recurrence following primary resection could show the utility of a synthetic biomarker approach before transitioning to early detection applications. As the field advances towards human testing, it should be noted that a number of components and vehicles that form the 'parts' for synthetic biomarker generation are undergoing clinical evaluation, have a demonstrated safety record in humans or are approved by the FDA. Examples include protease-activated substrates used in imaging probes for intraoperative detection of tumour margins[146–148], linkers for masked antibodies (NCT03993379 and NCT03013491)[149–151], activation domains for T cell engagers[152,153] (NCT03577028) and antibody–drug conjugates[154]. Similarly, for genetically encoded synthetic biomarkers, numerous clinical trials have highlighted the safety and utility of attenuated bacteria as vehicles for targeting tumours and delivering therapy[155]. These precedents provide a broader understanding of the embodiments of synthetic biomarkers that will be safe and well tolerated by humans, and anticipate the types of substrate sequences that will be selective for different types of human tumours.

Charting the course ahead

Although the nascent field of synthetic biomarkers is exciting and full of promise, there are gaps in our current knowledge of cancer pathogenesis that need to be filled alongside addressing technical challenges to guide future advances. In particular, there is limited understanding of the biology of early lesions and when and how a precursor lesion

transitions into malignancy, yet such information is needed to guide sensor engineering strategies. This highlights a challenge in the synthetic biomarker field for early detection of cancers. The Cancer Genome Atlas (TCGA) has generated a tremendous knowledge base for the biomedical community but there is a bias towards advanced and locally advanced tumours. Therefore, there is a need for extensive profiling of early-stage and in situ tumours as well as lethal precursors that have a high propensity for malignancy. There is also an urgent unmet clinical need to detect aggressive cancers, and early detection efforts would be greatly bolstered by the ability to predict tumour aggressiveness and lethality. In addition, one has to be mindful of the fact that tumours are heterogeneous, and the biology can be very complex. For translation of some of these synthetic biomarker technologies into humans, one has to select the tumour system carefully to avoid a potentially harmful combination of unknown biology and an agent for which information on pharmacokinetics and long-term safety in a particular clinical setting is limited. The US National Cancer Institute (NCI) initiated the Human Tumor Atlas Network (HTAN) to create detailed molecular, cellular and spatial maps of a variety of precancers, in situ cancers and advanced cancers as a function of time[156]. This will lead to a profound understanding of how precancers transition to malignancy for those cancer types studied by the HTAN, and how invasive cancers progress, metastasize and respond to or develop resistance to treatment. The knowledge generated from in situ and early lesions will generate testable hypotheses and biological information that could be leveraged to develop synthetic diagnostic tools, while also unearthing new candidate endogenous biomarkers.

Multidisciplinary teams of bioengineers, biologists and clinicians should work together strategically and in an integrated manner to find the answers to a number of questions that include, but are not limited to, the following. For which early-stage tumours or precancer lesions of lethal potential are the biology and pathogenesis sufficiently understood to drive the engineering of sensors? How can machine learning support identification of key features within complex biological datasets to achieve the required predictive power for synthetic biomarkers? Which populations will benefit the most from early detection? What are the short-term or long-term tolerability profiles of bioengineered sensors? How often can patients be screened? What additional complications can result from patient comorbidities? Under what situations can the same probes be used for detecting cancer recurrence? How expensive will long-term surveillance of 'at risk' patients be compared with the current standard of care? How can patient and tumour heterogeneity be overcome to ensure diagnostic accuracy? Which proof-of-principle studies in humans are worth consideration? How can mathematical models assist in such endeavours? Although there may seem to be more unknowns than definitive answers at this time, we anticipate that solutions will emerge at an increasingly rapid pace through collective, multidisciplinary efforts, and the audacious and innovative visions of scientists.

Acknowledgements

Competing interests

G.A.K. is a co-founder of Glympse Bio, and consults for Glympse Bio and Satellite Bio. S.N.B. is a director of Vertex, is a co-founder of and consultant for Glympse Bio, Satellite Bio and CEND Therapeutics, is a consultant for Moderna, and receives sponsored research funds from Johnson & Johnson. S.G., C.P., S.S. and L.G. declare no conflicts of interest.

References

1. Siegel RL, Miller KD & Jemal A Cancer statistics, 2020. CA Cancer J. Clin 70, 7–30 (2020). [PubMed: 31912902]

2. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N. Engl. J. Med 365, 395–409 (2011). [PubMed: 21714641]

3. Ru Zhao Y et al. NELSON lung cancer screening study. Cancer Imaging 11 Spec. No. A, S79–S84 (2011). [PubMed: 22185865]

4. Siu AL Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann. Intern. Med 164, 279–296 (2016). [PubMed: 26757170]

5. Bibbins-Domingo K et al. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. JAMA 315, 2564–2575 (2016). [PubMed: 27304597]

6. Curry SJ et al. Screening for cervical cancer: US Preventive Services Task Force recommendation statement. JAMA 320, 674–686 (2018). [PubMed: 30140884]

7. Moyer VA Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann. Intern. Med 160, 330–338 (2014). [PubMed: 24378917]

8. Pinsky PF, Prorok PC & Kramer BS Prostate cancer screening -a perspective on the current state of the evidence. N. Engl. J. Med 376, 1285–1289 (2017). [PubMed: 28355509]

9. Cohen JD et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 359, 926–930 (2018). [PubMed: 29348365]

10. Au SH et al. Clusters of circulating tumor cells: a biophysical and technological perspective. Curr. Opin. Biomed. Eng 3, 13–19 (2017). [PubMed: 29226271]

11. Bettegowda C et al. Detection of circulating tumor DNA in early-and late-stage human malignancies. Sci. Transl. Med 6, 224ra224 (2014).

12. Maheswaran S et al. Detection of mutations in EGFR in circulating lung-cancer cells. N. Engl. J. Med 359, 366–377 (2008). [PubMed: 18596266]

13. Dawson SJ et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. N. Engl. J. Med 368, 1199–1209 (2013). [PubMed: 23484797]

14. Lennon AM et al. Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. Science 369, eabb9601 (2020). [PubMed: 32345712]

15. De Rubis G, Rajeev Krishnan S & Bebawy M Liquid biopsies in cancer diagnosis, monitoring, and prognosis. Trends Pharmacol. Sci 40, 172–186 (2019). [PubMed: 30736982]

16. Sokoll LJ et al. A prospective, multicenter, National Cancer Institute Early Detection Research Network study of [−2]proPSA: improving prostate cancer detection and correlating with cancer aggressiveness. Cancer Epidemiol. Biomarkers Prev 19, 1193–1200 (2010). [PubMed: 20447916]

17. Karlsen MA et al. Evaluation of HE4, CA125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) as diagnostic tools of epithelial ovarian cancer in patients with a pelvic mass. Gynecol. Oncol 127, 379–383 (2012). [PubMed: 22835718]

18. Jaiswal S et al. Age-related clonal hematopoiesis associated with adverse outcomes. N. Engl. J. Med 371, 2488–2498 (2014). [PubMed: 25426837]

19. Prensner JR, Rubin MA, Wei JT & Chinnaiyan AM Beyond PSA: the next generation of prostate cancer biomarkers. Sci. Transl. Med 4, 127rv123 (2012).

20. Hammarström S The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. Semin. Cancer Biol 9, 67–81 (1999). [PubMed: 10202129]

21. Imperiale TF, Ransohoff DF & Itzkowitz SH Multitarget stool DNA testing for colorectal-cancer screening. N. Engl. J. Med 371, 187–188 (2014).

22. Serganova I & Blasberg RG Molecular imaging with reporter genes: has its promise been delivered? J. Nucl. Med 60, 1665–1681 (2019). [PubMed: 31792128]

23. Gilad AA & Shapiro MG Molecular imaging in synthetic biology, and synthetic biology in molecular imaging. Mol. Imaging Biol 19, 373–378 (2017). [PubMed: 28213833]

24. Condeelis J & Weissleder R In vivo imaging in cancer. Cold Spring Harb. Perspect. Biol 2, a003848 (2010). [PubMed: 20861158]

25. Kwong GA et al. Mass-encoded synthetic biomarkers for multiplexed urinary monitoring of disease. Nat. Biotechnol 31, 63–70 (2013). [PubMed: 23242163] This study first describes the design of protease-activated synthetic biomarkers for non-invasive detection of colorectal cancer from urine in a mouse model.

26. Abbosh C et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 545, 446–451 (2017). [PubMed: 28445469]

27. Heitzer E, Haque IS, Roberts CES & Speicher MR Current and future perspectives of liquid biopsies in genomics-driven oncology. Nat. Rev. Genet 20, 71–88 (2019). [PubMed: 30410101] This comprehensive review highlights the opportunities as well as the many challenges that must be overcome before liquid biopsies can be widely used for cancer detection.

28. Fleischhacker M & Schmidt B Circulating nucleic acids (CNAs) and cancer — a survey. Biochim. Biophys. Acta Rev. Cancer 1775, 181–232 (2007).

29. Diehl F et al. Circulating mutant DNA to assess tumor dynamics. Nat. Med 14, 985–990 (2008). [PubMed: 18670422]

30. Erdi YE Limits of tumor detectability in nuclear medicine and PET. Mol. Imaging Radionucl. Ther 21, 23–28 (2012). [PubMed: 23486256]

31. Lutz AM, Willmann JK, Cochran FV, Ray P & Gambhir SS Cancer screening: a mathematical model relating secreted blood biomarker levels to tumor sizes. PLoS Med 5, e170 (2008). [PubMed: 18715113]

32. Hori SS & Gambhir SS Mathematical model identifies blood biomarker-based early cancer detection strategies and limitations. Sci. Transl. Med 3, 109ra116 (2011).In this study, the authors develop a mathematical model to determine how early a clinical blood biomarker can be used to detect cancer.

33. Hori SS, Lutz AM, Paulmurugan R & Gambhir SS A model-based personalized cancer screening strategy for detecting early-stage tumors using blood-borne biomarkers. Cancer Res 77, 2570–2584 (2017). [PubMed: 28283654]

34. Machiraju GB, Mallick P & Frieboes HB Multicompartment modeling of protein shedding kinetics during vascularized tumor growth. Sci. Rep 10, 16709 (2020). [PubMed: 33028917]

35. Birkbak NJ & McGranahan N Cancer genome evolutionary trajectories in metastasis. Cancer Cell 37, 8–19 (2020). [PubMed: 31935374]

36. Lambert AW, Pattabiraman DR & Weinberg RA Emerging biological principles of metastasis. Cell 168, 670–691 (2017). [PubMed: 28187288]

37. Massague J & Obenauf AC Metastatic colonization by circulating tumour cells. Nature 529, 298–306 (2016). [PubMed: 26791720]

38. Turajlic S & Swanton C Metastasis as an evolutionary process. Science 352, 169–175 (2016). [PubMed: 27124450]

39. Labidi-Galy SI et al. High grade serous ovarian carcinomas originate in the fallopian tube. Nat. Commun 8, 1093 (2017). [PubMed: 29061967]

40. Conner JR et al. Outcome of unexpected adnexal neoplasia discovered during risk reduction salpingo-oophorectomy in women with germ-line BRCA1 or BRCA2 mutations. Gynecol. Oncol 132, 280–286 (2014). [PubMed: 24333842]

41. Lopez-Giacoman S & Madero M Biomarkers in chronic kidney disease, from kidney function to kidney damage. World J. Nephrol 4, 57–73 (2015). [PubMed: 25664247]

42. Ishizawa T et al. Real-time identification of liver cancers by using indocyanine green fluorescent imaging. Cancer 115, 2491–2504 (2009). [PubMed: 19326450]

43. Hanahan D & Weinberg RA Hallmarks of cancer: the next generation. Cell 144, 646–674 (2011). [PubMed: 21376230]

44. Kwon EJ, Dudani JS & Bhatia SN Ultrasensitive tumour-penetrating nanosensors of protease activity. Nat. Biomed. Eng 1, 0054 (2017). [PubMed: 28970963]

45. Kessenbrock K, Plaks V & Werb Z Matrix metalloproteinases: regulators of the tumor microenvironment. Cell 141, 52–67 (2010). [PubMed: 20371345]

46. Ribatti D, Nico B, Crivellato E, Roccaro AM & Vacca A The history of the angiogenic switch concept. Leukemia 21, 44–52 (2007). [PubMed: 16990761]

47. Dudani JS, Ibrahim M, Kirkpatrick J, Warren AD & Bhatia SN Classification of prostate cancer using a protease activity nanosensor library. Proc. Natl Acad. Sci. USA 115, 8954–8959 (2018). [PubMed: 30126988]

48. Kirkpatrick JD et al. Urinary detection of lung cancer in mice via noninvasive pulmonary protease profiling. Sci. Transl. Med 12, eaaw0262 (2020). [PubMed: 32238573] This study demonstrates the use of a 14-plex library of synthetic biomarkers for early detection of lung cancer in genetically engineered mouse models.

49. Mac QD et al. Non-invasive early detection of acute transplant rejection via nanosensors of granzyme B activity. Nat. Biomed. Eng 3, 281–291 (2019). [PubMed: 30952979]

50. Lin KY, Kwong GA, Warren AD, Wood DK & Bhatia SN Nanoparticles that sense thrombin activity as synthetic urinary biomarkers of thrombosis. ACS Nano 7, 9001–9009 (2013). [PubMed: 24015809]

51. Kwong GA et al. Mathematical framework for activity-based cancer biomarkers. Proc. Natl Acad. Sci. USA 112, 12627–12632 (2015). [PubMed: 26417077] This study develops a physiologically based pharmacokinetic model to predict the performance of protease-activated synthetic biomarkers for early cancer detection in humans.

52. Dudani JS, Warren AD & Bhatia SN Harnessing protease activity to improve cancer care. Annu. Rev. Canc Biol 2, 353–376 (2018).

53. Soo Choi H et al. Renal clearance of quantum dots. Nat. Biotechnol 25, 1165 (2007). [PubMed: 17891134]

54. Dudani JS, Jain PK, Kwong GA, Stevens KR & Bhatia SN Photoactivated spatiotemporally-responsive nanosensors of in vivo protease activity. ACS Nano 9, 11708–11717 (2015). [PubMed: 26565752]

55. Wittrup KD, Thurber GM, Schmidt MM & Rhoden JJ Practical theoretic guidance for the design of tumor-targeting agents. Methods Enzymol 503, 255–268 (2012). [PubMed: 22230572]

56. Mac QD et al. Activity-based urinary biomarkers of response and resistance to checkpoint blockade immunotherapy. bioRxiv 10.1101/2020.12.10.420265 (2021).

57. Hao L, Zhao RT, Ngambenjawong C, Fleming HE & Bhatia SN CRISPR-Cas-amplified urine biomarkers for multiplexed and portable cancer diagnostics. bioRxiv 10.1101/2020.06.17.157180 (2020).

58. Warren AD et al. Disease detection by ultrasensitive quantification of microdosed synthetic urinary biomarkers. J. Am. Chem. Soc 136, 13709–13714 (2014). [PubMed: 25198059]

59. Warren AD, Kwong GA, Wood DK, Lin KY & Bhatia SN Point-of-care diagnostics for noncommunicable diseases using synthetic urinary biomarkers and paper microfluidics. Proc. Natl Acad. Sci. USA 111, 3671–3676 (2014). [PubMed: 24567404]

60. Chan LW et al. Engineering synthetic breath biomarkers for respiratory disease. Nat. Nanotechnol 15, 792–800 (2020). [PubMed: 32690884]

61. Loynachan CN et al. Renal clearable catalytic gold nanoclusters for in vivo disease monitoring. Nat. Nanotechnol 14, 883–890 (2019). [PubMed: 31477801]

62. Nishihara T et al. Beta-galactosidase-responsive synthetic biomarker for targeted tumor detection. Chem. Commun 54, 11745–11748 (2018).

63. Nishihara T et al. Synthetic biomarker design by using analyte-responsive acetaminophen. Chembiochem 18, 910–913 (2017). [PubMed: 28236354]

64. Fernandez-Garcia J, Altea-Manzano P, Pranzini E & Fendt SM Stable isotopes for tracing mammalian-cell metabolism in vivo. Trends Biochem. Sci 45, 185–201 (2020). [PubMed: 31955965]

65. Perets TT et al. Optimization of ^{13}C-urea breath test threshold levels for the detection of Helicobacter pylori infection in a national referral laboratory. J. Clin. Lab. Anal 33, e22674 (2019). [PubMed: 30221401]

66. Gorowska-Kowolik K, Chobot A & Kwiecien J ^{13}C methacetin breath test for assessment of microsomal liver function: methodology and clinical application. Gastroenterol. Res. Pract 2017, 7397840 (2017). [PubMed: 28757868]

67. Hoteit MA et al. Deterioration in liver function after liver-directed therapy for hepatocellular carcinoma measured by cholate clearance. GastroHep 2, 232–239 (2020).

68. Hanna GB, Boshier PR, Markar SR & Romano A Accuracy and methodologic challenges of volatile organic compound-based exhaled breath tests for cancer diagnosis: a systematic review and meta-analysis. JAMA Oncol 5, e182815 (2019). [PubMed: 30128487]

69. Djago F, Lange J & Poinot P Induced volatolomics of pathologies. Nat. Rev. Chem 5, 183–196 (2021).

70. Lange J et al. Volatile organic compound based probe for induced volatolomics of cancers. Angew. Chem. Int. Ed 58, 17563–17566 (2019).The authors of this study report the use of a deuterated metabolite that is released as a VOC in exhaled breath for cancer diagnosis in mice.

71. Bhang HE, Gabrielson KL, Laterra J, Fisher PB & Pomper MG Tumor-specific imaging through progression elevated gene-3 promoter-driven gene expression. Nat. Med 17, 123–129 (2011). [PubMed: 21151140] In this report of a vector-based synthetic biomarker, the authors demonstrate the use of a tumour-specific promoter to drive the production of a reporter to image disseminated cancer in mouse models of melanoma and breast cancer.

72. Browne AW et al. Cancer screening by systemic administration of a gene delivery vector encoding tumor-selective secretable biomarker expression. PLoS ONE 6, e19530 (2011). [PubMed: 21589655]

73. Warram JM et al. Systemic delivery of a breast cancer-detecting adenovirus using targeted microbubbles. Cancer Gene Ther 19, 545–552 (2012). [PubMed: 22653385]

74. Warram JM, Borovjagin AV & Zinn KR A genetic strategy for combined screening and localized imaging of breast cancer. Mol. Imaging Biol 13, 452–461 (2011). [PubMed: 20658194]

75. D'Souza AL et al. A strategy for blood biomarker amplification and localization using ultrasound. Proc. Natl Acad. Sci. USA 106, 17152–17157 (2009). [PubMed: 19805109]

76. Aalipour A et al. Engineered immune cells as highly sensitive cancer diagnostics. Nat. Biotechnol 37, 531–539 (2019). [PubMed: 30886438] This study describes engineered macrophages as immune cell sensors that detect cancer after infiltration by releasing a synthetic biomarker in response to metabolic polarization.

77. Ronald JA, Chuang H-Y, Dragulescu-Andrasi A, Hori SS & Gambhir SS Detecting cancers through tumor-activatable minicircles that lead to a detectable blood biomarker. Proc. Natl Acad. Sci. USA 112, 3068–3073 (2015). [PubMed: 25713388]

78. Niu G & Chen X Molecular imaging with activatable reporter systems. Theranostics 2, 413–423 (2012). [PubMed: 22539937]

79. Montano-Samaniego M, Bravo-Estupinan DM, Mendez-Guerrero O, Alarcon-Hernandez E & Ibanez-Hernandez M Strategies for targeting gene therapy in cancer cells with tumor-specific promoters. Front. Oncol 10, 605380 (2020). [PubMed: 33381459]

80. Tannous BA & Teng J Secreted blood reporters: insights and applications. Biotechnol. Adv 29, 997–1003 (2011). [PubMed: 21920429]

81. Chao CN et al. Gene therapy for human lung adenocarcinoma using a suicide gene driven by a lung-specific promoter delivered by JC virus-like particles. PLoS ONE 11, e0157865 (2016). [PubMed: 27322500]

82. Wu C et al. Combinatorial control of suicide gene expression by tissue-specific promoter and microRNA regulation for cancer therapy. Mol. Ther 17, 2058–2066 (2009). [PubMed: 19809402]

83. Jafri MA, Ansari SA, Alqahtani MH & Shay JW Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. Genome Med 8, 69 (2016). [PubMed: 27323951]

84. Jiang H et al. Arginine deiminase expressed in vivo, driven by human telomerase reverse transcriptase promoter, displays high hepatoma targeting and oncolytic efficiency. Oncotarget 8, 37694–37704 (2017). [PubMed: 28455966]

85. Kyo S, Takakura M, Fujiwara T & Inoue M Understanding and exploiting hTERT promoter regulation for diagnosis and treatment of human cancers. Cancer Sci 99, 1528–1538 (2008). [PubMed: 18754863]

86. Li C et al. MR molecular imaging of tumors based on an optimal hTERT promoter tyrosinase expression system. Oncotarget 7, 42474–42484 (2016). [PubMed: 27283901]

87. Berger J, Hauber J, Hauber R, Geiger R & Cullen BR Secreted placental alkaline phosphatase: a powerful new quantitative indicator of gene expression in eukaryotic cells. Gene 66, 1–10 (1988). [PubMed: 3417148]

88. Bettan M, Darteil R & Scherman D Secreted human placental alkaline phosphatase as a reporter gene for in vivo gene transfer. Anal. Biochem 271, 187–189 (1999). [PubMed: 10419635]

89. Bao R, Selvakumaran M & Hamilton TC Use of a surrogate marker (human secreted alkaline phosphatase) to monitor in vivo tumor growth and anticancer drug efficacy in ovarian cancer xenografts. Gynecol. Oncol 78, 373–379 (2000). [PubMed: 10985897]

90. Nilsson EE et al. An in vivo mouse reporter gene (human secreted alkaline phosphatase) model to monitor ovarian tumor growth and response to therapeutics. Cancer Chemother. Pharmacol 49, 93–100 (2002). [PubMed: 11862422]

91. Richter JR, Mahoney M, Warram JM, Samuel S & Zinn KR A dual-reporter, diagnostic vector for prostate cancer detection and tumor imaging. Gene Ther 21, 897–902 (2014). [PubMed: 25056609]

92. Hiramatsu N et al. Alkaline phosphatase vs luciferase as secreted reporter molecules in vivo. Anal. Biochem 339, 249–256 (2005). [PubMed: 15797565]

93. Tannous BA Gaussia luciferase reporter assay for monitoring biological processes in culture and in vivo. Nat. Protoc 4, 582–591 (2009). [PubMed: 19373229]

94. Wurdinger T et al. A secreted luciferase for ex vivo monitoring of in vivo processes. Nat. Methods 5, 171–173 (2008). [PubMed: 18204457]

95. Alessandrini F, Ceresa D, Appolloni I, Marubbi D & Malatesta P Noninvasive monitoring of glioma growth in the mouse. J. Cancer 7, 1791–1797 (2016). [PubMed: 27698917]

96. Chung E et al. Secreted Gaussia luciferase as a biomarker for monitoring tumor progression and treatment response of systemic metastases. PLoS ONE 4, e8316 (2009). [PubMed: 20016816]

97. Tseng AW, Akerstrom V, Chen C, Breslin MB & Lan MS Detection of neuroendocrine tumors using promoter-specific secreted Gaussia luciferase. Int. J. Oncol 48, 173–180 (2016). [PubMed: 26530405]

98. Liu SH et al. BIRC5 is a target for molecular imaging and detection of human pancreatic cancer. Cancer Lett 457, 10–19 (2019). [PubMed: 31059751]

99. Fang Y, Wolfson B & Godbey WT Non-invasive detection of bladder cancer via expression-targeted gene delivery. J. Gene Med 19, 366–375 (2017). [PubMed: 29024250]

100. Almeida AM, Queiroz JA, Sousa F & Sousa A Minicircle DNA: the future for DNA-based vectors? Trends Biotechnol 38, 1047–1051 (2020). [PubMed: 32409109]

101. Oliveira PH & Mairhofer J Marker-free plasmids for biotechnological applications -implications and perspectives. Trends Biotechnol 31, 539–547 (2013). [PubMed: 23830144]

102. Huyn ST et al. A potent, imaging adenoviral vector driven by the cancer-selective mucin-1 promoter that targets breast cancer metastasis. Clin. Cancer Res 15, 3126–3134 (2009). [PubMed: 19366829]

103. Wang T, Chen Y & Ronald JA A novel approach for assessment of prostate cancer aggressiveness using survivin-driven tumour-activatable minicircles. Gene Ther 26, 177–186 (2019). [PubMed: 30867586]

104. Reagan MR & Kaplan DL Concise review: mesenchymal stem cell tumor-homing: detection methods in disease model systems. Stem Cell 29, 920–927 (2011).

105. Liu L et al. Exogenous marker-engineered mesenchymal stem cells detect cancer and metastases in a simple blood assay. Stem Cell Res. Ther 6, 181 (2015). [PubMed: 26391980]

106. Karp JM & Leng Teo GS Mesenchymal stem cell homing: the devil is in the details. Cell Stem Cell 4, 206–216 (2009). [PubMed: 19265660]

107. Droujinine IA, Eckert MA & Zhao W To grab the stroma by the horns: from biology to cancer therapy with mesenchymal stem cells. Oncotarget 4, 651–664 (2013). [PubMed: 23744479]

108. Kouidhi S, Noman MZ, Kieda C, Elgaaied AB & Chouaib S Intrinsic and tumor microenvironment-induced metabolism adaptations of T cells and impact on their differentiation and function. Front. Immunol 7, 114 (2016). [PubMed: 27066006]

109. Somasundaram R et al. Tumor-associated B-cells induce tumor heterogeneity and therapy resistance. Nat. Commun 8, 607 (2017). [PubMed: 28928360]

110. Vitale M, Cantoni C, Pietra G, Mingari MC & Moretta L Effect of tumor cells and tumor microenvironment on NK-cell function. Eur. J. Immunol 44, 1582–1592 (2014). [PubMed: 24777896]

111. Smith TT et al. In situ programming of leukaemia-specific T cells using synthetic DNA nanocarriers. Nat. Nanotechnol 12, 813–820 (2017). [PubMed: 28416815]

112. Perez C, Gruber I & Arber C Off-the-shelf allogeneic T cell therapies for cancer: opportunities and challenges using naturally occurring "universal" donor T cells. Front. Immunol 11, 583716 (2020). [PubMed: 33262761]

113. Forbes NS Engineering the perfect (bacterial) cancer therapy. Nat. Rev. Cancer 10, 785–794 (2010). [PubMed: 20944664]

114. Zheng JH et al. Two-step enhanced cancer immunotherapy with engineered *Salmonella* Typhimurium secreting heterologous flagellin. Sci. Transl. Med 9, eaak9537 (2017). [PubMed: 28179508]

115. Jiang SN et al. Inhibition of tumor growth and metastasis by a combination of Escherichia coli-mediated cytolytic therapy and radiotherapy. Mol. Ther 18, 635–642 (2010). [PubMed: 20051939]

116. Zhou S, Gravekamp C, Bermudes D & Liu K Tumour-targeting bacteria engineered to fight cancer. Nat. Rev. Cancer 18, 727–743 (2018). [PubMed: 30405213]

117. Panteli JT, Van Dessel N & Forbes NS Detection of tumors with fluoromarker-releasing bacteria. Int. J. Cancer 146, 137–149 (2020). [PubMed: 31093970]

118. Panteli JT, Forkus BA, Van Dessel N & Forbes NS Genetically modified bacteria as a tool to detect microscopic solid tumor masses with triggered release of a recombinant biomarker. Integr. Biol 7, 423–434 (2015).

119. Danino T et al. Programmable probiotics for detection of cancer in urine. Sci. Transl. Med 7, 289ra284–289ra284 (2015).The authors of this study engineer the probiotic *E. coli* Nissle to colonize and report on the presence of liver tumours in mice by producing a colorimetric readout in urine.

120. Slomovic S, Pardee K & Collins JJ Synthetic biology devices for in vitro and in vivo diagnostics. Proc. Natl Acad. Sci. USA 112, 14429–14435 (2015). [PubMed: 26598662]

121. Danino T, Mondragon-Palomino O, Tsimring L & Hasty J A synchronized quorum of genetic clocks. Nature 463, 326–330 (2010). [PubMed: 20090747]

122. Prindle A et al. A sensing array of radically coupled genetic 'biopixels'. Nature 481, 39–44 (2011). [PubMed: 22178928]

123. Din MO et al. Synchronized cycles of bacterial lysis for in vivo delivery. Nature 536, 81–85 (2016). [PubMed: 27437587]

124. Chowdhury S et al. Programmable bacteria induce durable tumor regression and systemic antitumor immunity. Nat. Med 25, 1057–1063 (2019). [PubMed: 31270504]

125. Gurbatri CR et al. Engineered probiotics for local tumor delivery of checkpoint blockade nanobodies. Sci. Transl. Med 12, eaax0876 (2020). [PubMed: 32051224]

126. Ho CL et al. Engineered commensal microbes for diet-mediated colorectal-cancer chemoprevention. Nat. Biomed. Eng 2, 27–37 (2018). [PubMed: 31015663]

127. Brown PO & Palmer C The preclinical natural history of serous ovarian cancer: defining the target for early detection. PLoS Med 6, e1000114 (2009). [PubMed: 19636370]

128. Rakhit CP et al. Early detection of pre-malignant lesions in a *KRAS*^{G12D}-driven mouse lung cancer model by monitoring circulating free DNA. Dis. Model Mech 12, dmm036863 (2019). [PubMed: 30760495]

129. Whitney M et al. Parallel in vivo and in vitro selection using phage display identifies protease-dependent tumor-targeting peptides. J. Biol. Chem 285, 22532–22541 (2010). [PubMed: 20460372]

130. Ruoslahti E Tumor penetrating peptides for improved drug delivery. Adv. Drug Deliv. Rev 110–111, 3–12 (2017).

131. Jones S et al. Comparative lesion sequencing provides insights into tumor evolution. Proc. Natl Acad. Sci. USA 105, 4283–4288 (2008). [PubMed: 18337506]

132. Yachida S et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 467, 1114–1117 (2010). [PubMed: 20981102]

133. Meza R, Jeon J, Moolgavkar SH & Luebeck EG Age-specific incidence of cancer: phases, transitions, and biological implications. Proc. Natl Acad. Sci. USA 105, 16284–16289 (2008). [PubMed: 18936480]

134. Luebeck EG Cancer: genomic evolution of metastasis. Nature 467, 1053–1055 (2010). [PubMed: 20981088]

135. Ittmann M et al. Animal models of human prostate cancer: the consensus report of the new york meeting of the mouse models of human cancers consortium prostate pathology committee. Cancer Res 73, 2718–2736 (2013). [PubMed: 23610450]

136. Puente XS, Sanchez LM, Overall CM & Lopez-Otin C Human and mouse proteases: a comparative genomic approach. Nat. Rev. Genet 4, 544–558 (2003). [PubMed: 12838346]

137. Choi B, Rempala GA & Kim JK Beyond the Michaelis–Menten equation: accurate and efficient estimation of enzyme kinetic parameters. Sci. Rep 7, 17018 (2017). [PubMed: 29208922]

138. Keu KV et al. Reporter gene imaging of targeted T cell immunotherapy in recurrent glioma. Sci. Transl. Med 9, eaag2196 (2017). [PubMed: 28100832]

139. Widen JC et al. AND-gate contrast agents for enhanced fluorescence-guided surgery. Nat. Biomed. Eng 5, 264–277 (2021). [PubMed: 32989286] The authors of this article show that an AND gate optical imaging probe that requires two distinct protease cleavage events significantly increased specificity and sensitivity in the detection of tumour tissue.

140. Ronald JA, D'Souza AL, Chuang HY & Gambhir SS Artificial microRNAs as novel secreted reporters for cell monitoring in living subjects. PLoS ONE 11, e0159369 (2016). [PubMed: 27442530]

141. Roybal KT et al. Precision tumor recognition by T cells with combinatorial antigen-sensing circuits. Cell 164, 770–779 (2016). [PubMed: 26830879]

142. Morsut L et al. Engineering customized cell sensing and response behaviors using synthetic notch receptors. Cell 164, 780–791 (2016). [PubMed: 26830878]

143. Holt BA & Kwong GA Protease circuits for processing biological information. Nat. Commun 11, 5021 (2020). [PubMed: 33024117]

144. Cazanave S et al. SAT-281-protease activity sensors for non-invasive monitoring of NASH. J. Hepatol 70, e760 (2019).

145. Azeem R et al. Safety and tolerability in healthy volunteers of the Glympse bio test system-NASH diagnostic. Hepatology 72, 941A–942A (2020).

146. Whitley MJ et al. A mouse-human phase 1 co-clinical trial of a protease-activated fluorescent probe for imaging cancer. Sci. Transl. Med 8, 320ra324 (2016).

147. Unkart JT et al. Intraoperative tumor detection using a ratiometric activatable fluorescent peptide: a first-in-human phase 1 study. Ann. Surg. Oncol 24, 3167–3173 (2017). [PubMed: 28699134]

148. Smith BL et al. Feasibility study of a novel protease-activated fluorescent imaging system for real-time, intraoperative detection of residual breast cancer in breast conserving surgery. Ann. Surg. Oncol 27, 1854–1861 (2020). [PubMed: 31898104]

149. Desnoyers LR et al. Tumor-specific activation of an EGFR-targeting probody enhances therapeutic index. Sci. Transl. Med 5, 207ra144 (2013).

150. US National Library of Medicine. ClinicalTrials.gov, https://ClinicalTrials.gov/show/NCT03993379 (2019).

151. US National Library of Medicine. ClinicalTrials.gov, https://ClinicalTrials.gov/show/NCT03013491 (2017).

152. Austin RJ et al. TriTACs, a novel class of T-cell-engaging protein constructs designed for the treatment of solid tumors. Mol. Cancer Ther 20, 109–120 (2021). [PubMed: 33203731]

153. US National Library of Medicine. ClinicalTrials.gov, https://ClinicalTrials.gov/show/NCT03577028 (2018).

154. Horwitz S et al. Brentuximab vedotin with chemotherapy for CD30-positive peripheral T-cell lymphoma (ECHELON-2): a global, double-blind, randomised, phase 3 trial. Lancet 393, 229–240 (2019). [PubMed: 30522922]

155. Duong MT, Qin Y, You SH & Min JJ Bacteria–cancer interactions: bacteria-based cancer therapy. Exp. Mol. Med 51, 1–15 (2019).

156. Rozenblatt-Rosen O et al. The human tumor atlas network: charting tumor transitions across space and time at single-cell resolution. Cell 181, 236–249 (2020). [PubMed: 32302568]

157. Rajkomar A, Dean J & Kohane I Machine learning in medicine. N. Engl. J. Med 380, 1347–1358 (2019). [PubMed: 30943338] This article provides a conceptual overview of the use of machine learning and its applications in medicine.

158. Sidey-Gibbons JAM & Sidey-Gibbons CJ Machine learning in medicine: a practical introduction. BMC Med. Res. Methodol 19, 64 (2019). [PubMed: 30890124]

159. Madabhushi A & Lee G Image analysis and machine learning in digital pathology: challenges and opportunities. Med. Image Anal 33, 170–175 (2016). [PubMed: 27423409]

160. Cui M & Zhang DY Artificial intelligence and computational pathology. Lab. Invest 101, 412–422 (2021). [PubMed: 33454724]

161. Golub TR et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286, 531 (1999). [PubMed: 10521349]

162. Perakakis N, Yazdani A, Karniadakis GE & Mantzoros C Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics. Metabolism 87, A1–A9 (2018). [PubMed: 30098323]

163. Grapov D, Fahrmann J, Wanichthanarak K & Khoomrung S Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. OMICS 22, 630–636 (2018). [PubMed: 30124358]

164. Douglas GM et al. Multi-omics differentially classify disease state and treatment outcome in pediatric Crohn's disease. Microbiome 6, 13 (2018). [PubMed: 29335008]

165. Tasaki S et al. Multi-omics monitoring of drug response in rheumatoid arthritis in pursuit of molecular remission. Nat. Commun 9, 2755 (2018). [PubMed: 30013029]

166. Pavel AB, Sonkin D & Reddy A Integrative modeling of multi-omics data to identify cancer drivers and infer patient-specific gene activity. BMC Syst. Biol 10, 16–16 (2016). [PubMed: 26864072]

167. Breiman L Random forests. Mach. Learn 45, 5–32 (2001).

168. van der Maaten L & Hinton G Visualizing data using t-SNE. J. Mach. Learn. Res 9, 2579–2605 (2008).

Box 1 |

The application of machine learning to cancer

- Machine learning is a branch of artificial intelligence based on the theory that computers can learn from prior examples to perform tasks and predict outcomes rather than being explicitly programmed with rules to make decisions[157,158]. A key advantage of machine learning compared with human learning is that computers can learn from complex and massive amounts of data. For example, machine learning is being applied to wide-ranging areas in medicine from pathology for automated detection of cancer in digitized histology slides[159,160] to prediction of disease aggressiveness and patient outcomes from -omic datasets[161–166].

- Supervised learning is a type of machine learning algorithm whereby the model learns from prior examples by training on a range of input features (for example, biomarker levels, height and weight) associated with a known output label (for example, cancer)[157,158]. The trained model can then generalize the input-to-output mapping to predict the assignment of never-before-seen inputs to an output label. These predictions can result in discrete categories (for example, benign or malignant) or a continuous range (for example, a score from 0 to 100).

- A classifier is a supervised learning method that categorizes unlabelled data into one or more discrete categories, also referred to as 'classes', such as cancer stages. For example, a random forest classifier is a collection of a large number of randomly created decision trees in which each node in the decision tree works on a random subset of features to calculate the output. The predicted output class is based on the most popular prediction among individual decision trees[167].

- In contrast to supervised learning, unsupervised learning is a type of machine learning algorithm that draws inferences from unlabelled data without prior knowledge. Clustering tumour specimens based on RNA transcript levels by t-distributed stochastic neighbour embedding (t-SNE), for example, is a form of unsupervised learning since the data are categorized without the use of predefined labels[168].
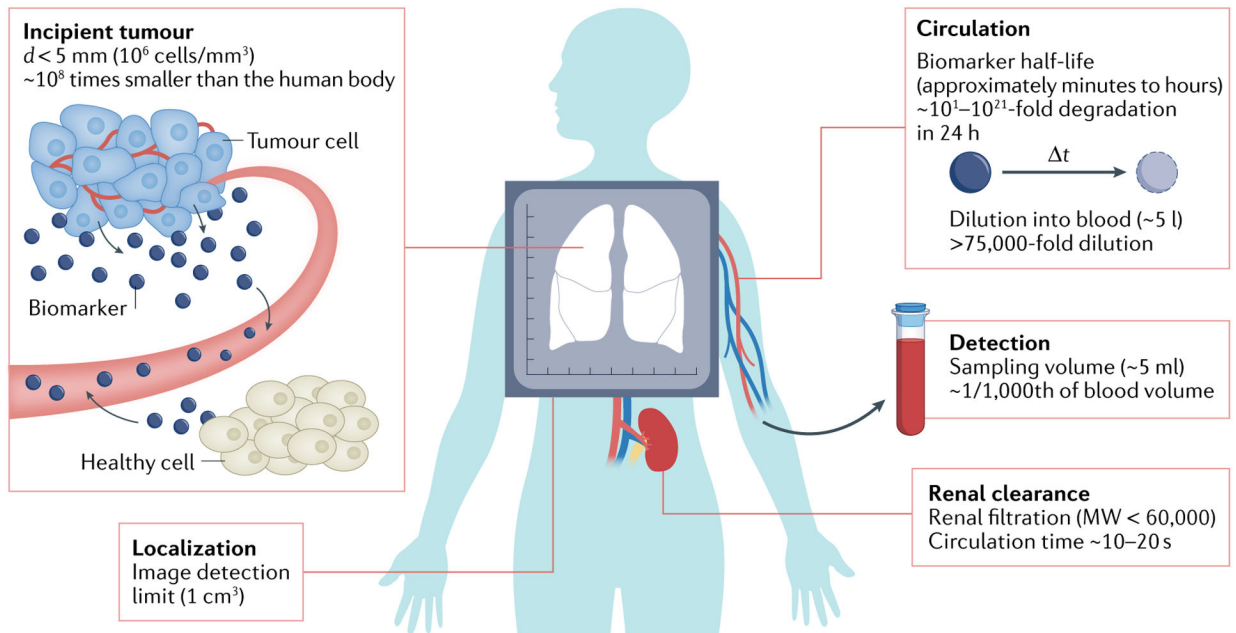
Fig. 1 |. Challenges associated with detecting early-stage tumours.
An early-stage tumour (smaller than 5 mm in diameter) is on average eight orders of magnitude smaller in volume than the human body. Several factors hinder the ability to detect biomarkers shed from tumours, including transport challenges from the tumour microenvironment (TME) into the circulation, an approximately five orders of magnitude-fold dilution into blood and short circulation times owing to degradation and renal filtration. These factors decrease the number of tumour-associated biomarkers (for example, cell-free nucleic acids, proteins, metabolites and circulating tumour cells) that can be found in a typical 5–10 ml blood draw, which represents only ~1/1,000th of the total circulation volume. d, diameter; MW, molecular weight; t, change in time.
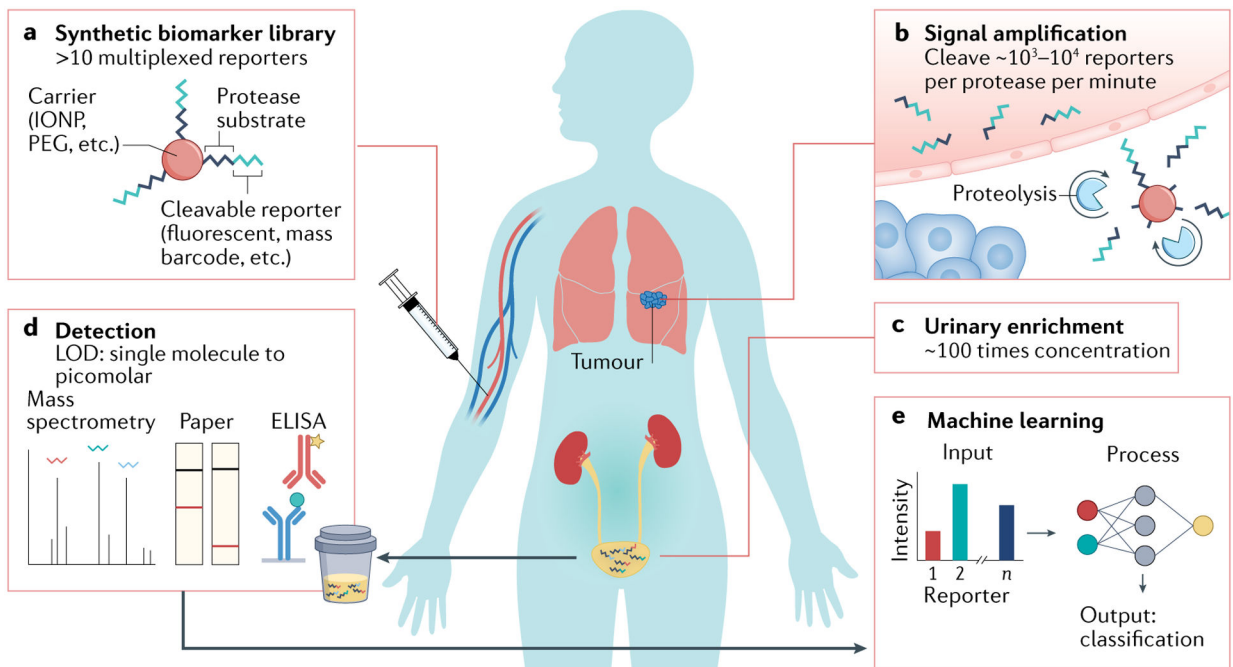
Fig. 2 |. Activity-based synthetic biomarkers enrich tumour protease signatures.
a | Synthetic biomarkers are activity-based sensors that comprise a biocompatible carrier (for example, iron oxide nanoparticles (IONPs), polyethylene glycol (PEG) or iron oxide nanoworms) coupled to peptide substrates for dysregulated proteases and a cleavable reporter (for example, mass-barcoded or fluorescent peptides). Peptide substrate libraries can be multiplexed by using orthogonal reporters. **b** | Following non-invasive delivery of a synthetic biomarker library (for example, by intravenous or intranasal administration), protease signatures are amplified by enzymatic turnover, resulting in the release of multiple reporters from each sensor upon proteolytic cleavage at the tumour site. **c,d** | The cleaved reporters are shed into the circulation, where they are further enriched by renal filtration (panel **c**) and detected in urine samples by several analytical platforms, including mass spectrometry, enzyme-linked immunosorbent assay (ELISA) and paper tests (panel **d**). **e** | Diagnosis is performed using machine learning-based classification algorithms. LOD, limit of detection; *n*, number.
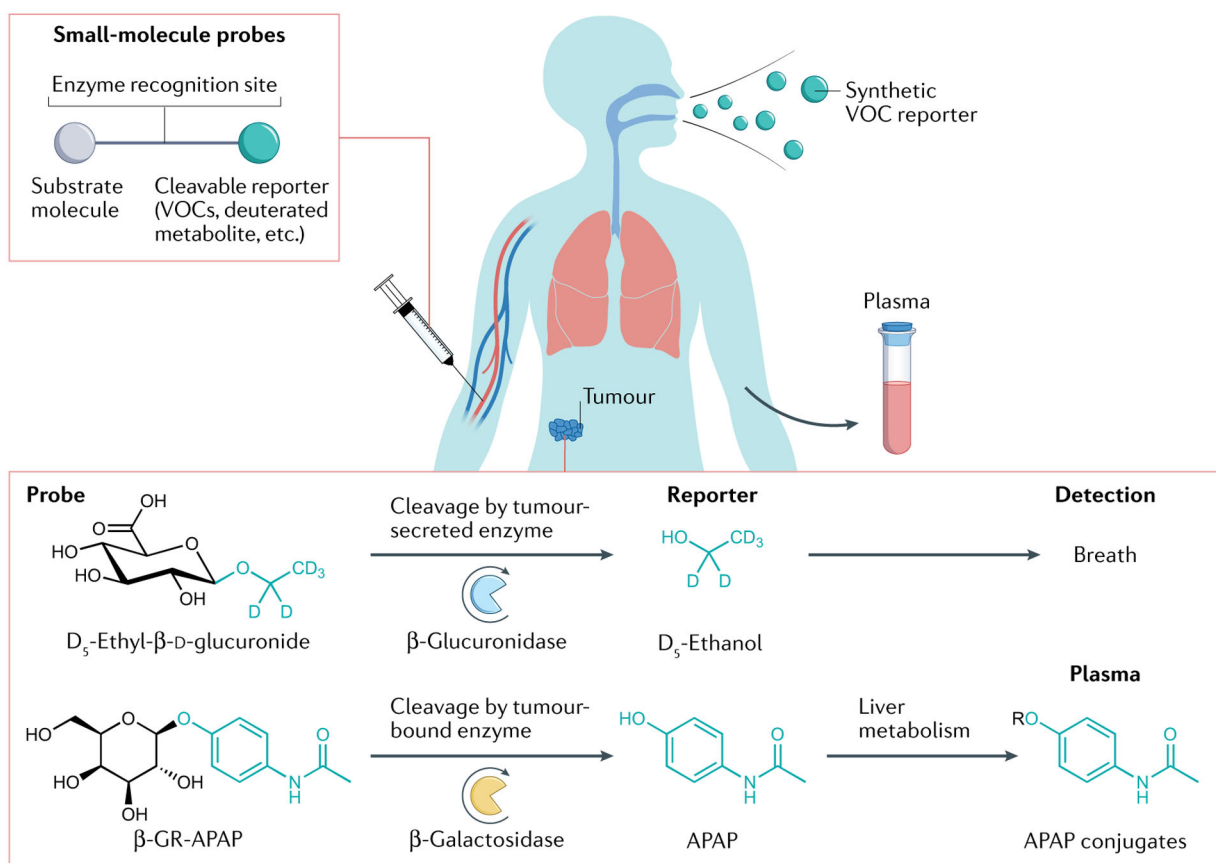
Fig. 3 |. Small-molecule probes sense tumour-associated enzymatic activity.
Small-molecule probes comprise an enzyme recognition site linked to a synthetic cleavable reporter, such as volatile organic compounds (VOCs) or stable isotope labels (for example, ^{13}C-methacetin or ^{13}C-cholate). Following systemic administration, tumour-associated enzymes convert the probes into synthetic biomarkers (for example, D_5-ethanol or acetaminophen (APAP) conjugates) whose abundance is detectable in breath or plasma samples. β-GR-APAP, β-galactosidase-responsive acetaminophen.
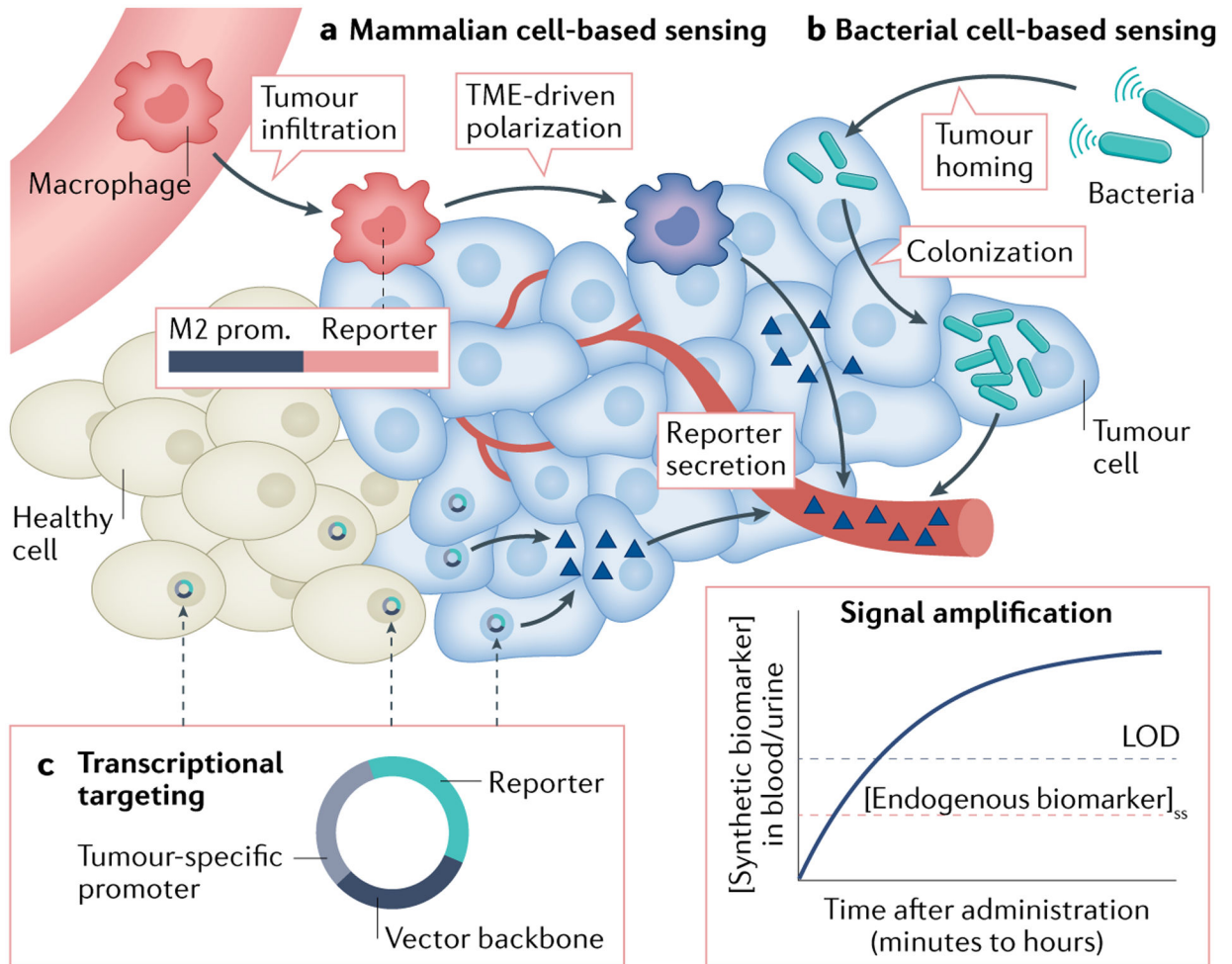
Fig. 4 |. Genetically encoded synthetic biomarkers leverage tumour-specific cues to achieve detectable signals.
Cells engineered with genetically encoded synthetic biomarkers exploit key features of the tumour microenvironment (TME) to trigger the secretion of detectable reporters. Secreted reporters can be detected in blood to indicate the presence of disease or they can be imaged to provide spatial information on tumour location or immune cell activation. **a** | Mammalian cell-based 'immunodiagnostics' exploit the metabolic alterations of tumour-infiltrating macrophages to trigger the production of a secreted biomarker by engineered macrophages. **b** | Bacteria, which colonize tumours owing to suppressed immunosurveillance and increased availability of nutrients in the necrotic tumour core, release programmed reporters at the site of the tumour. **c** | DNA vectors leverage tumour-associated gene expression patterns by encoding a secretable reporter transcriptionally targeted to cancer cells using tumour-specific promoters. LOD, limit of detection; prom., promoter; ss, steady state.
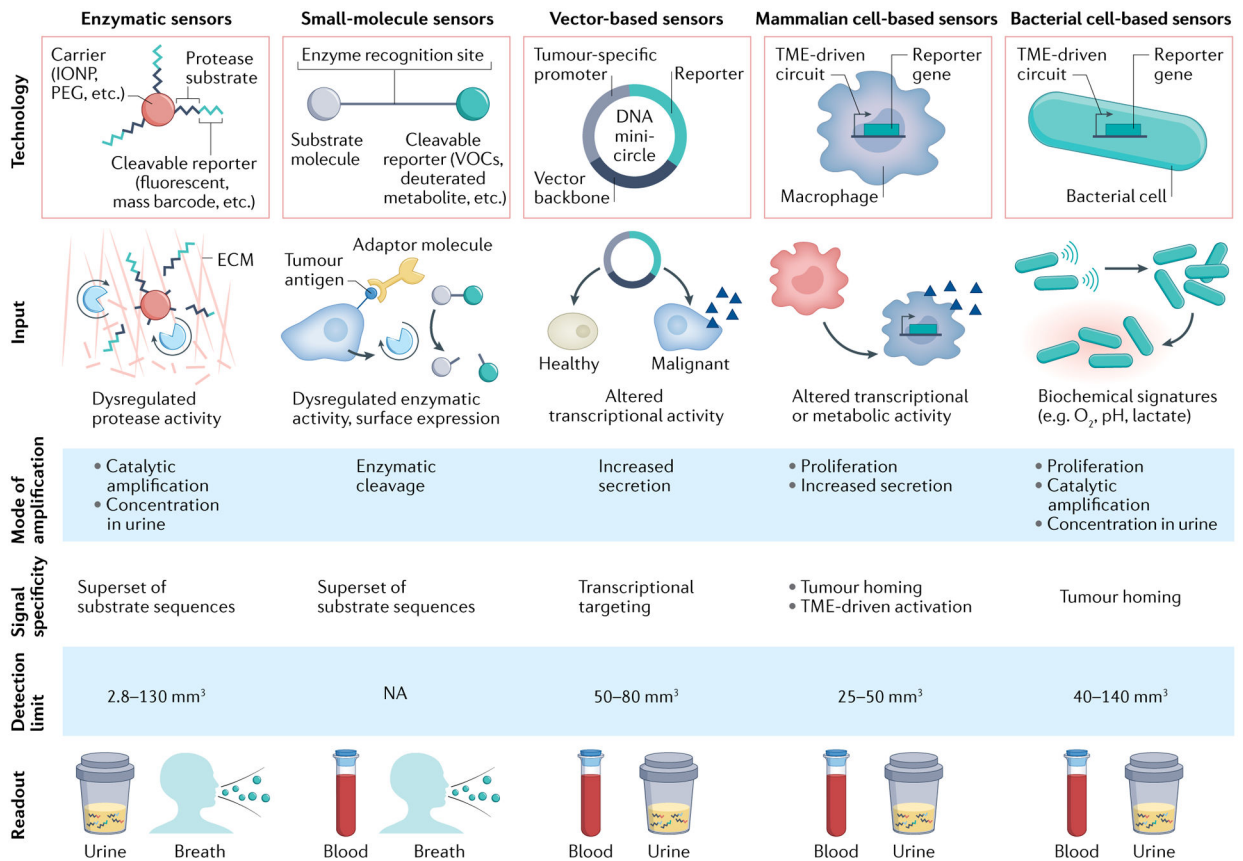
Fig. 5 |. Characteristics of synthetic biomarkers for early-stage cancer detection.
Enzymatic, small-molecule, DNA-based, mammalian cell-based and bacterial cell-based sensors leverage synthetic biomarkers to enhance early cancer detection. Each technology senses dysregulated activity (that is, the 'input') associated with the tumour microenvironment (TME), such as protease activity, metabolic activity or biophysical features. Through diverse modes of amplification and strategies for improving signal specificity, these approaches lower the limit of detection below current clinical thresholds ($\sim$1 cm^3). ECM, extracellular matrix; IONP, iron oxide nanoparticle; PEG, polyethylene glycol; VOCs, volatile organic compounds; NA, not available.