Review

Cancer Research

Androgen Receptor on the Move: Boarding the Microtubule Expressway to the Nucleus

Maria Thadani-Mulero[1], David M. Nanus[1,2], and Paraskevi Giannakakou[1,2]

Abstract

Recent studies have shown that the microtubule-stabilizing drug paclitaxel, which is commonly used for the treatment of prostate cancer, inhibits signaling from the androgen receptor by inhibiting its nuclear accumulation downstream of microtubule stabilization. This mechanism is independent of paclitaxel-induced mitotic arrest and could provide an alternative mechanism of drug action that can explain its clinical activity. In this review, we highlight the importance of signaling and trafficking pathways that depend on intact and dynamic microtubules, and, as such, they represent downstream targets of microtubule inhibitors. We showcase prostate cancer, which is driven by the activity of the androgen receptor, as recent reports have revealed a connection between the microtubule-dependent trafficking of the androgen receptor and the clinical efficacy of taxanes. Identification and further elucidation of microtubule-dependent tumor-specific pathways will help us better understand the molecular basis of clinical taxane resistance as well as to identify individual patients more likely to respond to treatment. *Cancer Res; 72(18); 4611–5. ©2012 AACR.*

Introduction

In 2012, prostate cancer will be diagnosed in more than 240,000 men, with approximately 28,000 deaths attributable to prostate cancer (1). Prostate cancer is a heterogeneous disease, driven primarily by androgen receptor (AR) signaling, and has been traditionally treated with androgen deprivation therapy (ADT). Although our understanding of the molecular basis of prostate cancer has significantly increased over the past decade, for men who develop metastases, the principle of ADT is essentially the same as it was first proposed 60 years ago: to interfere with androgen signaling (2). The goal of ADT is to block active AR signaling, either by eliminating the ligand or affecting the receptor directly. Although most patients with prostate cancer are initially sensitive to androgen withdrawal, loss of sensitivity to ADT occurs, leading to the development of castrate-resistant prostate cancer (CRPC; ref. 3). Similarly, even with the introduction of new and more effective therapies that target the androgen axis, such as the CYP17A1 inhibitor abiraterone, which targets the central synthesis of testosterone, and the AR antagonist MDV3100, ADT is not curative (3). The molecular disturbances that contribute to prostate cancer progression in the setting of castrate levels of circulating androgen have been reviewed elsewhere (4–7), but almost universally allow for the continued function of the AR as a transcription factor, resulting in androgen-driven prostate cancer growth. Consequently, targeting the androgen axis has remained a key concept in the development of novel therapeutic strategies.

In 2004, the combination of docetaxel plus prednisone was established as the standard of care for first-line treatment of patients with CRPC, making taxanes the first class of chemotherapy drugs shown to improve survival in CRPC (8, 9). At the cellular level, the taxanes (paclitaxel, docetaxel, and cabazitaxel) bind β-tubulin and stabilize microtubules, resulting in mitotic arrest and cell death (10, 11). Microtubules are dynamic cytoskeletal polymers critically important for several cellular functions, including structural support and the formation of the mitotic apparatus. During cell division, microtubule dynamics increase significantly (4–100 fold) to enable fast "search and chromosome capture" functions required for mitosis (10). Therefore, drugs that stabilize microtubules, like the taxanes, interfere with mitotic cell progression by suppressing microtubule dynamics. This key observation, supported by numerous *in vitro* studies, has led to the common belief that the clinical activity of taxanes stems from their antimitotic activity (12). However, this mechanism of action has not helped us understand the molecular basis of clinical response and resistance to taxane chemotherapy, as this model applies primarily to rapidly dividing cells and tissues. It is important to emphasize here that patients' tumors have significantly lower rates of cell division than cancer cells growing *in vitro*. For example, prostate cancer doubles every 33 to 577 days (13, 14), in contrast to the rapidly dividing prostate cancer cells grown in tissue culture with doubling times between 30 to 48 hours (15, 16); therefore, mitotic arrest alone cannot account for the therapeutic benefit of taxane-based chemotherapy (17, 18). Thus, the effects of taxanes on interphase microtubules and the cellular pathways that depend directly on intact microtubules

Authors' Affiliations: [1]Department of Medicine, Division of Hematology and Medical Oncology, Weill Cornell Medical College, and [2]Weill Cornell Cancer Center, New York, New York

Corresponding Author: Paraskevi Giannakakou, Weill Cornell Medical College, 1300 York Avenue, C610C, New York, NY 10065–4896. Phone: 212-746-3783; Fax: 212-746-6731; E-mail: pag2015@med.cornell.edu

doi: 10.1158/0008-5472.CAN-12-0783

could provide an alternate mechanism of action for this class of drugs. Although this notion challenges the existing paradigm of taxanes exerting their clinical activity exclusively through inhibition of mitosis, by shifting the focus to interphase microtubules, it provides a unique opportunity to dissect how these drugs work and why they are not effective in all tumor types and all patients. The question is then raised of what are the functions of interphase microtubules that are critical for the growth and survival of the tumor?

Interphase Microtubules as Targets for Taxane Chemotherapy

In interphase cells, microtubules cover the entire area of the cell's cytoplasm, originating from the microtubule-organizing center (MTOC), right outside the nucleus, and extending all the way to the plasma membrane, providing ample surface for protein–protein interactions. In epithelial cells, microtubules display an inherent polarity, having their slow-growing minus end embedded in the MTOC and their fast-growing plus end oriented toward the plasma membrane. This polarity is used by microtubule-based motor proteins, moving cargoes either toward the nucleus (dynein) or toward the plasma membrane (kinesins), thereby allowing for the directional flow of signal information within the cell, which ultimately dictates cell function (19–21). All of these qualities make microtubules centralized nodes of dynamic signaling pathways (22), which remain largely unexplored and their therapeutic potential unexploited.

We have focused our efforts into identifying pathways that depend on intact and dynamic microtubules, disruption of which by taxane treatment would be fatal for tumor cell survival. We have shown that the activity of certain transcription factors depends on the chemomechanics of the microtubule cytoskeleton and, therefore, these factors represent downstream targets of taxane activity. These factors include the tumor suppressor p53, which requires intact microtubules and the dynein minus-end–directed motor protein for trafficking and effective nuclear accumulation (23, 24), and the hypoxia-inducible factor-1α (HIF-1α), of which the activity is tightly regulated by microtubule dynamics through microtubule-dependent mRNA trafficking to sites of active protein translation (25, 26). Interestingly, this mechanism does not apply to renal cell carcinoma (RCC), in which HIF-1α regulation is independent of microtubules (27). These results can potentially explain the lack of taxane clinical activity in RCC; identification of the cellular factors that link HIF-1α to microtubules that are missing or are deregulated in RCC can provide a new therapeutic strategy for the treatment of this lethal disease.

Additional proteins of which the translocation is microtubule mediated are the retinoblastoma protein (28), the glucocorticoid receptor (29), and the parathyroid hormone receptor protein (PTHrP; ref. 30). Komlodi-Pasztor and colleagues provide a detailed list of other proteins that traffic on or associate with microtubules (14).

In prostate cancer, specifically, we and others have recently shown that taxane chemotherapy impairs AR signaling activity, not through mitosis, but by impairing AR nuclear translocation and inhibiting subsequent transcriptional activation of androgen response element (ARE)–containing target genes (31, 32). Other recent studies showed that paclitaxel-induced inhibition of AR activity is mediated by FOXO1, an AR-suppressive nuclear transcription factor (33), and that docetaxel treatment can downregulate the expression of AR and prostate-specific antigen in prostate cancer cell lines (34).

Prompted by the established clinical activity of taxanes in CRPC, together with the fact that AR continues to drive disease progression despite prior antiandrogen therapies (35), we set out to investigate the role of tubulin and the impact of microtubule-targeting drugs on AR trafficking and signaling.

The results presented in our recent study (31) provide a mechanistic insight for the clinical activity of taxanes in CRPC by revealing an unconventional link between a nuclear transcription factor and the chemomechanics of the microtubule cytoskeleton. As illustrated in Fig. 1, AR associates with microtubules and is trafficked toward the nucleus with the aid of the minus-end–directed motor protein dynein. It is well established that upon ligand binding, AR dimerizes and the ligand–receptor complex translocates to the nucleus (36). However, the mechanism enabling this translocation was previously unknown. The recent studies (31, 32) identify microtubules as the "highway tracks" that enable the rapid and targeted nuclear "delivery" of AR, which is required for its transcriptional activity. What remains to be solved, however, is whether AR binds microtubules directly, remaining tethered in the cytoplasm and associating with dynein only after ligand binding, or whether AR associates with dynein in the cytoplasm and, following ligand binding, the ligand–receptor complex gets recruited to the microtubule for trafficking. Recent data from our laboratory support the first model, as we show that AR association with microtubules is diminished in the presence of ligand (Fig. 2) and that the coprecipitation of AR with dynein is enhanced following ligand stimulation (30). These results suggest that unliganded AR, at steady state, is tethered to the microtubule cytoskeleton and that, upon ligand binding, the complex associates with dynein and is released from the microtubule. *In vitro* studies using recombinant AR protein and purified microtubules are required to further investigate AR-binding affinity to microtubules. Regardless of direct or indirect binding of AR to the microtubule, dynein's function is critical for the trafficking not only of AR but also other proteins for which nuclear localization is critical to their respective physiologic roles, such as p53 (23), retinoblastoma protein (28), and PTHrP (30). With the implication of dynein in this mechanism, it has also become important to decipher the role of the dynein accessory proteins that mediate cargo recognition specific for AR. This information will enable the development of dynein small-molecule inhibitors, which in combination with taxane chemotherapy, could prove significantly beneficial to patients with CRPC.

Using the Microtubule–AR Axis to Understand Clinical Taxane Resistance in Prostate Cancer

The model presented in Fig. 1 provides a basis for understanding clinical taxane resistance in prostate cancer.

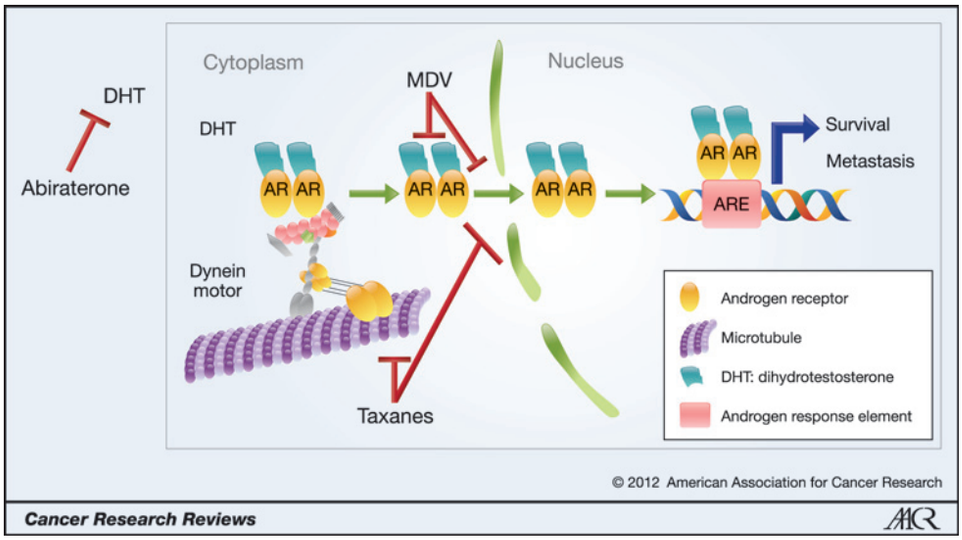


Figure 1. Proposed model of taxane mechanism of action in prostate cancer. This model represents a novel mechanism of action for taxanes in prostate cancer, which implicates this class of drugs in critical interphase cellular functions such as AR intracellular transport and signaling. In the model, AR associates with microtubules and translocates to the nucleus via the motor protein dynein. This transport is made possible because of the inherent polarity of microtubules, which is recognized by the minus-end–directed motor protein dynein, to transport cargoes toward the nucleus. Taxanes, which bind to and hyperstabilize microtubules, inhibit this trafficking and subsequently prevent AR from reaching the nucleus and activating target genes. This mechanism of action predicts that the combination of a taxane with an inhibitor of AR ligand synthesis (i.e., abiraterone) or with inhibitors of AR–ligand interaction (i.e., MDV3100) would be synergistic in the clinical setting, as there will be inhibition of the AR signaling axis by 2 different but converging pathways. Additionally, the model predicts that a small-molecule inhibitor of dynein would similarly impair AR nuclear accumulation and would also be synergistic in combination with a taxane. Finally, the model suggests that a putative small-molecule inhibitor targeting the interaction between AR and microtubules or dynein could be used therapeutically for CRPC treatment.

Darshan and colleagues (31) showed that perturbation of the microtubule–AR axis is an important determinant of taxane activity, independent of mitosis, as AR cytoplasmic sequestration in circulating tumor cells isolated from patients with CRPC significantly correlated with clinical response to taxane chemotherapy.

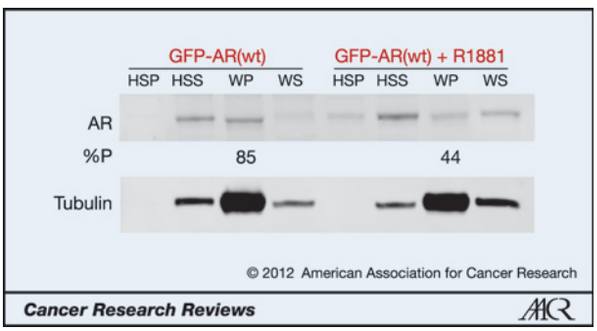


Figure 2. Ligand treatment decreases the association of AR with microtubules. A microtubule cosedimentation assay using whole-cell lysate from PC3:mCh-tub cells transfected with GFP–AR[(wild-type (wt)] was carried out in the presence or absence of the synthetic dihydrotestosterone analogue (R1881, 10 nmol/L). Briefly, a total of 1 mg of precleared cell extract [high-speed supernatant (HSS)] from the transfected cells was incubated for 30 min with 10 μmol/L exogenous purified bovine brain tubulin and subjected to a cycle of polymerization with 20 μmol/L paclitaxel at 37°C. The samples were centrifuged at 100,000*g* to separate the microtubule polymers [warm pellet (WP)] from the soluble tubulin dimers [warm supernatant (WS)], resolved by SDS-PAGE and immunoblotted for the presence of AR and tubulin. Note that R1881 treatment decreases the amount of AR protein that cosediments with the microtubule polymer, as can be seen by the shift from 85% to 44% of AR in the WP [%$P = 100^{*}$WP/(WP+WS)]. Protein quantification was done using ImageJ (NIH) software. Tubulin was detected as microtubule polymers in the WP fraction in both conditions. HSP, high-speed pellet.

Taxanes bind β-tubulin, suppress microtubule dynamics, and hyperstabilize the microtubule cytoskeleton by inducing the formation of microtubule bundles, a hallmark of effective drug-target engagement. Microtubule bundling is the first cellular insult that leads to disruption of downstream pathways. This mechanism implies that taxane chemotherapy should be most effective against tumor types in which microtubule-dependent pathways drive tumor progression, like the AR pathway in prostate cancer.

Despite the success of taxanes in CRPC treatment, their efficacy varies from patient to patient, whereas it remains unclear why individual patients respond to paclitaxel but not docetaxel and vice versa, even though these drugs share the same mechanism of action and a common binding site on β-tubulin. In 2010, the docetaxel analogue cabazitaxel was approved by the U.S. Food and Drug Administration for the treatment of patients with CRPC who previously failed docetaxel-based therapy (37). This approval highlights, once again, the activity of this class of drugs in CRPC, while raising the question of finding what the molecular basis of clinical taxane resistance is. According to the model presented here, clinical taxane resistance could arise as a result of the following: (i) impaired drug uptake, potentially due to the presence of P-glycoprotein or other drug transporters; (ii) impaired binding to β-tubulin, possibly due to the presence of tubulin mutations at the drug-binding site or overexpression of βIII tubulin isotype; (iii) the presence of AR mutations or splice variants that do not require microtubule-based transport; and (iv) dysregulation of dynein–cargo interaction.

For the first possibility, limited studies have not suggested a significant correlation between P-glycoprotein expression and response to taxane treatment in patients with prostate cancer (38). Similarly, tubulin alterations, such as β-tubulin mutations or altered isotype expression, have not been associated with response to taxane-based therapy in CRPC either (39). Conversely, alterations of AR have been extensively studied in CRPC, albeit not in the context of taxane resistance. Recent studies have shown the presence of alternatively spliced AR variants, such as ARv567 and AR-V7, which arise following castration (40–43). These variants lack the ligand-binding domain, are insensitive to ADT, and are constitutively active in the nucleus, which allows for continuous AR transcriptional activity. The ARv567 variant was shown to be present in 59% of patients with CRPC and to arise in response to ADT or to the newer AR-targeted therapies, such as abiraterone (44). The frequency of this molecular alteration in CRPC makes it imperative to determine whether these variants are under microtubule control, similar to wild-type AR, and whether they would respond to taxane treatment. Our model predicts that any AR variant lacking the microtubule- or dynein-binding domain would be insensitive to taxane treatment and, thereby, has the potential to serve as a predictive biomarker of clinical taxane activity. Finally, for the dysregulation of dynein–cargo interaction, our model predicts that any aberration and/or mutation in the dynein motor protein that impairs cargo (AR) recognition and/or transport should lead to taxane resistance.

Therapeutic Implications and Perspectives

The model presented herein suggests that simultaneous targeting of different pathways that inhibit AR signaling may result in greater or more durable antitumor effects. Specifically, combination of a taxane, which interferes with AR nuclear translocation, with an inhibitor of androgen synthesis (e.g., abiraterone) or an inhibitor of AR–ligand interaction, such as MDV3100, which also inhibits AR nuclear accumulation (by yet-to-be defined mechanism; ref. 45), could yield enhanced therapeutic efficacy. To this end, a phase I trial is currently testing the combination of docetaxel and abiraterone in patients with CRPC. More importantly, understanding the precise mechanisms by which prostate cancer cells circumvent AR signaling inhibition will allow development of novel, more targeted therapies. For instance, the development of a small-molecule inhibitor of dynein, or of the dynein–AR interaction, could be added to the space of CRPC therapy. Similarly, the recent identification of an *N*-terminal–targeted AR inhibitor, which has the potential to target both AR wild-type and variants as the *N*-terminal domain is conserved (46, 47), should be synergistic in combination with a taxane. A deeper understanding of the mechanisms used by a prostate cancer cell to bypass AR inhibition as well as the cellular factors that regulate AR signaling in CRPC is required to develop approaches that will allow men to live with metastatic prostate cancer beyond the 1 to 2 years typically associated with responses to chemotherapy after ADT.

Summary

In summary, our group's work highlights the importance of microtubule-dynein–dependent trafficking for transcription factors, such as AR, that require rapid and targeted nuclear translocation upon specific stimuli. In addition, this work challenges the existing paradigm whereby the clinical activity of taxanes in prostate cancer is attributed solely to the drugs' antimitotic effects and it highlights the therapeutic importance of the signaling events that are impaired downstream of drug-induced microtubule disruption.

Disclosure of Potential Conflicts of Interest

D.M. Nanus and P. Giannakakou are paid consultants for Sanofi-Aventis. No potential conflicts of interest were disclosed by the other author.

Authors' Contributions

Conception and design: D.M. Nanus
Acquisition of data: M. Thadani-Mulero, P. Giannakakou
Analysis and interpretation of data: P. Giannakakou
Writing, review, and/or revision of the manuscript: M. Thadani-Mulero, D. M. Nanus, P. Giannakakou

Grant Support

This work was supported by grants from the U.S. NIH (R01 CA137020-01 and U54 CA143876, P. Giannakakou), the NIH Physical Sciences Oncology Center at Cornell (P. Giannakakou and D.M. Nanus), and the Weill Cornell Clinical and Translational Science Center (D.M. Nanus and P. Giannakakou), by a Creativity Award from the Prostate Cancer Foundation (P. Giannakakou and D.M. Nanus), and support from the Genitourinary Oncology Research Fund (D.M. Nanus).

Received February 28, 2012; revised April 13, 2012; accepted May 9, 2012; published online September 17, 2012.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin 2012;62:10–29.
2. Huggins C. The treatment of cancer of the prostate: (The 1943 Address in Surgery before the Royal College of Physicians and Surgeons of Canada). Can Med Assoc J 1944;50:301–7.
3. Massard C, Fizazi K. Targeting continued androgen receptor signaling in prostate cancer. Clin Cancer Res 2011;17:3876–83.
4. Attard G, Richards J, de Bono JS. New strategies in metastatic prostate cancer: targeting the androgen receptor signaling pathway. Clin Cancer Res 2011;17:1649–57.
5. Bryce A, Ryan CJ. Development and clinical utility of abiraterone acetate as an androgen synthesis inhibitor. Clin Pharmacol Ther 2012;91:101–8.
6. Ryan CJ, Tindall DJ. Androgen receptor rediscovered: the new biology and targeting the androgen receptor therapeutically. J Clin Oncol 2011;29:3651–8.
7. Harris WP, Mostaghel EA, Nelson PS, Montgomery B. Androgen deprivation therapy: progress in understanding mechanisms of resistance and optimizing androgen depletion. Nat Clin Pract Urol 2009;6:76–85.
8. Tannock IF, de Wit R, Berry WR, Horti J, Pluzanska A, Chi KN, et al.TAX 327 Investigators. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. N Engl J Med 2004; 351:1502–12.
9. Petrylak DP, Tangen CM, Hussain MH, Lara PN Jr, Jones JA, Taplin ME, et al. Docetaxel and estramustine compared with mitoxantrone

and prednisone for advanced refractory prostate cancer. N Engl J Med 2004;351:1513–20.
10. Jordan MA, Wilson L. Microtubules as a target for anticancer drugs. Nat Rev Cancer 2004;4:253–65.
11. Torres K, Horwitz SB. Mechanisms of Taxol-induced cell death are concentration dependent. Cancer Res 1998;58:3620–6.
12. Jackson JR, Patrick DR, Dar MM, Huang PS. Targeted anti-mitotic therapies: can we improve on tubulin agents? Nat Rev Cancer 2007;7:107–17.
13. Berges RR, Vukanovic J, Epstein JI, CarMichel M, Cisek L, Johnson DE, et al. Implication of cell kinetic changes during the progression of human prostatic cancer. Clin Cancer Res 1995;1:473–80.
14. Komlodi-Pasztor E, Sackett D, Wilkerson J, Fojo T. Mitosis is not a key target of microtubule agents in patient tumors. Nat Rev Clin Oncol 2011;8:244–50.
15. Ghosh AK, Steele R, Ray RB. Knockdown of MBP-1 in human prostate cancer cells delays cell cycle progression. J Biol Chem 2006;281: 23652–7.
16. Butterworth KT, McCarthy HO, Devlin A, Ming L, Robson T, McKeown SR, et al. Hypoxia selects for androgen independent LNCaP cells with a more malignant geno- and phenotype. Int J Cancer 2008;123:760–8.
17. Komlodi-Pasztor E, Sackett DL, Fojo AT. Inhibitors targeting mitosis: tales of how great drugs against a promising target were brought down by a flawed rationale. Clin Cancer Res 2012;18:51–63.
18. Gascoigne KE, Taylor SS. Cancer cells display profound intra- and interline variation following prolonged exposure to antimitotic drugs. Cancer Cell 2008;14:111–22.
19. Gundersen GG, Cook TA. Microtubules and signal transduction. Curr Opin Cell Biol 1999;11:81–94.
20. Asrih M, Pellieux C, Papageorgiou I, Lerch R, Montessuit C. Role of ERK1/2 activation in microtubule stabilization and glucose transport in cardiomyocytes. Am J Physiol Endocrinol Metab 2011;301:E836–43.
21. Babich A, Burkhardt JK. Lymphocyte signaling converges on microtubules. Immunity 2011;34:825–7.
22. Etienne-Manneville S. From signaling pathways to microtubule dynamics: the key players. Curr Opin Cell Biol 2010;22:104–11.
23. Giannakakou P, Sackett DL, Ward Y, Webster KR, Blagosklonny MV, Fojo T. p53 is associated with cellular microtubules and is transported to the nucleus by dynein. Nat Cell Biol 2000;2:709–17.
24. Giannakakou P, Nakano M, Nicolaou KC, O'Brate A, Yu J, Blagosklonny MV, et al. Enhanced microtubule-dependent trafficking and p53 nuclear accumulation by suppression of microtubule dynamics. Proc Natl Acad Sci U S A 2002;99:10855–60.
25. Mabjeesh NJ, Escuin D, LaVallee TM, Pribluda VS, Swartz GM, Johnson MS, et al. 2ME2 inhibits tumor growth and angiogenesis by disrupting microtubules and dysregulating HIF. Cancer Cell 2003;3: 363–75.
26. Carbonaro M, O'Brate A, Giannakakou P. Microtubule disruption targets HIF-1alpha mRNA to cytoplasmic P-bodies for translational repression. J Cell Biol 2011;192:83–99.
27. Carbonaro M, Escuin D, O'Brate A, Thadani-Mulero M, Giannakakou P. Microtubules regulate hypoxia-inducible factor-1a protein trafficking and activity: implications for taxane therapy. J Biol Chem 2012; 287:11859–69.
28. Roth DM, Moseley GW, Glover D, Pouton CW, Jans DA. A microtubule-facilitated nuclear import pathway for cancer regulatory proteins. Traffic 2007;8:673–86.
29. Harrell JM, Murphy PJ, Morishima Y, Chen H, Mansfield JF, Galigniana MD, et al. Evidence for glucocorticoid receptor transport on microtubules by dynein. J Biol Chem 2004;279:54647–54.
30. Lam MH, Thomas RJ, Loveland KL, Schilders S, Gu M, Martin TJ, et al. Nuclear transport of parathyroid hormone (PTH)-related protein is dependent on microtubules. Mol Endocrinol 2002;16:390–401.
31. Darshan MS, Loftus MS, Thadani-Mulero M, Levy BP, Escuin D, Zhou XK, et al. Taxane-induced blockade to nuclear accumulation of the androgen receptor predicts clinical responses in metastatic prostate cancer. Cancer Res 2011;71:6019–29.
32. Zhu ML, Horbinski CM, Garzotto M, Qian DZ, Beer TM, Kyprianou N. Tubulin-targeting chemotherapy impairs androgen receptor activity in prostate cancer. Cancer Res 2010;70:7992–8002.
33. Gan L, Chen S, Wang Y, Watahiki A, Bohrer L, Sun Z, et al. Inhibition of the androgen receptor as a novel mechanism of taxol chemotherapy in prostate cancer. Cancer Res 2009;69:8386–94.
34. Kuroda K, Liu H, Kim S, Guo M, Navarro V, Bander NH. Docetaxel down-regulates the expression of androgen receptor and prostate-specific antigen but not prostate-specific membrane antigen in prostate cancer cell lines: implications for PSA surrogacy. Prostate 2009;69:1579–85.
35. Chen CD, Welsbie DS, Tran C, Baek SH, Chen R, Vessella R, et al. Molecular determinants of resistance to antiandrogen therapy. Nat Med 2004;10:33–9.
36. Feldman BJ, Feldman D. The development of androgen-independent prostate cancer. Nat Rev Cancer 2001;1:34–45.
37. de Bono JS, Oudard S, Ozguroglu M, Hansen S, Machiels JP, Kocak I, et al.TROPIC Investigators. Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. Lancet 2010;376:1147–54.
38. Hamidovic A, Hahn K, Kolesar J. Clinical significance of ABCB1 genotyping in oncology. J Oncol Pharm Pract 2010;16:39–44.
39. Kavallaris M. Microtubules and resistance to tubulin-binding agents. Nat Rev Cancer 2010;10:194–204.
40. Hu R, Dunn TA, Wei S, Isharwal S, Veltri RW, Humphreys E, et al. Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. Cancer Res 2009;69:16–22.
41. Guo Z, Yang X, Sun F, Jiang R, Linn DE, Chen H, et al. A novel androgen receptor splice variant is up-regulated during prostate cancer progression and promotes androgen depletion-resistant growth. Cancer Res 2009;69:2305–13.
42. Sun S, Sprenger CC, Vessella RL, Haugk K, Soriano K, Mostaghel EA, et al. Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. J Clin Invest 2010;120:2715–30.
43. Watson PA, Chen YF, Balbas MD, Wongvipat J, Socci ND, Viale A, et al. Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. Proc Natl Acad Sci U S A 2010;107:16759–65.
44. Mostaghel EA, Marck BT, Plymate SR, Vessella RL, Balk S, Matsumoto AM, et al. Resistance to CYP17A1 inhibition with abiraterone in castration-resistant prostate cancer: induction of steroidogenesis and androgen receptor splice variants. Clin Cancer Res 2011;17:5913–25.
45. Tran C, Ouk S, Clegg NJ, Chen Y, Watson PA, Arora V, et al. Development of a second-generation antiandrogen for treatment of advanced prostate cancer. Science 2009;324:787–90.
46. Andersen RJ, Mawji NR, Wang J, Wang G, Haile S, Myung JK, et al. Regression of castrate-recurrent prostate cancer by a small-molecule inhibitor of the amino-terminus domain of the androgen receptor. Cancer Cell 2010;17:535–46.
47. Sadar MD. Small molecule inhibitors targeting the "achilles' heel" of androgen receptor activity. Cancer Res 2011;71:1208–13.

Androgen Receptor on the Move: Boarding the Microtubule Expressway to the Nucleus

Maria Thadani-Mulero, David M. Nanus and Paraskevi Giannakakou

Cancer Res 2012;72:4611-4615.

Integrating liquid biopsies into the management of cancer

Giulia Siravegna[1–3], Silvia Marsoni[1], Salvatore Siena[4,5] and Alberto Bardelli[1,3]

Abstract | During cancer progression and treatment, multiple subclonal populations of tumour cells compete with one another, with selective pressures leading to the emergence of predominant subclones that replicate and spread most proficiently, and are least susceptible to treatment. At present, the molecular landscapes of solid tumours are established using surgical or biopsy tissue samples. Tissue-based tumour profiles are, however, subject to sampling bias, provide only a snapshot of tumour heterogeneity, and cannot be obtained repeatedly. Genomic profiles of circulating cell-free tumour DNA (ctDNA) have been shown to closely match those of the corresponding tumours, with important implications for both molecular pathology and clinical oncology. Analyses of circulating nucleic acids, commonly referred to as 'liquid biopsies', can be used to monitor response to treatment, assess the emergence of drug resistance, and quantify minimal residual disease. In addition to blood, several other body fluids, such as urine, saliva, pleural effusions, and cerebrospinal fluid, can contain tumour-derived genetic information. The molecular profiles gathered from ctDNA can be further complemented with those obtained through analysis of circulating tumour cells (CTCs), as well as RNA, proteins, and lipids contained within vesicles, such as exosomes. In this Review, we examine how different forms of liquid biopsies can be exploited to guide patient care and should ultimately be integrated into clinical practice, focusing on liquid biopsy of ctDNA — arguably the most clinically advanced approach.

[1]*Candiolo Cancer Institute – Fondazione del Piemonte per l'Oncologia (FPO), Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Strada Provinciale 142, Km 3.95, 10060 Candiolo, Torino, Italy.*
[2]*Fondazione Italiana per la Ricerca sul Cancro (FIRC) Institute of Molecular Oncology (IFOM), Via Adamello 16, 20139 Milano, Italy.*
[3]*Department of Oncology, University of Turin, Regione Gonzole 10, 10043 Orbassano, Torino, Italy.*
[4]*Niguarda Cancer Center, ASST Grande Ospedale Metropolitano Niguarda, Piazza dell'Ospedale Maggiore 3, 20162 Milano, Italy.*
[5]*Department of Oncology and Haematology–Oncology, University of Milan, Via Festa del Perdono 7, 20122 Milano, Italy.*

Correspondence to A.B.
alberto.bardelli@unito.it

doi:10.1038/nrclinonc.2017.14
Published online 2 Mar 2017

Cancers arise owing to the accumulation of molecular alterations in genes that control cell survival, growth, proliferation, and differentiation within the nascent tumour. Currently, the molecular profile of cancers is typically assessed using DNA and/or RNA obtained from a fragment of the primary tumour or a single metastatic lesion; therapeutic strategies are subsequently defined according to the molecular profile of the tissue. Importantly, however, the molecular profile of tumours evolves dynamically over time. The ability of tumours to evolve in response to a wide variety of endogenous and exogenous selective pressures has several implications: firstly, the genetic make-up of individual cancers is highly heterogeneous; secondly, within a single patient, distinct metastatic lesions can be molecularly divergent; thirdly, therapeutic stress exerted on tumour cells, particularly by targeted drugs, can dynamically modify the genomic landscape of tumours[1–5]. Of note, human blood samples contain materials — including cell-free DNA (cfDNA) and RNA (cfRNA); proteins; cells; and vesicles (such as exosomes) — that can originate from different tissues, including cancers. Indeed, the rapid turnover of cancer cells is postulated to result in the constant release of tumour-derived nucleic acids and vesicles into the circulation, and viable tumour cells can also separate from the tumour to enter the bloodstream. Thus, the ability to detect and characterize circulating cell-free tumour DNA (ctDNA) and/or tumour-derived RNA (predominantly microRNAs (miRNAs)), and circulating tumour cells (CTCs), has enabled clinicians to repeatedly and non-invasively interrogate the dynamic evolution of human cancers (FIG. 1). The possibility of probing the molecular landscape of solid tumours via a blood draw, with major implications for research and patient care, has attracted remarkable interest among the oncology community; the term 'liquid biopsy' is often used to describe this approach[6].

Many studies have illustrated the potential of liquid biopsy approaches to determine the genomic profile of patients with cancer, monitor treatment responses and quantify minimal residual disease, and assess the emergence of therapy resistance. In addition to blood, several other body fluids such as urine[7], saliva[8], pleural effusions[9], and cerebrospinal fluid (CSF)[10], as well as stool[11], have been shown to contain tumour-derived genetic material, and our ability to exploit liquid biopsies for diagnostic purposes will further expand in the future. The analysis of liquid biopsy specimens is, however, challenging because

Key points

- Patient selection is central to the success of targeted therapy; identification of tumour-specific molecular landscapes is pivotal to guiding treatment choices
- The genomic landscape of each individual tumour is heterogeneous and changes over time as a result of the Darwinian clonal evolution imposed on cancer cells by selective pressures, including targeted therapy
- Longitudinal surveillance of clonal evolution is essential for precision medicine, but cannot be effectively achieved using tissue biopsy specimens, owing to sampling issues
- The blood of patients with cancer contains diverse tumour-derived materials, including circulating cell-free tumour DNA (ctDNA), circulating tumour cells, and exosomes
- The sampling and analysis of ctDNA or other circulating tumour components present in biological fluids, termed 'liquid biopsy', enables minimally invasive monitoring of tumour evolution over time in the clinic
- Two different liquid biopsy companion diagnostic tests for *EGFR* mutations in plasma ctDNA have been approved by the regulatory agencies in Europe and the USA for the selection of patients with non-small-cell lung cancer for anti-EGFR treatment in clinical practice

ctDNA is fragmented and highly under-represented compared with germ line cfDNA[12], and only a limited number of CTCs can be isolated from a blood sample[13]. Thus, analysis of tumour material obtained by liquid biopsies requires highly sensitive assays, which became available only within the past 5 years[14]. Herein, we provide a brief overview of the various types of tumour-derived material that can be sampled using liquid biopsies. Subsequently, we discuss how different forms of liquid biopsy can be exploited in patient care and should ultimately be integrated into clinical practice, focusing primarily on those related to ctDNA that, arguably, have the greatest clinical utility at present.

Blood-based liquid biopsies

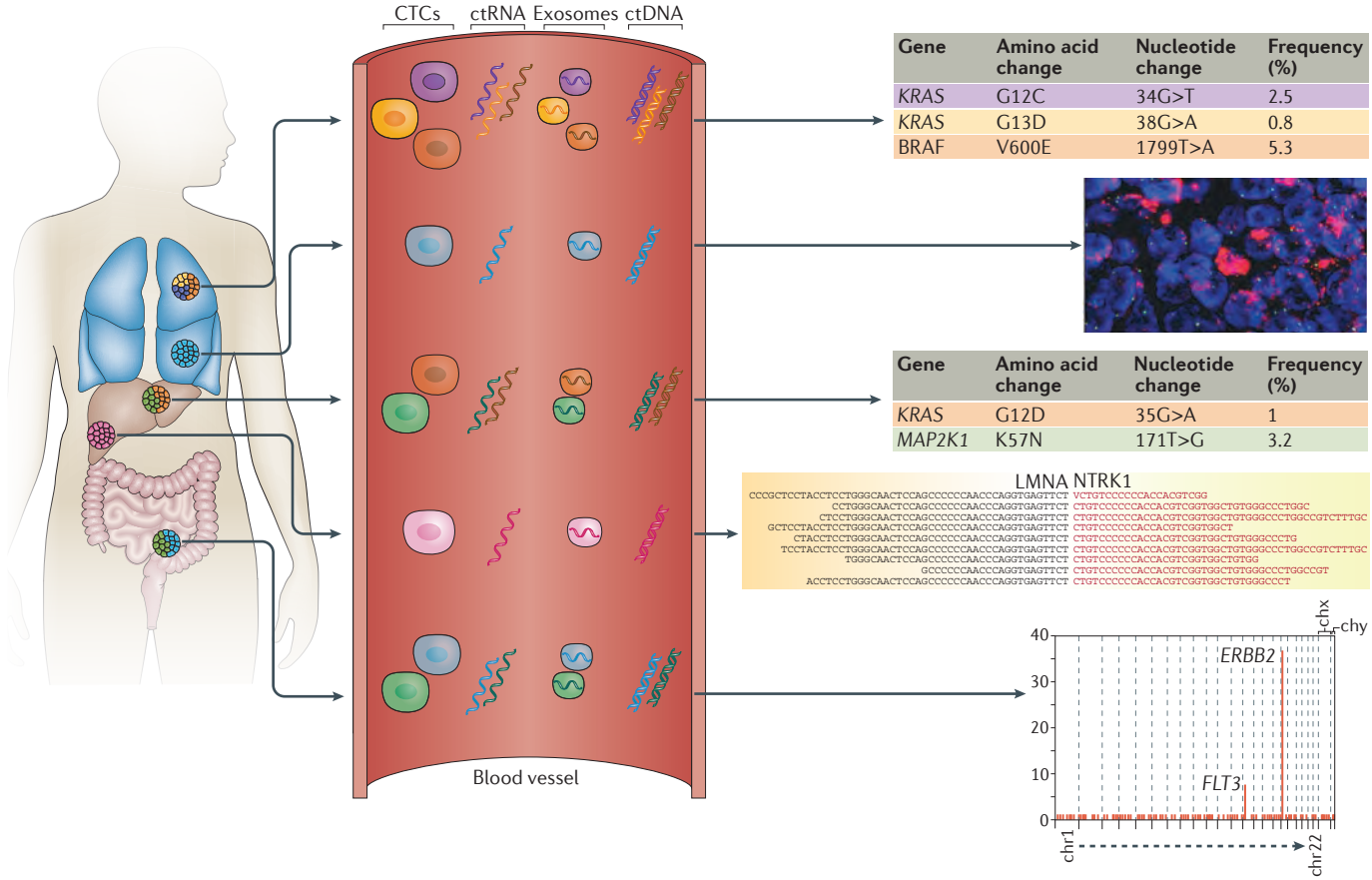
A range of tumour components can be isolated from the blood. Molecular analysis of these different components can provide distinct and complementary information (TABLE 1).

CTCs: content analysis and experimental models. CTCs are tumour cells that have, presumably, intravasated or been passively shed from the primary tumour and/or metastatic lesions into the bloodstream. The existance of CTCs was first reported in 1869 by the Australian physician Thomas Ashworth[15], but their clinical utility was not appreciated until the late 1990s[16]. CTCs can be isolated from the blood of patients with cancer, either as single cells or in cell clusters, and the number of CTCs detected has been associated with treatment outcomes and overall survival[17]. The abundance of CTCs in the blood is low (approximately 1 cell per 1×10^9 blood cells in patients with metastatic cancer), however, and varies between tumour types[18].

Multiple technologies have been implemented to isolate CTCs[13,19–22], and have been extensively reviewed elsewhere[23]. Briefly, CTCs can be isolated through negative-enrichment based on their size and other biophysical properties, or by positive enrichment using markers commonly expressed on the surface of these cells, such as epithelial cell adhesion molecule (EpCAM)[23]. Of note, we still lack markers that enable CTCs to be distinguished from nonmalignant epithelial cells, although direct imaging has been used in combination with functional assays to identify CTCs. Size-based selection methods exploit the fact that CTCs are usually larger than normal blood cells[24,25]. Importantly, however, this approach probably results in considerable loss of CTCs that could be overcome by using cocktails of antibodies for positive selection of these cells. Moreover, various approaches, including protein, DNA, and RNA analyses, can be applied to explore the content and facilitate the identification of CTCs[21].

Interestingly, once isolated *ex vivo*, CTCs can be molecularly characterized and used in functional assays *in vitro* and *in vivo*, in order to provide insights into the cancer biology, as reported by several groups[26,27]. Notably, Yu and colleagues[28] isolated CTCs from six patients with metastatic luminal breast cancer and cultured them *in vitro* for >6 months. Establishing cell lines of CTCs derived from other cancer types has, however, proved challenging. For instance, creation of permanent cell lines using CTCs isolated from patients with colorectal cancer is ineffective, with only one succesful example reported to date[29]. By contrast, efforts to expand CTCs in immunocompromised mice have met with considerable success. Xenograft models have been established using CTCs obtained from patients with metastatic luminal breast cancer[30,31]. Moreover, Hodgkinson and colleagues[32] have demonstrated that CTCs from patients with either chemosensitive or chemorefractory small-cell lung cancer (SCLC) form tumours in immunocompromised mice, and that the response to chemotherapy in these CTC-derived xenografts (CDXs) models seem to mirror that of the donor patient. Importantly, whole-genome sequencing (WGS) analysis revealed that the isolated CTCs and corresponding CDX had comparable genomic profiles[32]. Using this approach, blood samples collected from a patient before and after drug-resistant disease relapse could potentially be used to generate CDX models that could subsequently be characterized and compared in order to search for new druggable targets; routine *in vivo* testing of the corresponding targeted therapies would also be possible, thus facilitating the development of personalized medicine strategies. The timescale required to produce such models will almost certainly preclude application of the findings to tailor therapy for the original CTC donor, but the information gleaned could be used to generate hypotheses for future reseach. In addition to CDX models, CTCs isolated from patients with prostate cancer have been successfully cultured as 3D organoids, which might recapitulate the molecular complexity of different prostate cancer subtypes[33].

Remarkably, Khoo and colleagues[34] demonstrated an efficient approach for evaluating treatment sensitivity using patient-derived CTCs cultured in microfabricated tapered microwells coupled to a microfluidics platform. Drug screening was performed without pre-enrichment of CTC clusters, enabling rapid feedback (after 2 weeks) and, therefore, immediate intervention upon detection of drug resistance or tolerance[34]. The CTC-culturing procedure was clinically validated using 73



Gene	Amino acid change	Nucleotide change	Frequency (%)
KRAS	G12C	34G>T	2.5
KRAS	G13D	38G>A	0.8
BRAF	V600E	1799T>A	5.3

Gene	Amino acid change	Nucleotide change	Frequency (%)
KRAS	G12D	35G>A	1
MAP2K1	K57N	171T>G	3.2

Figure 1 | **Liquid biopsies capture the molecular heterogeneity of metastatic cancers.** Liquid biopsies of tumour components in the blood, including circulating cell-free tumour DNA (ctDNA) and RNAs (ctRNA), exosomes, and circulating tumour cells (CTCs), can be leveraged to capture the molecular heterogeneity of distinct tumour lesions harbouring different genetic alterations. The tables illustrate the detection of point mutations in oncogenes (G12D, G12C, and G13D mutations in *KRAS*; V600E mutation of *BRAF*; and K57N mutation of *MAP2K1* (MEK1)) through candidate-gene or next-generation sequencing analysis of ctDNA in plasma. Identification of a *LMNA–NTRK* gene fusion is depicted in the sequence alignment. Gene copy-number variations can also be detected using fluorescence *in situ* hybridization analysis of CTCs isolated from the blood, as shown in the fluorescence micrograph, and via whole-exome analysis of blood-derived tumour DNA by next-generation sequencing (as demonstrated for *FLT3* and *ERBB2 (HER2)* amplification).

blood samples from 55 patients with early stage or metastatic breast cancer, and in 24 samples used for drug screening; the researchers observed that the potential for CTC-cluster formation was inversely correlated with increasing drug concentrations, and proposed that increases in the drug concentrations needed to prevent CTC-cluster formation during the course of treatment might portend the onset of drug resistance or tolerance[34].

Exosomes. Several types of microvesicles can be released from non-neoplastic and tumour cells[35]. Extracellular vesicles (EVs) can be classified into two groups: the first comprises microvesicles shed directly from the cell membrane via budding; the second consists of exosomes, which are exuded via exocytosis when multivesicular bodies (MVBs) fuse with the plasma membrane[36]. Exosomes were first described in 1983 by Pan and Johnstone[37], and are EVs of 40–100 nm in size that are released by several cell types into the extracellular space and a variety of body fluids[38,39]. Blood cells, endothelial cells, immunocytes, platelets, and smooth-muscle cells are known to release exosomes[40,41]. Exosomes can be extracted from body fluids by normal density-gradient centrifugation[42]. Alternatively, exosomes can be isolated through ultracentrifugation, visualized by transmission microscopy, or selected based on the presence of specific protein markers, such as the tetraspanin proteins CD63, CD9, and CD81 (REFS 38,39). Exosomes have important roles in exchanging molecular information between cells: they have been shown to contain proteins as well as a range of nucleic acids, including DNA, mRNAs, and miRNAs, suggesting that they can modulate the activity of the recipient cells[38,39]. Given their content, exosomes and other EVs could potentially be exploited as cancer biomarkers. Moreover, exosomal miRNAs seem to be involved in disease progression; for example, they can stimulate angiogenesis and promote metastasis[43]. Harvesting of exosomes from biological fluids enables the isolation and subsequent analysis of mRNA, and thus the detection of mutations, splice variants, and

gene fusions, as well as gene-expression profiling[44]. In comparison with ctDNA fragments, of which only two copies are essentially present in the tumour cell of origin, mRNA originating from a highly expressed gene could occur in thousands of copies per cell and might be shed into the circulation (within EVs or as cfRNA) at higher concentrations; therefore, analysis of exosomal mRNA might be advantageous, especially in patients with limited amounts of detectable ctDNA.

Circulating RNAs. In 1996, the detection of circulating tumour-associated mRNA was first described in the blood of patients with melanoma[45], and soon after, other forms of RNAs — mostly miRNAs and long noncoding RNAs (lncRNAs) — were identified in the circulation of patients with solid cancers[46]. The presence of tumour-derived mRNA in blood is of clinical relevance for a number of reasons, including the identification of tumour-specific gene-expression profiles. Somatic mutations in DNA only represent a subset of the molecular alterations associated with cancers, and do not fully recapitulate changes in gene-expression profiles that might result from epigenetic alterations, the effects of miRNAs, or other mechanisms. If feasible, therefore, blood-based RNA profiling of cancers could provide highly valuable information.

miRNAs are the most abundant cfRNA molecules in the blood, and can be carried in exosomes[46], apoptotic bodies, protein–miRNA complexes[47], and tumour-educated platelets (TEP)[48]. The landscape of miRNAs in blood seems to correlate with that of the solid tumours from which they originate[49]. The amount and composition of exosomal miRNAs differs between patients with cancer and individuals without this disease, implicating miRNAs as potential non-invasive diagnostic biomarkers[46,50,51]. Whether cfRNA originates preferentially from tumour cells, or whether cells of the haematopoietic system make a strong contribution to cfRNA levels (possibly owing to response of immune cells to the disease[52]), remains controversial. Most of the analyses of mRNA and miRNA in blood remain exploratory, and validation in clinical studies with standardized protocols is required to substantiate the value of cfRNAs in the clinical setting (BOX 1). The clinical implications of cfRNA are currently unclear; an in-depth discussion of this topic is beyond the scope of this Review, and this subject has been comprehensively reviewed in this journal and in others[53,54]. An example of clinical utility is, however, provided by the NETest, which can provide insight into the activity of neuroendocrine tumours via targeted expression profiling of 51 genes using mRNA isolated from peripheral blood samples[55].

ctDNA: mechanisms of release, characteristics, quantity, and quality. The existence of cfDNA was first described in 1948 by Mandel and Metais[56], who detected non-cell-bound nucleic acids in the bloodstream of individuals with cancer. Information about the origin and release of cfDNA in such individuals is limited, although inflammation, autoimmunity, smoking, pregnancy, exercise, and heart dysfunction have been shown to contribute to the presence of cfDNA[57].

In 1977, Leon and co-workers[58] reported a higher amount of cfDNA in patients with cancer than in individuals without this disease. This seminal discovery triggered further research: Stroun and colleagues[59] discovered that tumour-related genetic alterations were present in the cfDNA of patients with cancer, and research by other groups confirmed that various types of neoplastic genomic alterations (that is, mutations in oncogenes and/or tumour-suppressor genes[60], microsatellite instability[61], and epigenetic changes[62]) can be detected in 'ctDNA' — as part of the total cfDNA pool, but specifically derived from tumours. The exact mechanisms by which ctDNA is released into the blood remain to be clarified. In addition to apoptosis, DNA fragments can be released from cancer cells via other mechanisms, including necrosis[63]. Indeed, tumour masses often have high levels of necrosis, which might lead to the release of tumour DNA in body fluids. Furthermore, macrophages seem to have a role in the release of tumour DNA fragments in the circulation, via phagocytosis of necrotic neoplastic cells[64]. WGS of plasma cfDNA isolated from pregnant women revealed a fragmentation pattern (~142 bp) reminiscent of that of nuclease-cleaved nucleosomes[65], which suggests that cfDNA is generated by apoptotic degradation of cellular DNA. In other studies, the size distribution of cfDNA in healthy individuals and patients with cancer has been compared, also revealing an enrichment of fragments the size of single or multiple nucleoprotein complexes[66], which further suggests apoptosis is the main source of both total cfDNA and ctDNA[12].

Once in the circulation, clearance of cfDNA is rapid and occurs via the kidneys, liver, and spleen[67]; the half-life of fetal cfDNA in pregnant women is approximately 16 minutes[67]. How pharmacological treatments, inflammation, and circadian rhythms contribute to cfDNA

Table 1 | **Comparison between the applications of ctDNA, CTCs, and exosomes**

	ctDNA/RNA	CTCs	Exosomes
Potential to fully recapitulate spatial and temporal tumour heterogeneity	Yes[3,4,164]	No	No
Assesment of pre/post-analytical variability	Yes[12,68]	Yes[201]	Yes[35]
Detection of somatic mutations, InDels, copy-number alterations and gene-fusions	Yes[1–3,7,10,11,60,64,71,72,75,92,125,129–132,148,149,151–156,158,189,202]	Yes[13,19,21]	Yes[203,204]
Evaluation of methylation patterns	Yes[137–142,145,146]	Yes[205]	Yes[206]
Analysis of mRNA/miRNA/lncRNA/RNA splice variants	Yes[45,49]	Yes[20]	Yes[40,43,46,51]
Analysis of RNA expression	No	Yes[19,207]	Yes[50,86]
Cell morphology and functional studies *ex vivo*	No	Yes[26–34]	No
Demonstration of signal colocalization	No	Yes[121]	No
Proteomics analysis	No	Yes[116–118]	Yes[50]

'Yes' indicates that the approach is feasible, possible, and/or published studies are available; 'No' indicates that the application is not feasible and/or no studies are available. CTCs, circulating tumour cells; ctDNA, circulating tumour DNA; InDels, DNA insertions and/or deletions; lncRNA, long noncoding RNA; mRNA, messenger RNA; miRNA, microRNA.

Box 1 | **Standardization issues for implementing ctDNA analysis in the clinic**

Pre-analytical variability

- Use specialized collecting tubes (such as Streck tubes) to stabilize blood samples at room temperature and prevent lysis of white blood cells[12]
- Define the optimal time period between blood draws and plasma processing, as well as centrifugation conditions to reach the maximum final ctDNA yield[200]
- Define quantification methods (for example, using fluorescent dyes, spectrophotometry, or qPCR)[199]
- Define ctDNA isolation protocols to reach the maximum final ctDNA yield

Analytical variability

- Intrinsic PCR errors
- Nonuniform genomic coverage
- Technological errors (for example, related to next-generation sequencing platforms)

Biological variability

- Take into account both spatial and temporal tumour heterogeneity
- Currently, the clinical sensitivities of ctDNA assays are challenging to determine, owing to variability in the tumour stages and types assessed, the sample-processing techniques used, and the targeted molecular alterations across different studies

clearance is not entirely understood. For ctDNA analyses, plasma samples are preferable to serum samples[12]: the overall quantity of cfDNA is 2–24-fold higher in the serum[68], probably owing to extensive contamination of DNA released from immune cells that are lysed during the clotting process; therefore, plasma has proved to be a superior source of ctDNA owing to the lower background levels wild-type DNA.

Liquid biopsy of other body fluids

Several studies have revealed the presence of tumour-derived nucleic acids in other body fluids, such as urine[69], saliva[8,70], and CSF[71,72] (FIG. 2). In patients with lung cancer, ctDNA has also been detected in bronchial washings and pleural fluids[73,74]. Indeed, the localization of the primary tumour and of any metastatic lesions seems to have a major effect on the abundence of ctDNA in different body fluids[8,75].

Urine. More than 20 years ago, Zhang and co-workers[76] reported that DNA fragments corresponding to the Y-chromosome gene *SRY* could be detected in cell-free urine supernatant from female recipients of renal transplants from a male donor; however, why genomic DNA can be detected in the urine after renal filtration was not demonstrated until 10 years later[77], when analyses demonstrated that cfDNA in urine, also known as transrenal DNA (tr-DNA), is present as a result of renal clearance of cfDNA from the blood (FIG. 2). In principle, therefore, urine could be a useful source of tumour-derived DNA, particularly considering that urine collection is simple and non-invasive, and the fact that we know very little about the properties of tr-DNA in patients with cancer is somewhat surprising. Glomerular filtration of plasma is highly selective; only molecular complexes <6.4 nm in diameter and with a molecular weight of <70 kDa, equating to ~100 bp of DNA, can transit into the lumen of the nephron. The clinical condition of individual patients probably also affects the passage of nucleic acids from the blood into the urine. For example, smoking has been shown to influence the amount of tr-DNA found in urine[78].

Limited data are available on the correlation between tumour-derived nucleic acids in urine and in matched blood samples from patients with urogenital cancers, although the final concentrations of tumour-derived tr-DNA in the urine after glomerular filtration do not seem to reflect transport of ctDNA from the blood[79]. Instead, the vast majority of urinary tumour tr-DNA fragments detected in patients with urogenital cancers are thought to result from shedding of tumour cells or their breakdown products directly into the urinary tract.

According to their size, tr-DNA fragments in urine can be divided in two main groups: high-molecular-weight DNA (≥1 kbp) and low-molecular-weight DNA (<100 bp)[77,80–82]. Next-generation sequencing (NGS) approaches have been applied to assess the size of tr-DNA fragments in the urine of pregnant women, revealing that most maternal and fetal tr-DNA fragments contain fewer than 100 bp, with a peak at 29 bp[83]. As in blood, DNase I is the major DNA-hydrolysing enzyme present in urine, but the activity of this enzyme in urine is more than 100-fold higher than that in the blood, potentially explaining the greater fragmentation of tr-DNA versus cfDNA[84].

Owing to a high concentration of RNA-hydrolysing enzymes, mRNA is not preserved in urine; however, miRNAs are more resistant to nucleases owing to their small size (20–25 nucleotides) and can be detected in urine[85]. Moreover, miRNAs are more stable than mRNAs, as they are often present within EVs, such as exosomes[86], or bound to the protein argonaute-2 (REF. 87).

Importantly, with regard to clinical relevance, tumour-specific genetic alterations commonly detected in plasma ctDNA, such as point mutations and methylation profiles, can also be detected in urinary tr-DNA from patients with cancer[88,89]. For instance, in one study[77], *KRAS* mutations in tr-DNA were detected in five of eight patients with stage IV pancreatic cancer, and in four of five patients with stage III–IV colorectal adenocarcinoma. Reckamp and co-workers[7] reported the ability to detect mutations in exons 19, 20, and 21 of *EGFR* in both plasma and matched urine samples from patients enrolled on the phase I/II TIGER-X trial (NCT01526928) of rociletinib in previously treated patients with *EGFR*-mutated advanced-stage non-small-cell lung cancer (NSCLC). Using the mutation status of 60 evaluable tissue samples as a reference, the sensitivity of *EGFR*-mutation detection in urine was 72% (34 of 47 patients) for T790M, 75% (12 of 16 patients) for L858R, and 67% (28 of 42 patients) for exon 19 deletions, with higher sensitivities — comparable to those achieved with plasma — reported for samples that met the recommended volume[7]. Remarkably, combined urine and plasma testing resulted in the identification of 12 additional T790M-positive patients, in whom T790M mutations were missed using tumour tissue analysis[7], indicating that urine ctDNA analysis might provide complementary information about a patient's mutational status that is not captured by plasma or tissue tests.

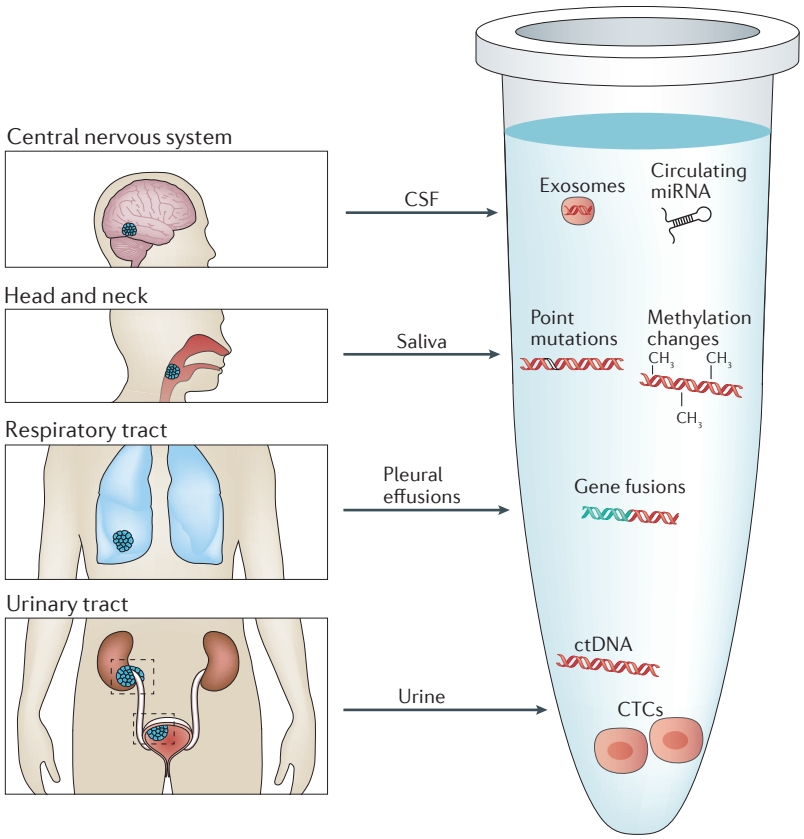


Figure 2 | **Body fluids as a source of tumour-derived molecular information.** Schematic representation of different body fluids (other than blood) that can contain tumour-derived molecular information, specifically, urine, cerebrospinal fluid (CSF), saliva, and pleural effusions. The localization of the primary tumour and of any metastatic lesions influences the presence of circulating tumour-derived nucleic acids, cells, and microvesicles in individual body fluids. As indicated, point mutations, gene fusions, and methylation changes associated with the cancer can be detected through analysis of DNA or mRNA derived from these blood-borne tumour materials. CTCs, circulating tumour cells; ctDNA, circulating cell-free tumour DNA; miRNA, microRNA.

At present, quantification of tumour-derived tr-DNA in urine is technically challenging, owing mainly to the low amounts present, although the continuous development of DNA-amplification and sequencing technologies will probably facilitate this approach. Moreover, we should be mindful of the possibility that liquid biopsy of urine could be preferable to the use of other body fluids, as this approach is a truly non-invasive alternative to biopsy sampling and can be performed at home by the patients themselves. Liquid biopsy of tr-DNA offers the fascinating possibility of monitoring minimal residual disease (MRD) after surgery with curative intent.

Cerebrospinal fluid. CSF is secreted by the choroid plexuses of the ventricles of the brain[90]. Given that CSF is in direct contact with all the cells of the central nervous system (CNS), including any cancer cells present, this biological fluid has been exploited to profile 'ctDNA' in patients with brain tumours[71,91] (FIG. 2). Of note, although ctDNA can be detected in plasma from most patients with metastatic cancer, individuals with brain tumours, including those with high-grade gliomas and medulloblastomas, have only low levels of ctDNA in their blood, probably owing to the presence of the blood–brain barrier[92]. Lumbar puncture to obtain CSF is an invasive procedure, but is routinely performed in patients with brain tumours (including meningiomas, glioblastomas and medulloblastomas) as well as those with leptomeningeal carcinomatosis and lymphomas[93]. Evidence suggests that CSF is also suitable for analysis of tumour-derived DNA in patients with tumours that have metastasized to the brain[71]. Importantly, De Mattos-Arruda and colleagues[10] have observed a strong association of the tumour type (grade III or IV glioma, medulloblastoma, or ependymoma) and tumour location (proximity to a CSF reservoir or the cortical surface) with the presence of tumour-derived DNA in the CSF. Considering, however, that CSF quickly circulates throughout the ventricles and spinal reservoirs[94], liquid biopsies of CSF obtained through lumbar puncture could potentially be exploited for ctDNA analyses.

The potential of CSF-ctDNA for characterizing and monitoring brain tumours has been investigated, in comparison with the use of plasma ctDNA[10]. Matched samples of CSF, plasma, and tumour tissue DNA from patients with glioblastoma, medulloblastoma, or brain metastases from lung or breast cancer were analysed[10]. Notably, ctDNA derived from tumours located in the CNS was found to be more abundant in CSF than in plasma, and CSF-ctDNA could be used to detect somatic mutations as well as to longitudinally monitor tumour burden[10]. Additional studies are needed to compare ctDNA and cytology analysis of CSF from patients with cancer, but the two approaches will probably complement the use of other biomarkers, radiological imaging, and clinical parameters[95,96].

miRNAs can also be identified in CSF samples; in a few studies, analysis of circulating free miRNA has been exploited to diagnose primary CNS lymphoma[97], and exosome-derived miRNA has been used as a biomarker for monitoring therapeutic responses in patients with glioblastoma[98]. Interestingly, CSF exosomes seem to be enriched for miRNAs relative to other types of CSF EVs (specifically, microvesicles), whereas the distribution of miRNA between different plasma EV types is less predictable[99].

Saliva. Few studies have examined the presence of tumour-derived DNA in saliva (FIG. 2). Notably, however, Wang and colleagues[8] analysed DNA in saliva for potential tumour biomarkers in 93 patients with head and neck squamous-cell carcinoma (HNSCC). The authors hypothesized that tumour DNA might be released from the basal side of HNSCC cells[8]. Thus, saliva and plasma samples were screened for human papillomavirus (HPV) genes or somatic mutations in genes or genomic regions commonly altered in HNSCC (such as *TP53*, *PIK3CA*, *CDKN2A*, *HRAS*, and *NRAS*) using multiplex PCR and massively parallel sequencing, and when both plasma and saliva were tested in combination, ctDNA was identified in 96% of patients independent of tumour site[8]. In saliva, however, tumour DNA was found in 100% of patients with oral cancers, compared

with 47–70% of patients with cancers of the other sites (oropharynx, larynx, orypopharynx), while ctDNA was found in plasma from 80% and 100% of these patient groups, respectively[8]. The authors, therefore, concluded that, whereas plasma is enriched for tumour DNA from other sites, saliva is enriched for tumour DNA originating from the oral cavity, making it a valuable biomarker for the detection of oral HNSCC[8]. Importantly, in this proof-of-principle study[8], even early stage oral cancers were associated with highly detectable levels of tumour-derived DNA in saliva. This finding demonstrates the importance of examining specific bodily fluids according to the anatomical location of the tumour in order to achieve optimal sensitivity of ctDNA detection. If confirmed in other studies, saliva-based tests could be incorporated into diagnostic practice, to complement routine examinations, as well as disease monitoring and clinical decision-making. Saliva also contains EVs (including exosomes) that can carry distinct miRNAs implicated in oral cancers, such as oral squamous-cell carcinoma, and their identification holds promise for both diagnostic and prognostic biomarker assessments[100,101].

Pleural effusion fluids and bronchial washings. Pleural effusion fluid[9,73] and bronchial washing samples collected with physiological saline solutions are currently used in diagnosing cancers of the respiratory system. In this setting, detection of *EGFR* mutations through cytology approaches is feasible, although often difficult owing to the limited number of cancer cells that are usually available for analysis. In an alternative approach, Kimura *et al.*[9] assessed the feasibility of detection of activating *EGFR* exon 18–21 mutations in cfDNA present in pleural effusion fluid from patients with NSCLC. Their findings indicate that the *EGFR* mutational status can be accurately ascertained using tumour-derived cfDNA from pleural effusion fluid, and is correlated with responsiveness to EGFR tyrosine-kinase inhibitors (TKIs)[9].

In another study[74], the feasibility of identifying *EGFR* mutations in tumour-derived DNA collected through bronchial washings, termed cytology cell-free DNA (ccfDNA), was examined. The results demonstrated the high sensitivity and specificity (88% and 100%, respectively) of this approach compared with the analysis of DNA from tumour tissue[74], suggesting that activating *EGFR* mutations can be accurately detected in ccfDNA. Thus, ccfDNA might be a valuable alternative to cytological samples, although larger investigations are needed to validate this diagnostic approach.

A limited number of studies have investigated the diagnostic, prognostic, or predictive value of miRNAs in pleural effusion fluid from patients with NSCLC[102,103]. In one study[102,103], the authors found that a signature comprising five miRNAs in the effusion samples was predictive of the overall survival of patients with NSCLC and malignant pleural effusion.

Technologies to analyse ctDNA

As discussed, circulating cfDNA is primarily composed of germ line DNA that originates from normal cells, with a relatively small and highly variable fraction of ctDNA present in patients with cancer. As such, the sensitivity of traditional approaches to DNA analysis (such as Sanger sequencing) is insufficient for detection of somatic mutations in plasma ctDNA from patients with cancer. To overcome these limitations, digital-PCR-based technologies with a high level of analytical sensitivity and specificity have been developed, enabling high-throughput, targeted amplification of the mutant gene of interest on the background of abundant wild-type alleles, reaching limits of detection below 0.0001% (TABLE 2) — which is mandatory for the detection of rare aberrations in ctDNA[104].

In addition, nontargeted genome-wide analyses enable the identification of tumour-specific alterations without prior knowledge of the aberrations likely to be present in the tumour (TABLE 2); therefore, such approaches can be exploited for *de novo* discovery of genetic changes underlying therapy resistance and the identification of new actionable targets in patients with cancer. NGS is a technique that involves immobilization of DNA fragments on a solid support and reading of the sequence as a part of a DNA-synthesis process[105]. Using NGS, millions of ctDNA sequences can be produced in a single reaction, and are subsequently aligned and compared with a reference genome or to the germ line DNA obtained from the same patient (that is, from nonmalignant tissue, typically peripheral blood mononuclear cells), making it possible to identify nucleotide changes (variants or mutations) relative to the reference sequence. Several NGS-based methods have now been devised that enable the detection of not only point mutations and insertions, deletions or rearrangements, but also copy-number alterations and gene fusions.

In the future, digital PCR and NGS will probably both be used complementarily in liquid biopsy analyses. The former approach enables dynamic profiling of individual mutations, but requires *a priori* knowledge of the mutant allele, whereas the latter technique enables the discovery of novel mutated variants, but has higher costs and cannot be readily applied to monitor patients longitudinally[106]. Further discussions of the technical aspects of liquid biopsy are beyond the scope of this Review.

Clinical applications of liquid biopsies

The potential of liquid biopsy assays is far reaching, and their wide-ranging clinical applications are only starting to emerge. Liquid biopsies can be exploited for diagnostic purposes, to identify and track tumour-specific alterations during the course of the disease, and to guide therapeutic decisions[107]. Clinical implementation will only be achievable, however, if standardized procedures are defined and large validation studies are performed (BOX 1).

CTCs as biomarkers and their clinical utility. At present, the clinical value of CTC analysis remains controversial, although evidence indicates that the abundance of tumour cells in the blood of patients with cancer has prognostic value, and that CTC numbers after treatment can be predictive of response to therapy and, thus, treatment outcomes[108,109]. These findings must be considered

with caution, however, because CTC numbers are highly variable between different tumour types, and are subject to biases relating to the variety of CTC-detection methods used. Moreover, the correlation between the number of CTCs detected and patient survival is far from being defined[108], and this limitation is likely to be overcome only by combining different technologies to improve assay performance.

The only FDA-approved platform for the isolation and enumeration of CTCs in patients with metastatic breast, colorectal, or prostate cancer is CellSearch[110–112], a semi-automated system that enables positive selection of CTCs based on the expression of the epithelial marker EpCAM and the lack of expression of the leukocyte-specific molecule CD45. Using CellSearch, CTCs have been shown to be present in the peripheral blood of patients with most types of carcinomas (including those of the prostate, breast, ovary, colorectum, or lung), but not in the blood of those without cancer[113]. In a separate study, the numbers of CTCs detected at baseline and in subsequent blood draws were found to be good predictors of the progression-free survival (PFS) and overall survival of patients with metastatic breast cancer[114]. Data suggest that CellSearch can also be applied in the nonmetastatic setting, as long as expert training and central image review is performed, with the detection of CTCs in blood from patients with stage I–III breast cancer independently predicting worse disease-free, metastasis-free, breast-cancer-specific, and overall survival, compared with the outcomes of patients without detectable CTCs[115].

The biology and clinical implications of CTCs will hopefully become better understood with the availability of more-comprehensive molecular characterization and functional analyses. RNA *in situ* hybridization enables differentiation of epithelial from mesenchymal cancer cells according to the expression levels of markers specific for these cell types[116]; importantly, RNA can also be extracted from CTCs and sequenced, and in patients with prostate cancer, this approach has enabled the identification of specific gene fusions (such as *TMPRSS2–ERG* fusions)[117]. In another study[118], microfluidic-based single-cell mRNA-expression analysis enabled transcriptional profiling of 87 cancer-associated and reference genes in individual CTCs from patients with breast cancer, revealing elevated expression of genes associated with metastasis and with epithelial-to-mesenchymal transition, as well as providing insights into tumour heterogeneity.

Moreover, CTCs can be exploited to investigate the presence of drug targets, as a surrogate for tumour biopsy specimens[119]. In this regard, intratumoural expression of programmed cell death 1 ligand 1 (PD-L1) has been highlighted as a key factor that prevents the

Table 2 | **Comparison of technologies for ctDNA analysis**

Approach	Method	Technology	LoD	Advantages	Disadvantages
Candidate-gene analysis	qPCR*	PNA clamp-PCR‡ (REF. 104)	0.1%	• Rapid • High sensitivity • Suitable for the detection of specific point mutations, copy-number variations, short indels, and gene fusions • No bioinformatic analysis • Cost-effective	• Only enables monitoring of known mutations
		LNA/DNA-PCR‡ (REF. 208)	0.1%		
		ARMS[209]	0.05–0.1%		
		COLD-PCR‡ (REF. 210)	0.1–0.01%		
	Digital PCR	BEAMing[211]	0.01%		
		ddPCR[212–214]	0.001%		
	NA	InPlex§ (REFS 133,215,**216**)	<0.01%		
		Endpoint PCR¶ (REF. 156)	<0.0001%		
Deep-sequencing	Targeted	AmpliSeq[217]	>2%	• Does not require any prior knowledge of the molecular alteration	• Longer time needed to obtain, process and analyse results than that needed for candidate-gene analysis • Bioinformatic expertise required • Expensive
		TAm-Seq[218]	>2%		
		SAFE-SeqS[219]	0.1%		
		Guardant360 digital sequencing test[197]	<0.1%		
		CAPP-Seq[134]	0.01%		
		iDES[220]	<0.01%		
		PARE[221]	0.001%		
	WES (nontargeted)	NA	>1–3%		
	WGS (nontargeted)	Digital karyotyping[221–223]	0.001%		
		PARE[221,224,225]	0.001%		

ARMS, amplification refractory mutation system; BEAMing, beads, emulsion, amplification, magnetics; CAPP-Seq, cancer personalized profiling by deep sequencing; COLD-PCR, complete enrichment coamplification at lower denaturation temperature PCR; ctDNA, circulating cell-free tumour DNA; ddPCR, droplet digital PCR; EMA, European Medicines Agency; FDA, US Food and Drug Administration; iDES, integrated digital error suppression; LNA/DNA-PCR, locked nucleic acids/DNA chimera PCR; LoD, limit of detection; NA, not applicable; PARE, parallel analysis of RNA ends; PNA clamp-PCR, peptide nucleic acids clamp PCR; qPCR, quantitative PCR; SAFE-SeqS, safe-sequencing system; TAm-Seq, tagged-amplicon deep sequencing; WES, whole-exome sequencing; WGS, whole-genome sequencing. *The therascreen EGFR RGQ PCR Kit and cobas EGFR Mutation Test v2 are qPCR assays approved by the EMA and FDA, respectively, for the analysis of plasma ctDNA for *EGFR* mutations that determine eligibility of patients with non-small-cell lung cancer for treatment with EGFR tyrosine-kinase inhibitors. ‡Semiquantitative technologies. §InPlex allele-specific blocker qPCR involves the construction of original and specific PCR primers. ¶Endpoint PCR incorporates an increased number of cycles, which enables an amplification plateau to be reached.

immune system from destroying cancer cells[120]; therefore, characterization of tumour cells for PD-L1 expression will probably be fundamental to the success of this form of immunotherapy. Mazel *et al.*[121] have provided evidence that PD-L1 is frequently expressed on CTCs in the blood of patients with hormone-receptor-positive, HER2-negative breast cancer. PD-L1-expressing CTCs were present in 11 of 16 (68.8%) patients with tumour cells detectable in the blood, with the fraction of PD-L1-positive CTCs varying from 0.2–100%. This CTC-based PD-L1 assay might be used in future clinical trials for stratification of patients and monitoring of response to immune-checkpoint blockade.

ctDNA in cancer diagnosis. Quantitative analysis of cfDNA can be used to assess tumour burden — with diagnostic implications. For example, the amount of cfDNA present in plasma is substantially higher in patients with cancer than in healthy individuals or in patients with benign diseases, and seems to increase with tumour stage (and presumably, therefore, tumour volume)[122,123]. Moreover, cfDNA measurements could potentially be used to determine if a patient is disease free after curative surgery[122–124] (and ctDNA analysis will be more useful in this regard: see 'MRD monitoring and early diagnosis of relapse — ctDNA as prognostic biomarker'). In the diagnosis of cancer, the absolute levels of cfDNA in the circulation provide limited information[122–124]; however, when levels of cfDNA are coupled with identification of somatic mutations (that is, focusing on ctDNA), they provide valuable diagnostic information[124,125]. Somatic mutations are tumour specific and, as a result, evaluation of these aberrations in ctDNA offers the potential for better diagnostic accuracy than can be achieved with standard protein biomarkers, such as carcinoembryonic antigen (CEA)[122–124]. A number of studies have demonstrated the potential of liquid biopsy assays in the early diagnosis of cancer. For example, ctDNA analysis of Epstein–Barr virus (EBV) has been used for the early detection of nasopharyngeal carcinoma[126]. Specifically, Chan *et al.*[126] screened 1,318 asymptomatic volunteers and detected viral DNA in 69, with further investigation uncovering the presence of nasopharyngeal cancers in three of the EBV-positive patients, demonstrating the utility of ctDNA-based screening. In a separate study, *KRAS* mutations were detected in 13 out of 1,098 healthy volunteers, and within 25 months, six of the 13 individuals were found to have cancer (of the bladder or respiratory apparatus)[127]. An open question is how liquid-biopsy-based approaches to early detection of cancer might affect patient outcomes. Improvements would be expected through detection of tumours at an earlier stage, which are generally more amenable to curative treatment. Importantly, however, many precancerous benign conditions have been shown to carry common mutations shared with malignant tumours; therefore, highly sensitive ctDNA analyses could lead to high rates of false positives and, consequently, to overdiagnosis and overtreatment, similar to the results obtained by highly sensitive radiological screening[128].

Importantly, the results of several studies have demonstrated a high concordance between mutational profiles of candidate genes in matched tumour and plasma DNA samples from patients with breast cancer[92,129,130], colorectal cancer[1,92,131–133], or NSCLCs[134,135]. For example, Bettegowda *et al.*[92] characterized a large population of patients with different tumour types using both digital PCR and NGS approaches; mutant ctDNA was identified in 75% of 640 patients with advanced-stage pancreatic, ovarian, bladder, gastroesophageal, breast, hepatocellular, colorectal, or head and neck cancers, or melanoma, and in >50% of patients with early stage cancers[92]. In a study focused on pancreatic ductal adenocarcinoma, actionable somatic mutations and gene amplifications were identified in nearly 30% of patients, with allele frequencies of mutant *KRAS* in plasma cfDNA varying from undetectable to 87.7%[136]. In an ongoing Canadian study (NCT02251314; Supplementary information S1 (table)), *BRAF* mutations in pre-mortem and postmortem plasma ctDNA, and in matched tumour DNA samples from patients with melanoma are being evaluated in order to understand the quantitative relationship between ctDNA and total tumour burden. In addition to the potential to non-invasively estimate tumour burden, the correlation between mutations present in ctDNA and tumour tissue samples is increasingly important in diagnosing specific molecular tumour subtypes, with implications for precision medicine, as discussed extensively in the following sections.

Methylation profiles in ctDNA — predicting response to chemotherapy. Promoter hypermethylation at specific CpG sites associated with tumour-suppressor genes occurs in many cancers; therefore, methylated ctDNA is a promising biomarker. Several studies have compared aberrant methylation in tumour tissues and matched ctDNA from blood samples, in settings such as lung, gastrointestinal[137], breast[138], ovarian, prostate[139], testicular[140], and head and neck cancer, and in most cases a good correlation was reported[141].

Detection of promoter hypermethylation in ctDNA might have higher sensitivity than analyses of instability in microsatellite DNA (at unique GT/CA repeats, either mononucleotide microsatellite sites, such as BAT25 and BAT26, or dinucleotide sites, such as D2S123, D5S346, and D17S250), which could potentially be further improved if combined with mutational analysis[142]. A seminal study that included both methylation and copy-number analysis at a genome-wide level was carried out by Lo's group[143]. The results indicate that copy-number alterations could be inferred from bisulfite DNA-sequencing data in patients with nonmetastatic cancer, with a diagnostic sensitivity and specificity of 74% and 94%, respectively.

Methylation of the *MGMT* gene (encoding 6-*O*-methylguanine-DNA methyltransferase, an enzyme involved in repairing alkylated guanine in DNA) in ctDNA was first investigated in patients with glioblastoma multiforme (GBM): Balaña *et al.*[144] reported a high level of concordance between *MGMT* promoter

methylation in matched tissue and serum DNA, and a potential correlation between *MGMT* promoter methylation and clinical response to treatment. In particular, increased serum *MGMT* promoter methylation was shown to predict a better response and time to progression after treatment with cytotoxic alkylating agents[144]. Of note, *MGMT* silencing by promoter methylation has been used to identify patients with GBM or metastatic colorectal cancer who are most likely to respond to the alkylating agents dacarbazine or temozolomide[145]. Assessment of the prognostic and predictive value of *MGMT* promoter methylation testing in plasma ctDNA and tumour tissue samples using a technology named 'methyl-BEAMing' (beads, emulsion, amplification, magnetics) revealed that this approach is more specific than other techniques, such as methylation-specific PCR and bisulfite pyrosequencing[146]. Of note, the quantitative assays methyl-BEAMing and bisulfite pyrosequencing outperformed methylation-specific PCR in patients with metastatic colorectal cancer, thus enabling better prediction of treatment response and PFS.

MRD monitoring and early diagnosis of relapse — ctDNA as prognostic biomarker. In principle, liquid biopsy approaches might be well-suited to measuring MRD, as residual tumour components can be detected with high sensitivity, and data from proof-of-concept studies have shown that ctDNA levels can be used to monitor MRD following surgery or other curative treatments[17,147]. Beaver *et al.*[148] were able to detect *PIK3CA* mutations in DNA from plasma samples obtained before surgery in 93% of patients with localized breast cancer and a limited tumour burden; matched blood samples were also collected after surgery from 10 patients (within 14 days for five patients, per protocol, and between 15 and 72 days after surgery in the others), and ctDNA remained detectable in five of these patients, suggesting incomplete eradiction of disease after surgery alone, and predicting the development of recurrent disease in one patient. Similarly, in a prospective cohort of 55 patients with early stage breast cancer receiving neoadjuvant chemotherapy[149], detection of ctDNA in plasma after treatment predicted metastatic relapse with a high level of accuracy; further longitudinal mutation tracking increased the sensitivity for prediction of relapse, anticipating clinical relapse by a median of 7.9 months.

Indeed, ctDNA surveillance, aimed at identifying recurrence in patients with no evidence of disease after primary treatment with curative intent, is a key use of liquid biopsy in clinical trials (Supplementary information S1 (table)). The power of NGS analysis of ctDNA in disease surveillance has been evaluated in a retrospective analysis[150] in 107 patients with diffuse large-B-cell lymphoma (DLBCL) who had achieved complete remission with the EPOCH regimen (comprising etoposide, prednisone, vincristine cyclophosphamide and doxorubicin) in any of three different trials that used EPOCH as a control treatment. All three trial protocols included the longitudinal collection of plasma samples at pivotal therapeutic timepoints. Clonotypes were retrospectively defined in tumour tissues for 86% of the patients. In the same patients, the clonotype was assessed by liquid-biopsy-based assays, as a measure of disease recurrence, and the findings were compared with those of CT and/or PET imaging (the gold standards for disease assessment) at matched timepoints. After a median follow-up duration of 11 years, patients who developed detectable clonotypic ctDNA had a hazard ratio for clinical progression 228 times greater than that of patients with undetectable ctDNA[150]. The positive and negative predictive values (PPV and NPV) of the liquid biopsy approach were 88% and 98%, respectively, and recurrence was identified a median of 3.5 months before clinical evidence of disease, suggesting that interim liquid biopsy is a promising biomarker to identify patients at high risk of treatment failure[150].

In a landmark study, Diehl and co-workers[131] used a BEAMing approach to detect mutational frequencies as low as 0.01% in cfDNA from patients with colorectal cancer, and those with MRD detected generally had disease relapse within 1 year of localized surgery. More recently, Tie and colleagues[151] reported the results of a prospective trial evaluating the relationship of postoperative ctDNA levels with tumour recurrence in patients with stage II colorectal cancer; the recurrence rate was >10-fold higher in patients with detectable postoperative ctDNA than that of patients in whom ctDNA was undetectable, suggesting that the presence of ctDNA could be a predictive biomarker. As for the early diagnosis of cancer, whether identification of MRD-positive patients at increased risk of relapse can improve patient outcomes through early diagnosis of relapse and proactive 'consolidative' or 'rescue' treatment remains to be clarified. Monitoring of MRD also raises the possibility of de-escalating treatment in MRD-negative patients.

Evaluating response and predicting resistance to therapy. Similarly to monitoring MRD and thus relapse risk after definitive treatments, liquid biopsies can be applied to the monitoring of response and/or resistance to systemic therapy. For example, Dawson and colleagues[130] exploited WGS and candidate-gene analysis of ctDNA from patients with metastatic breast cancer to detect mutations in *PIK3CA* and *TP53*, which had previously been identified in tumour tissue. Furthermore, the amount of mutant ctDNA was better correlated with responses to standard treatment compared with serum CA15-3 levels or CTC enumeration[130]. High ctDNA levels were also correlated with unfavourable overall survival, and, importantly, ctDNA assessment was able to provide the earliest measure of treatment response in 53% of the patients analysed using all the monitoring modalities tested (increased ctDNA levels were detectable on average 5 months before the detection of progressive disease using imaging)[130]. In addition, Schiavon and colleagues[152] analysed *ESR1* mutations in ctDNA to demonstrate the evolution of resistance during therapy in 171 patients with advanced-stage breast cancer. *ESR1* mutations were detected only in patients with oestrogen receptor (ER)-positive disease previously treated with aromatase inhibitors, and were associated with a substantially shorter PFS on subsequent therapy[152].

In a proof-of-concept study in patients with metastatic breast or ovarian cancer, Murtaza and co-workers[153] performed whole-exome sequencing of plasma ctDNA collected from patients before initiating treatment and at disease recurrence, enabling the identification of mutations that were associated with the emergence of drug resistance.

The utility of ctDNA analysis in monitoring therapeutic response has also been reported in patients with NSCLC[154]: a reduction in the levels of ctDNA harbouring *EGFR* mutations that reflect sensitivity to EGFR TKIs was observed in 96% of the patients after the first treatment cycle, which provided an early indication of treatment response; however, the *EGFR* T790M mutation, which is a 'gatekeeper' mutation associated with resistance to first-generation EGFR TKIs, was detected in ctDNA before clinical disease progression. Moveover, ctDNA sequencing led to the discovery of novel mutations responsible for resistance to third-generation EGFR TKIs, which also target mutant EGFRs harbouring the T790M mutation[154].

Similarly, liquid biopsies have been successfully applied to identify mechanisms of resistance to antibody-mediated EGFR blockade in patients with metastatic colorectal cancer. *KRAS* mutations could be detected in plasma at high levels in almost 40% of patients after 6 months of treatment with cetuximab or panitumumab, whereas tissue biopsy samples obtained before the initiation of treatment did not reveal any mutations in this gene[2,155]. Interestingly, the emergence of resistant *KRAS*-mutated clones could be detected up to 10 months before radiographic confirmation of disease progression[155]. Indeed, Diaz and co-workers[2] uncovered the presence of multiple *KRAS* mutations in the circulation of single patients with metastatic colorectal cancer receiving panitumumab treatment. Importantly, mathematical modelling applied to their ctDNA-analysis data indicated that colorectal cancers probably contain resistant *KRAS*-mutant cells before therapy[155], with subsequent selection of these resistant cells under therapeutic pressure, in a Darwinian manner. *MET* amplification has also been detected in the plasma ctDNA of patients with metastatic colorectal cancer and acquired resistance to EGFR blockade[156].

Mutations in the EGFR extracellular domain (ECD) that negate binding of cetuximab and panitumumab have been detected in patients with colorectal cancer who achieved a partial response or stable disease after treatment with these anti-EGFR antibodies[157]. Interestingly, the same patients then received treatment with MM-151, a novel mixture of oligoclonal anti-EGFR antibodies with affinity that is not affected by ECD mutations, which led to reductions in the levels of ctDNA containing *EGFR* ECD mutations[157]. Notably, the reduction in the allelic frequency of the *EGFR* p.G465E mutation observed in one patient anticipated a marked reduction in tumour volume that was measured about 5 weeks later using CT[157], thus confirming the sensitivity of the *EGFR*-ECD-mutant clones to MM-151. These findings illustrate the potential of ctDNA analyses for the monitoring of disease evolution to guide therapeutic decisions.

The detection of oncogenic gene rearrangements in ctDNA is another exciting area of research. For example, Russo *et al.*[158] longitudinally monitored plasma ctDNA to assess *LMNA–NTRK1* gene-fusion status in a patient with metastatic colorectal cancer during treatment with the TRK tyrosine-kinase receptor inhibitor entrectinib. In addition to monitoring response and resistance, this liquid biopsy approach enabled characterization of previously unknown mutations that confer resistance to entrectinib: two novel NTRK1 kinase domain mutations (G595R and G667C) that were undetectable in the ctDNA obtained before treatment became detectable as resistance developed[158].

Indeed, liquid biopsies can be used to monitor dynamic clonal evolution in response to selective pressures exerted by therapy (FIG. 3). For example, ctDNA has been analysed in order to genotype colorectal cancers and track clonal evolution at multiple points before, during, and after treatment with anti-EGFR antibodies[1]. This study revealed that the abundance of *KRAS*-mutant clones that often emerge during EGFR blockade declines upon withdrawal of anti-EGFR antibody therapy[1], indicating that clonal evolution continues beyond clinical progression. Of note, ctDNA profiles of individuals who benefit from multiple challenges with anti-EGFR antibodies exhibit pulsatile levels of mutant *KRAS*, providing a molecular explanation for the efficacy of rechallenge therapies based on EGFR blockade[1]. Beyond colorectal cancer, evidence of successful rechallenge strategies with targeted therapies can be found in patients with other tumour types, particularly melanoma or NSCLC[159–163], but the molecular evolution has not been documented using liquid biopsy approaches.

Importantly, ctDNA analyses have been instrumental in demonstrating that responses to targeted therapies can be driven by distinct resistance mechanisms arising within separate tumour lesions in the same patient[3,164]. This finding suggests that evaluations of tissue and liquid biopsy samples should be integrated with radiological imaging to monitor the effect of individual oncogenic alterations on lesion-specific treatment responses. Murtaza and colleagues[4] reported an extensive analysis of matched tumour biopsy specimens and plasma samples collected from a patient with ER+/HER2+ metastatic breast cancer first treated with tamoxifen and trastuzumab followed by lapatinib; tumour evolution was followed for more than 3 years. Remarkably, the findings of plasma ctDNA mutational analysis reflected the clonal hierarchy determined via sequencing of multiregional tumour tissues, and enabled tracking of varying treatment responses across different lesions[4]. The results of these studies further suggest that mutations that occurred early during tumour development (clonal or 'truncal' mutations) are ideal candidates for monitoring tumour burden using ctDNA, as they are present in essentially all cancer cells in a single patient. Subclonal or 'branch' mutations that arose later in the tumour's phylogeny might, however, have utility in informing patient stratification and, consequently, in guiding the choice of targeted treatment.

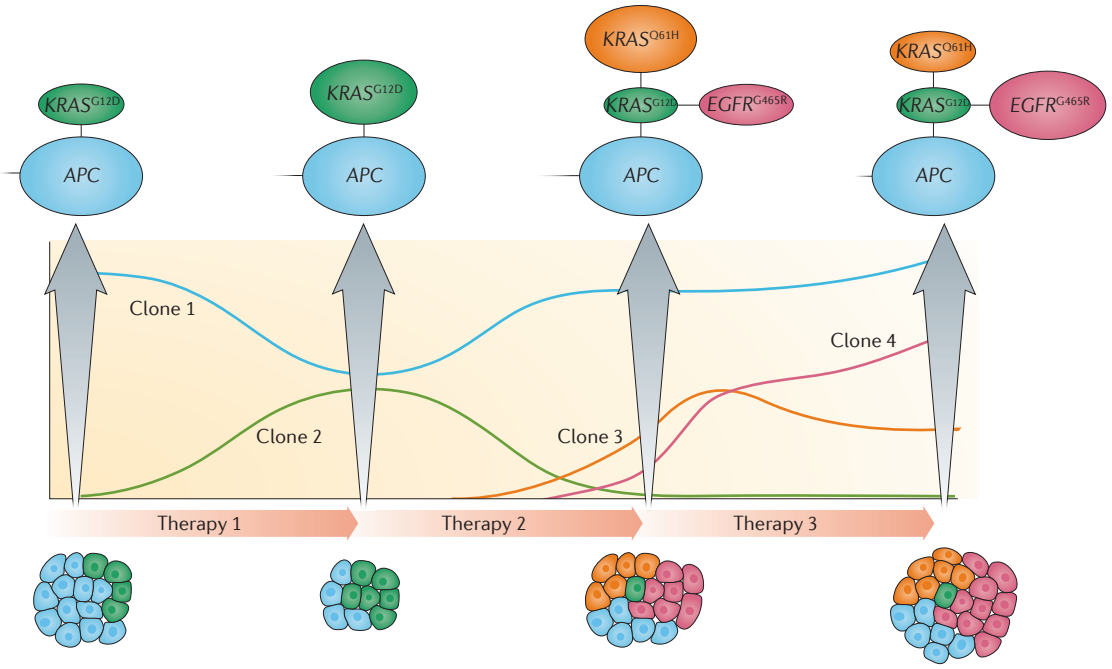


Figure 3 | **Monitoring clonal evolution using liquid biopsies.** Treatment with targeted therapies places tumour cells under selective pressure, thereby triggering clonal evolution that can be captured using liquid biopsy approaches. The data obtained can provide important insights into mechanisms of resistance, and can guide therapeutic decision-making. This schematic depicts the dynamic changes in the abundance of tumour-cell subclones harbouring different mutations in a patient with metastatic colorectal cancer treated with anti-EGFR antibodies. Monitoring of a clonal mutation in *APC* that is present in essentially all tumour cells (clone 1, blue line) tracks tumour burden, while subclonal mutations (*KRAS*^{G12D}, *KRAS*^{Q61H}, and *EGFR*^{G465R} in clones 2, 3 and 4, respectively) provide a measure of clonal evolution during therapy. Subclonal mutations can be lesion-specific, leading to variations in responses at different disease sites. In this hypothetical patient, primary treatment with the anti-EGFR antibody leads to a substantial decline in tumour burden through targeting of the majority of tumour cells, but outgrowth of the resistant *KRAS*^{G12D}-mutant subclone eventually causes tumour regrowth, necessitating a therapeutic switch. The change in therapy decreases the size of the *KRAS*^{G12D}-cell population, but resistant subclones (clones 3 and 4) with other mutations expand and drive tumour growth. A third-line of treatment restricts the growth of clone 3, but clone 4 continues to proliferate.

Liquid biopsy in the clinical setting

Liquid biopsy brings to the clinic the precious asset of convenience. Patient selection lies at the core of successful targeted therapy and precision medicine. At present, elucidation of molecular landscapes of tumour tissue is pivotal to guide treatment choices in clinical practice and in the context of clinical trials. Tissue biopsies are, however, fraught with issues. For example, accessibility is a problem when tumour lesions are in difficult anatomical locations, such as the mid-lung or retroperitoneum. Moreover, safety can be questionable, for instance, regarding sampling of tumours surrounding major vessels or in eloquent regions of the brain, or in patients with major comorbidities. In addition, assessments of tumour specimens can be restricted by time-related hindrances, including the reliability of molecular tests using years-old archival samples, and the unfeasibility of performing multiple longitudinal tests to follow tumour evolution and thereby expose mechanisms of secondary resistance to treatment. Evaluation of tumour-derived cfDNA in plasma, or other body fluids, clearly offers the potential to overcome these barriers. On the other hand, a limitation of liquid biopsy approaches is that transcriptome profiling and gene-expression analysis are not possible, as cfRNA is present in insufficient amounts in plasma and other body fluids[165]. CTCs could, however, be amenable to transcriptomic studies, especially in those tumour types for which gene-expression profiles have an established role in molecular diagnosis[166] and/or treatment choices for advanced-stage[167] or early stage disease settings[168], such as breast cancer. CTCs can also be exploited to understand the plasticity of tumour biology and guide treatment decisions based on the phenotypic status of the tumour cells, rather than their genotype, as indicated by the findings of Jordan and colleagues[169]. This group showed that CTCs from patients with originally ER+/HER– breast cancer can interconvert to a HER2+ phenotype under cytotoxic treatment, with activation of different signalling pathways in each population, without acquisition of additional genetic aberrations. Of note, HER– CTCs had activation of Notch and DNA-damage-repair pathways, therefore, displaying insensitivity to cytotoxic chemotherapy, but sensitivity to Notch-targeted drugs. By contrast, HER+ CTCs had reduced activation of these pathways, but were more proliferative. Thus, coexistance or interconversion between these states might enable tumour cells to overcome different stressors.

The integration of liquid biopsy assays into the management of patients with cancer is dependent on the accuracy of ctDNA sequencing in blood samples, defined as the positive predictive value (PPV) of plasma ctDNA levels versus 'gold standard' sequencing of matched tissue DNA. The clinical utility of liquid-biopsy-based techniques for treatment decision-making is well-exemplified by the determination of *EGFR* mutations in plasma ctDNA to guide the use of EGFR TKIs in patients with advanced-stage NSCLC — a clinical setting in which invasive diagnostic procedures are often deemed impossible. The first attempt at tissue–plasma pairwise comparison of *EGFR* mutations was reported a decade ago[170]. Since then, reports from numerous studies, including IGNYTE[171] and ASSESS[172] (two large, multinational, diagnostic, noncomparative intervention trials), and two meta-analyses[173,174], have confirmed that *EGFR*-sensitizing mutations are detectable in ctDNA with high specificity (>93%), but improvable sensitivity (<70%), making plasma genotyping of NSCLCs a viable alternative to tissue-based genotyping when the latter is not feasible[175].

Two randomized trials of first-generation EGFR TKIs have addressed the clinical utility of liquid biopsy and ctDNA analysis for *EGFR* genotyping. EURTAC[176] was the first trial to investigate ctDNA as a surrogate for *EGFR* testing of tissue biopsy samples and the effect of the liquid biopsy approach on outcomes. The trial randomly assigned 173 patients with activating *EGFR* mutations to receive erlotinib or a platinum-based chemotherapy doublet, with analysis of *EGFR* mutations in serum or plasma DNA included as a secondary end point[176]. The authors developed a multiplex real-time PCR assay for *EGFR* exon 19 deletions and L858R mutation, and were able to show that blood and tissue biomarkers had comparable predictive power[176]. Specifically, they found that *EGFR* L858R mutation, whether detected in tissue DNA or ctDNA, negatively affected median PFS and overall survival in erlotinib-treated patients compared with the outcomes of patients with tumours harbouring exon 19 deletions, as expected (L858R mutation is an established predictor of unfavourable outcome[177]). Mok and colleagues[178] used the cobas EGFR Blood Test to evaluate predictive biomarkers in the FASTACT-2 trial of induction chemotherapy plus either erlotinib or placebo, followed by erlotinib or placebo maintenance treatment. This test, at the time still in development, enabled the detection of G719A/C/S mutations in exons 18; 29 deletions in exon 19; S768I and T790M mutations and five insertions in exon 20; and L858R and L861Q mutations in exon 21 (REF. 179). In the planned retrospective analysis by Mok *et al.*[178], testing of samples from approximately 50% of the randomized patients revealed that baseline *EGFR*-mutation-positive patients who became *EGFR*-negative in plasma ctDNA at the end of the induction period had better outcomes, in terms of PFS and overall survival, than those who remained EGFR-mutation-positive.

Perhaps the most-valuable asset of a blood-based *EGFR*-mutation test is the capacity to monitor for resistance-associated mutations, such as the T790M gatekeeper mutation, because of their impact on ongoing treatment decision-making and patient survival. Importantly, the standard of care in this clinical space is rapidly changing owing to the availability of third-generation EGFR TKIs designed to have low activity against wild-type EGFR while retaining activity against EGFR isoforms with activating mutations as well as those harbouring some resistant mutations (including the T790M mutation)[180,181]. In a seminal study by Oxnard *et al.*[182], the investigators retrospectively established that patients with T790M-positive tumours treated with osimertinib, one such third-generation EGFR TKI that binds irreversibly to EGFRs harbouring common activating mutations and the T790M resistance mutation, had equivalent outcomes whether the mutation was detected in DNA from the tumour tissue or the blood. This finding is important because repeated tissue biopsy sampling is seldom feasible in patients with NSCLC. More importantly, the Oxnard group prospectively confirmed and expanded these findings in patients with newly diagnosed NSCLC and those with acquired resistance to EGFR-TKI therapy[183]. All patients underwent concomitant tissue and blood DNA genotyping using the cobas Blood Test, followed by longitudinal blood-based monitoring using the same test for *EGFR* exon 19 deletions, L858R and T790M mutations, and/or all *KRAS* codon 12 (G12X) alterations. The PPV of plasma ctDNA analysis was 100% for *EGFR* 19 deletions, *EGFR* L858R, and *KRAS* G12X, and 79% for T790M; sensitivity for the *EGFR* mutations was in the range of 74–88%, and was 64% for the *KRAS* mutations. Interestingly, sensitivity for both *EGFR* and *KRAS* was higher among patients with multiple metastatic sites and in those with hepatic or bone metastases than in patients with a single metastatic site outside of these organs. The assay also enabled detection of *EGFR* T790M mutations missed using tissue genotyping owing to tumour heterogeneity in patients with resistant disease[183]. The high specificity and a short median turnaround time of testing (12 days; range 1–54 days) make this assay a promising tool for guiding precision medicine[183].

Early primary tumour detection (diagnosis) might well be the holy grail of liquid biopsy, coupled with assays such as those used to detect *EGFR* mutations in NSCLC; however, this topic can only be addressed by very large trials in (apparently) healthy individuals, and is being pursued by several screening trials worldwide. In the USA and Korea, the umbrella LUNAR trial[184] is expected to enrol thousands of individuals to demonstrate, firstly, the feasibility and, secondly, the efficacy of early detection of breast, ovarian, lung, colorectal and pancreatic cancers. The LUNAR trial will integrate the use of cfDNA analyses, imaging, and germ line risk assessment, and has already collected samples from multiple trial sites, with reporting of pilot data expected late in 2016. In China, a cohort of 1,000 people with hepatitis B will undergo longitudinal liquid biopsy sampling to prove that aberrant shortening of cfDNA molecules can be used to indicate the presence of liver tumours before an ultrasonography-based clinical diagnosis[185].

On the other hand, the US FDA is investigating the manufacturers of the multiple-gene CancerIntercept Detect test, because the manufacturer was marketing the

test directly to consumers (at a cost of US$699); the FDA declared that the test constitutes a medical device that is not approved for direct marketing nor has it been adequately clinically validated within a clinical trial[186]. Indeed, liquid biopsy technologies are considered by regulatory agencies, such as the FDA and the European Medicine Agency (EMA), as diagnostic *in vitro* devices and are, therefore, subject to approval based on safety and effectiveness criteria. To accelerate development, adoption, and reimbursement of liquid biopsy assays of blood (and other bodily fluids), these generic rules must be translated into standards for clinical concordance, validation, and utility studies (BOX 1). The FDA have published a draft guidance document to facilitate this process[187]. The position of the FDA regarding liquid biopsy technologies is crucial in view of the fact that Illumina, a prominant manufacturer of NGS apparatus, and other large companies, such as Johnson & Johnson, Qiagen, and Roche, as well as a plethora of at least 30 other new companies, are steadfastly developing diagnostic blood tests for multiple cancer types. In February 2016, the business consulting firm RNCOS estimated that the global liquid biopsy industry "could cross the US$1 billion mark by 2020" (REF. 188).

To date, however, liquid biopsy has entered clinical practice only for the management of NSCLC. In January 2015, the EMA granted *In Vitro* Diagnostic Medical Device (IVD) marketing approval in Europe to the therascreen EGFR RGQ PCR Kit, and the following year, the FDA approved the cobas EGFR Mutation Test v2 (an updated version of the prior cobas EGFR Mutation Test) for use in the USA. Both of these IVDs can detect *EGFR* mutations in plasma ctDNA with comparable accuracy to that of bidirectional Sanger sequencing of DNA from tumour tissue specimens, as established within the framework of large clinical trials of small-molecule EGFR TKIs in patients with *EGFR*-mutated NSCLC.

The therascreen assay enables the detection of 19 different exon 19 deletions and three distinct exon 20 insertions (although it cannot be used to distinguish which specific deletion or insertion is present), as well as L858R, L861Q, G719X, S768I, and T790M mutations in *EGFR*. In the Lung-LUX3 trial[189], in which the efficacy of frontline chemotherapy was compared with that of afatinib (an irreversible second-generation EGFR TKI), the overall percentage agreement (OPA) between the therascreen test and the tissue-based test results was 92.2% (95% CI 89.0–94.8%); for 27 of 28 discordant sample pairs (96%), the presence of an *EGFR* mutation was detected in plasma ctDNA, but not in DNA from tumour tissue samples[190]. The clinical utility of the therascreen test was confirmed in a phase IV trial of gefitinib[191], in which 12 of 201 patients with no tumour tissue available for genotyping were found to harbour *EGFR* mutations using this liquid biopsy assay. On the other hand, the test had a 34% false-negative rate (36 out of 105 patients tested postive for *EGFR*-mutations in tissue DNA, but had no *EGFR* mutations detected in plasma cfDNA)[191]. Understandably, therefore, the EMA has amended the drug label for gefitinib to state that the detection of *EGFR* mutations in ctDNA should only be attempted for patients without an evaluable tumour sample[192].

The cobas EGFR Mutation Test v2 enables identification of 42 *EGFR* mutations, including exon 19 deletions, exon 20 insertions, and S768I, L861Q, L585R and T790M mutations, in both tissue and plasma[193,194]. The plasma test, however, is FDA-approved as a companion diagnostic for determining the eligibility of patients with NSCLC for erlotinib treatment based only on the presence of *EGFR* exon 19 deletions and the L858R substitution mutation; patients who test negative for these alterations must undergo routine biopsy sampling and testing for *EGFR* mutations in formalin-fixed paraffin-embedded tissue samples[194]. The approval of this test was based on the results of the phase III ENSURE trial (NCT01342965) conducted in China and Southern Asia to evaluate the efficacy and safety of erlotinib versus that of chemotherapy as first-line treatment in 217 patients with stage IIIB/IV NSCLC. Plasma samples were available for 517 of 610 (86.0%) of the patients screened for inclusion, and 214 of the 217 patients (98.6%) ultimately enrolled in this study[194]. Plasma samples were negative for *EGFR* exon 19 deletions and L585R mutation in 98.2% (95% CI 95.4–99.3%) of tissue-negative patients; however, only 76.7% (95% CI 70.5–81.9%) of the patients who tested positive for *EGFR* exon 19 deletions and the L858R mutation in tissue also had these alterations detected in plasma samples[193,194]. This relatively low positive percentage agreement (76.7%) between the tissue and plasma test results is what prompted the FDA to specify that patients without detectable mutations in plasma should undergo retesting using tumour tissue (if available)[193,194]. The efficacy of erlotinb based on the detection of *EGFR* mutations in plasma using the cobas EGFR Mutation Test v2 was also evaluated by 'bridging' to the drug efficacy results of the same trial based on tissue assessment using the cobas EGFR Mutation Test v1 (REF. 194); among the patients who tested positive for exon 19 deletion and/or an L858R mutations in plasma, those treated with erlotinib had improved progression-free survival (PFS) compared with those who received chemotherapy[194]. The cobas EGFR Mutation Test v2 has also been approved by the FDA for the detection of T790M mutation in tissue samples[193], but not yet in plasma samples. Interestingly, the performance of the cobas EGFR Mutation Test v2 in detecting the T790M resistance mutation in ctDNA has been evaluated in a pooled retrospective analysis of two single-arm phase II registration studies of osimertinib (AURA extension, NCT01802632; AURA2, NCT02094261)[195]; the positive percentage agreement and negative percentage agreement between the findings of tissue and plasma testings was 61.4% and 78.6%, respectively[195], which is not dissimilar to the results obtained for detection of exon 19 deletions and L585R mutations in ctDNA using the same test[193,194].

Notably, in a study conducted at the Dana-Farber Cancer Institute (Boston, Massachusetts, USA), the clinical utility of a similar approach to liquid biopsy analysis of *EGFR* mutations (as well as *KRAS* mutation), at diagnosis or relapse, was prospectively validated in 180 patients with advanced-stage NSCLC. Mutations were detected using a digital PCR platform with high specificity (>79%) and with a median turnaround time of only 3 days, compared

with 12 days and 27 days for tissue genotyping of newly diagnosed and relapsed tumours, respectively[196]. Outside of NSCLC, eligibility for treatment with targeted therapies is increasingly based on molecular inclusion and/or exclusion criteria for many cancer types, including melanoma, and colorectal and breast cancers. At present, most decisions are informed by data from tumour tissue specimens, but in the future, such assessments could potentially be replaced with liquid biopsy approaches, which might be of particular importance when surgery is not indicated and tissue biopsy samples are difficult to obtain, or when genetic heterogeneity could compound decision-making based on limited tissue specimens. Extensive research is needed, however, to prove the utility and reliability of ctDNA analysis in such diverse settings.

Of note, the preliminary results from a large validation trial of Guardant360 (REF. 197), a digital DNA-sequencing technology test encompassing a broad panel of 70 genes, including all of the current clinically actionable genes with approved targeted drugs, have been presented at the 2016 ASCO Annual Meeting[198]. The study included 15,191 patients with advanced-stage lung, breast, colorectal, or other cancers[198]. Accuracy was assessed indirectly by comparing the frequencies of specific changes in ctDNA with tumour-tissue-based data from patients included in The Cancer Genome Atlas (TCGA), and directly for a subset of almost 400 patients — mostly with NSCLC or colorectal cancer — using matched plasma and tissue samples. Correlations between the TCGA and ctDNA data ranged from 92–99% across multiple cancer genes and different classes of alterations. The *EGFR* T790M resistance mutation, however, was only detected in plasma DNA from patients treated with EGFR TKIs — that is, it was not detectable in pretreatment samples. The matched plasma–tissue comparison showed a PPV of 87% that, importantly, increased to 98% when the plasma and tissue samples were collected <6 months apart. The Guardant360 test is quite sensitive (mutated ctDNA accounting for 0.4% of the total cfDNA in the blood could be detected), and enabled the detection of resistance mutations in *EGFR*, *ALK*, and *KRAS* that were not detectable in the matched tumour biopsy samples in almost one-third of the patients. Overall, the test identified a molecular alteration in 64% of the patients, including 362 patients with NSCLC with no available tumour tissue, which could potentially be targeted using an FDA-approved drug or an experimental drug currently being tested in a clinical trial[198]. In several instances, mutations were identified at progression after an intial response to a targeted agent that can be successfully inhibited with additional lines of therapy, or lead to therapeutic switching, emphasizing the potential clinical significance of repeated assessment of ctDNA for targetable mutations throughout the disease course. For example, *EGFR* mutations that emerge during the treatment of patients with NSCLC with erlotinib or genfitib can be successfully targeted with rociletinib; similarly, second-generation inhibitors are available for *ALK* mutations that emerge after initial ALK blockade. The same principles apply to other mutations associated with a range of solid and haematological cancers.

Conclusions

Applications of liquid biopsies in oncology have emerged and developed at an incredible rate over the past 5 years. Many clinical trials performed in patients with solid tumours now incorporate longitudinal blood collections; however, most of the studies have included only small cohorts of patients. Implementation of liquid biopsy approaches in clinical practice will occur only after extensive controlled studies are performed. Currently, more than 60 trials (Supplementary information S1 (table)), with a projected accrual of more than 20,000 patients across 11 cancer types, are addressing the challenges posed by liquid biopsy. Limitations to be tackled include: the standardization of the blood collection procedure to improve the stability of samples at room temperature, thus reducing preanalytical variability[12]; defining ctDNA quantification methods[199]; standardization of ctDNA isolation to improve yield; and improving the sensitivity of ctDNA detection for rare molecular alterations in order to anticipate drug resistance (BOX 1).

Exploiting liquid biopsy approaches in patient screening could provide a more comprehensive view of tumour characteristics, including aggressiveness and the overall molecular landscape. Comprehensive implementation in the clinical setting will probably include the initial analysis of a subset of candidate genes for known hotspot mutations using PCR-based approaches, as a benchmark test and to confirm the presence of sufficient amounts and quality of ctDNA, followed by broader ctDNA sequencing to uncover actionable therapeutic targets.

In conclusion, the next generation of 'liquid biopsy' studies will be key to definitively establishing the clinical applicability of blood-based genomic profiling. Liquid biopsy approaches will probably provide improved diagnostic power, but a key question remains: will liquid-biopsy-driven treatments translate into improved outcomes?

1. Siravegna, G. *et al.* Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. *Nat. Med.* **21**, 795–801 (2015).
2. Diaz, L. A. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
3. Russo, M. *et al.* Tumor heterogeneity and lesion-specific response to targeted therapy in colorectal cancer. *Cancer Discov.* **6**, 147–153 (2015).
4. Murtaza, M. *et al.* Multifocal clonal evolution characterized using circulating tumour DNA in a case of metastatic breast cancer. *Nat. Commun.* **6**, 8760 (2015).
5. Morelli, M. P. *et al.* Characterizing the patterns of clonal selection in circulating tumor DNA from patients with colorectal cancer refractory to anti-EGFR treatment. *Ann. Oncol.* **26**, 731–736 (2015).
6. Alix-Panabières, C., Schwarzenbach, H. & Pantel, K. Circulating tumor cells and circulating tumor DNA. *Annu. Rev. Med.* **63**, 199–215 (2012).
7. Reckamp, K. L. *et al.* A highly sensitive and quantitative test platform for detection of NSCLC EGFR mutations in urine and plasma. *J. Thorac. Oncol.* **11**, 1690–1700 (2016).
8. Wang, Y. *et al.* Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas. *Sci. Transl Med.* **7**, 293ra104 (2015).
9. Kimura, H. *et al.* EGFR mutation status in tumour-derived DNA from pleural effusion fluid is a practical basis for predicting the response to gefitinib. *Br. J. Cancer* **95**, 1390–1395 (2006).
10. De Mattos-Arruda, L. *et al.* Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat. Commun.* **6**, 8839 (2015).
11. Diehl, F. *et al.* Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. *Gastroenterology* **135**, 489–498 (2008).
12. El Messaoudi, S., Rolet, F., Mouliere, F. & Thierry, A. R. Circulating cell free DNA: preanalytical considerations. *Clin. Chim. Acta* **424**, 222–230 (2013).
13. Yu, M., Stott, S., Toner, M., Maheswaran, S. & Haber, D. A. Circulating tumor cells: approaches to isolation and characterization. *J. Cell Biol.* **192**, 373–382 (2011).

14. Diaz, L. A. & Bardelli, A. Liquid biopsies: genotyping circulating tumor DNA. *J. Clin. Oncol.* **32**, 579–586 (2014).
15. Ashworth, T. R. A case of cancer in which cells similar to those in the tumours where seen in the blood after death. *Australian Med. J.***14**, 146–147 (1869).
16. Krebs, M. G., Hou, J. M., Ward, T. H., Blackhall, F. H. & Dive, C. Circulating tumour cells: their utility in cancer management and predicting outcomes. *Ther. Adv. Med. Oncol.* **2**, 351–365 (2010).
17. Krebs, M. G. *et al.* Molecular analysis of circulating tumour cells — biology and biomarkers. *Nat. Rev. Clin. Oncol.* **11**, 129–144 (2014).
18. Haber, D. A. & Velculescu, V. E. Blood-based analyses of cancer: circulating tumor cells and circulating tumor DNA. *Cancer Discov.* **4**, 650–661 (2014).
19. Lianidou, E. S. & Markou, A. Circulating tumor cells in breast cancer: detection systems, molecular characterization, and future challenges. *Clin. Chem.* **57**, 1242–1255 (2011).
20. Alix-Panabières, C. & Pantel, K. Circulating tumor cells: liquid biopsy of cancer. *Clin. Chem.* **59**, 110–118 (2013).
21. Pantel, K., Brakenhoff, R. H. & Brandt, B. Detection, clinical relevance and specific biological properties of disseminating tumour cells. *Nat. Rev. Cancer* **8**, 329–340 (2008).
22. Joosse, S. A. & Pantel, K. Biologic challenges in the detection of circulating tumor cells. *Cancer Res.* **73**, 8–11 (2013).
23. Alix-Panabières, C. & Pantel, K. Challenges in circulating tumour cell research. *Nat. Rev. Cancer* **14**, 623–631 (2014).
24. Lin, H. K. *et al.* Portable filter-based microdevice for detection and characterization of circulating tumor cells. *Clin. Cancer Res.* **16**, 5011–5018 (2010).
25. Paterlini-Brechot, P. & Benali, N. L. Circulating tumor cells (CTC) detection: clinical impact and future directions. *Cancer Lett.* **253**, 180–204 (2007).
26. Zhang, L. *et al.* The identification and characterization of breast cancer CTCs competent for brain metastasis. *Sci. Transl Med.* **5**, 180ra148 (2013).
27. Ameri, K. *et al.* Circulating tumour cells demonstrate an altered response to hypoxia and an aggressive phenotype. *Br. J. Cancer* **102**, 561–569 (2010).
28. Yu, M. *et al.* Cancer therapy. *Ex vivo* culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science* **345**, 216–220 (2014).
29. Cayrefourcq, L. *et al.* Establishment and characterization of a cell line from human circulating colon cancer cells. *Cancer Res.* **75**, 892–901 (2015).
30. Baccelli, I. *et al.* Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat. Biotechnol.* **31**, 539–544 (2013).
31. Rossi, E. *et al.* Retaining the long-survive capacity of circulating tumor cells (CTCs) followed by xeno-transplantation: not only from metastatic cancer ofthe breast but also of prostate cancer patients. *Oncoscience* **1**, 49–56 (2014).
32. Hodgkinson, C. L. *et al.* Tumorigenicity and genetic profiling of circulating tumor cells in small-cell lung cancer. *Nat. Med.* **20**, 897–903 (2014).
33. Gao, D. *et al.* Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159**, 176–187 (2014).
34. Khoo, B. L. *et al.* Liquid biopsy and therapeutic response: circulating tumor cell cultures for evaluation of anticancer treatment. *Sci. Adv.* **2**, e1600274 (2016).
35. Cocucci, E., Racchetti, G. & Meldolesi, J. Shedding microvesicles: artefacts no more. *Trends Cell Biol.* **19**, 43–51 (2009).
36. Heijnen, H. F., Schiel, A. E., Fijnheer, R., Geuze, H. J. & Sixma, J. J. Activated platelets release two types of membrane vesicles: microvesicles by surface shedding and exosomes derived from exocytosis of multivesicular bodies and α-granules. *Blood* **94**, 3791–3799 (1999).
37. Pan, B. T. & Johnstone, R. M. Fate of the transferrin receptor during maturation of sheep reticulocytes *in vitro*: selective externalization of the receptor. *Cell* **33**, 967–978 (1983).
38. Simons, M. & Raposo, G. Exosomes — vesicular carriers for intercellular communication. *Curr. Opin. Cell Biol.* **21**, 575–581 (2009).
39. Mathivanan, S., Ji, H. & Simpson, R. J. Exosomes: extracellular organelles important in intercellular communication. *J. Proteomics* **73**, 1907–1920 (2010).
40. Liao, J., Liu, R., Yin, L. & Pu, Y. Expression profiling of exosomal miRNAs derived from human esophageal cancer cells by Solexa high-throughput sequencing. *Int. J. Mol. Sci.* **15**, 15530–15551 (2014).
41. Wieckowski, E. & Whiteside, T. L. Human tumor-derived versus dendritic cell-derived exosomes have distinct biologic roles and molecular profiles. *Immunol. Res.* **36**, 247–254 (2006).
42. Lobb, R. J. *et al.* Optimized exosome isolation protocol for cell culture supernatant and human plasma. *J. Extracell. Vesicles* **4**, 27031 (2015).
43. Zhang, J. *et al.* Exosome and exosomal microRNA: trafficking, sorting, and function. *Genomics Proteomics Bioinformatics* **13**, 17–24 (2015).
44. Rani, S. *et al.* Isolation of exosomes for subsequent mRNA, microRNA, and protein profiling. *Methods Mol. Biol.* **784**, 181–195 (2011).
45. Stevens, G. L., Scheer, W. D. & Levine, E. A. Detection of tyrosinase mRNA from the blood of melanoma patients. *Cancer Epidemiol. Biomarkers Prev.* **5**, 293–296 (1996).
46. Taylor, D. D. & Gercel-Taylor, C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol. Oncol.* **110**, 13–21 (2008).
47. Vickers, K. C., Palmisano, B. T., Shoucri, B. M., Shamburek, R. D. & Remaley, A. T. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell Biol.* **13**, 423–433 (2011).
48. Joosse, S. A. & Pantel, K. Tumor-educated platelets as liquid biopsy in cancer patients. *Cancer Cell* **28**, 552–554 (2015).
49. Mitchell, P. S. *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl Acad. Sci. USA* **105**, 10513–10518 (2008).
50. Skog, J. *et al.* Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat. Cell Biol.* **10**, 1470–1476 (2008).
51. Rabinowits, G., Gerçel-Taylor, C., Day, J. M., Taylor, D. D. & Kloecker, G. H. Exosomal microRNA: a diagnostic marker for lung cancer. *Clin. Lung Cancer* **10**, 42–46 (2009).
52. Garcĩa-Olmo, D. C., Picazo, M. G., Toboso, I., Asensio, A. I. & Garcĩa-Olmo, D. Quantitation of cell-free DNA and RNA in plasma during tumor progression in rats. *Mol. Cancer* **12**, 8 (2013).
53. Ono, S., Lam, S., Nagahara, M. & Hoon, D. S. Circulating microRNA biomarkers as liquid biopsy for cancer patients: pros and cons of current assays. *J. Clin. Med.* **4**, 1890–1907 (2015).
54. Schwarzenbach, H., Nishida, N., Calin, G. A. & Pantel, K. Clinical relevance of circulating cell-free microRNAs in cancer. *Nat. Rev. Clin. Oncol.* **11**, 145–156 (2014).
55. Wren Laboratories, LLC. NETest Supporting Data. *Wrenlaboratories.com* http://www.wrenlaboratories.com/provider/netest-supporting-data/ (2017).
56. Mandel, P. & Metais, P. Les acides nucléiques du plasma sanguin chez l'homme [French]. *C. R. Seances Soc. Biol. Fil.* **142**, 241–243 (1948).
57. Swarup, V. & Rajeswari, M. R. Circulating (cell-free) nucleic acids — a promising, non-invasive tool for early detection of several human diseases. *FEBS Lett.* **581**, 795–799 (2007).
58. Leon, S. A., Shapiro, B., Sklaroff, D. M. & Yaros, M. J. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res.* **37**, 646–650 (1977).
59. Stroun, M. *et al.* Neoplastic characteristics of the DNA found in the plasma of cancer patients. *Oncology* **46**, 318–322 (1989).
60. Wang, J. Y. *et al.* Molecular detection of *APC*, *K-ras*, and *p53* mutations in the serum of colorectal cancer patients as circulating biomarkers. *World J. Surg.* **28**, 721–726 (2004).
61. Shaw, J. A. *et al.* Microsatellite alterations plasma DNA of primary breast cancer patients. *Clin. Cancer Res.* **6**, 1119–1124 (2000).
62. Fujiwara, K. *et al.* Identification of epigenetic aberrant promoter methylation in serum DNA is useful for early detection of lung cancer. *Clin. Cancer Res.* **11**, 1219–1225 (2005).
63. Jahr, S. *et al.* DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61**, 1659–1665 (2001).
64. Diehl, F. *et al.* Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl Acad. Sci. USA* **102**, 16368–16373 (2005).
65. Lo, Y. M. *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci. Transl Med.* **2**, 61ra91 (2010).
66. Heitzer, E. *et al.* Establishment of tumor-specific copy number alterations from plasma DNA of patients with cancer. *Int. J. Cancer* **133**, 346–356 (2013).
67. Lo, Y. M. *et al.* Rapid clearance of fetal DNA from maternal plasma. *Am. J. Hum. Genet.* **64**, 218–224 (1999).
68. Jung, M., Klotzek, S., Lewandowski, M., Fleischhacker, M. & Jung, K. Changes in concentration of DNA in serum and plasma during storage of blood samples. *Clin. Chem.* **49**, 1028–1029 (2003).
69. Millholland, J. M., Li, S., Fernandez, C. A. & Shuber, A. P. Detection of low frequency *FGFR3* mutations in the urine of bladder cancer patients using next-generation deep sequencing. *Res. Rep. Urol.* **4**, 33–40 (2012).
70. Li, Y., Zhou, X., St John, M. A. & Wong, D. T. RNA profiling of cell-free saliva using microarray technology. *J. Dent. Res.* **83**, 199–203 (2004).
71. Pan, W., Gu, W., Nagpal, S., Gephart, M. H. & Quake, S. R. Brain tumor mutations detected in cerebral spinal fluid. *Clin. Chem.* **61**, 514–522 (2015).
72. De Mattos-Arruda, L. *et al.* Capturing intra-tumor genetic heterogeneity by *de novo* mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Ann. Oncol.* **25**, 1729–1735 (2014).
73. Soh, J. *et al.* Usefulness of *EGFR* mutation screening in pleural fluid to predict the clinical outcome of gefitinib treated patients with lung cancer. *Int. J. Cancer* **119**, 2353–2358 (2006).
74. Kawahara, A. *et al.* Epidermal growth factor receptor mutation status in cell-free DNA supernatant of bronchial washings and brushings. *Cancer Cytopathol.* **123**, 620–628 (2015).
75. Wang, Y. *et al.* Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. *Proc. Natl Acad. Sci. USA* **112**, 9704–9709 (2015).
76. Zhang, J. *et al.* Presence of donor- and recipient-derived DNA in cell-free urine samples of renal transplantation recipients: urinary DNA chimerism. *Clin. Chem.* **45**, 1741–1746 (1999).
77. Botezatu, I. *et al.* Genetic analysis of DNA excreted in urine: a new approach for detecting specific genomic DNA sequences from cells dying in an organism. *Clin. Chem.* **46**, 1078–1084 (2000).
78. Simkin, M., Abdalla, M., El-Mogy, M. & Haj-Ahmad, Y. Differences in the quantity of DNA found in the urine and saliva of smokers versus nonsmokers: implications for the timing of epigenetic events. *Epigenomics* **4**, 343–352 (2012).
79. Hoque, M. O. *et al.* Quantitative detection of promoter hypermethylation of multiple genes in the tumor, urine, and serum DNA of patients with renal cancer. *Cancer Res.* **64**, 5511–5517 (2004).
80. Bryzgunova, O. E. *et al.* Isolation and comparative study of cell-free nucleic acids from human urine. *Ann. N. Y. Acad. Sci.* **1075**, 334–340 (2006).
81. Su, Y. H. *et al.* Human urine contains small, 150 to 250 nucleotide-sized, soluble DNA derived from the circulation and may be useful in the detection of colorectal cancer. *J. Mol. Diagn.* **6**, 101–107 (2004).
82. Melkonyan, H. S. *et al.* Transrenal nucleic acids: from proof of principle to clinical tests. *Ann. N. Y. Acad. Sci.* **1137**, 73–81 (2008).
83. Tsui, N. B. *et al.* High resolution size analysis of fetal DNA in the urine of pregnant women by paired-end massively parallel sequencing. *PLoS ONE* **7**, e48319 (2012).
84. Nadano, D., Yasuda, T. & Kishi, K. Measurement of deoxyribonuclease I activity in human tissues and body fluids by a single radial enzyme-diffusion method. *Clin. Chem.* **39**, 448–452 (1993).
85. Mall, C., Rocke, D. M., Durbin-Johnson, B. & Weiss, R. H. Stability of miRNA in human urine supports its biomarker potential. *Biomark. Med.* **7**, 623–631 (2013).
86. Li, M. *et al.* Analysis of the RNA content of the exosomes derived from blood serum and urine and itspotential as biomarkers. *Phil. Trans. R. Soc. B* http://dx.doi.org/10.1098/rstb.2013.0502 (2014).
87. Turchinovich, A., Weiz, L., Langheinz, A. & Burwinkel, B. Characterization of extracellular circulating microRNA. *Nucleic Acids Res.* **39**, 7223–7233 (2011).
88. Eisenberger, C. F. *et al.* Diagnosis of renal cancer by molecular urinalysis. *J. Natl Cancer Inst.* **91**, 2028–2032 (1999).
89. Goessl, C., Müller, M., Straub, B. & Miller, K. DNA alterations in body fluids as molecular tumor markers for urological malignancies. *Eur. Urol.* **41**, 668–676 (2002).

90. Damkier, H. H., Brown, P. D. & Praetorius, J. Cerebrospinal fluid secretion by the choroid plexus. *Physiol. Rev.* **93**, 1847–1892 (2013).
91. Segal, M. B. Extracellular and cerebrospinal fluids. *J. Inherit. Metab. Dis.* **16**, 617–638 (1993).
92. Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl Med.* **6**, 224ra24 (2014).
93. von Hoff, K. & Rutkowski, S. Medulloblastoma. *Curr. Treat. Options Neurol.* **14**, 416–426 (2012).
94. Chamberlain, M. C., Kormanik, P. A. & Glantz, M. J. A comparison between ventricular and lumbar cerebrospinal fluid cytology in adult patients with leptomeningeal metastases. *Neuro Oncol.* **3**, 42–45 (2001).
95. Bougel, S. *et al.* Methylation of the *hTERT* promoter: a novel cancer biomarker for leptomeningeal metastasis detection in cerebrospinal fluids. *Clin. Cancer Res.* **19**, 2216–2223 (2013).
96. Samuel, N., Remke, M., Rutka, J. T., Raught, B. & Malkin, D. Proteomic analyses of CSF aimed at biomarker development for pediatric brain tumors. *J. Neurooncol.* **118**, 225–238 (2014).
97. Baraniskin, A. *et al.* Identification of microRNAs in the cerebrospinal fluid as marker for primary diffuse large B-cell lymphoma of the central nervous system. *Blood* **117**, 3140–3146 (2011).
98. Saadatpour, L. *et al.* Glioblastoma: exosome and microRNA as novel diagnosis biomarkers. *Cancer Gene Ther.* **23**, 415–418 (2016).
99. Akers, J. C. *et al.* miRNA contents of cerebrospinal fluid extracellular vesicles in glioblastoma patients. *J. Neurooncol.* **123**, 205–216 (2015).
100. Gallo, A. & Alevizos, I. Isolation of circulating microRNA in saliva. *Methods Mol. Biol.* **1024**, 183–190 (2013).
101. Gallo, A., Tandon, M., Alevizos, I. & Illei, G. G. The majority of microRNAs detectable in serum and saliva is concentrated in exosomes. *PLoS ONE* **7**, e30679 (2012).
102. Han, H. S. *et al.* Downregulation of cell-free miR-198 as a diagnostic biomarker for lung adenocarcinoma-associated malignant pleural effusion. *Int. J. Cancer* **133**, 645–652 (2013).
103. Wang, T. *et al.* Cell-free microRNA expression profiles in malignant effusion associated with patient survival in non-small cell lung cancer. *PLoS ONE* **7**, e43268 (2012).
104. Oh, J. E. *et al.* Detection of low-level *KRAS* mutations using PNA-mediated asymmetric PCR clamping and melting curve analysis with unlabeled probes. *J. Mol. Diagn.* **12**, 418–424 (2010).
105. Reis-Filho, J. S. Next-generation sequencing. *Breast Cancer Res.* **11** (Suppl. 3), S12 (2009).
106. Cai, X., Janku, F., Zhan, Q. & Fan, J. B. Accessing genetic information with liquid biopsies. *Trends Genet.* **31**, 564–575 (2015).
107. Siravegna, G. & Bardelli, A. Genotyping cell-free tumor DNA in the blood to detect residual disease and drug resistance. *Genome Biol.* **15**, 449 (2014).
108. Nagaiah, G. & Abraham, J. Circulating tumor cells in the management of breast cancer. *Clin. Breast Cancer* **10**, 209–216 (2010).
109. Toss, A., Mu, Z., Fernandez, S. & Cristofanilli, M. CTC enumeration and characterization: moving toward personalized medicine. *Ann. Transl Med.* **2**, 108 (2014).
110. Cristofanilli, M. *et al.* Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *N. Engl. J. Med.* **351**, 781–791 (2004).
111. Cohen, S. J. *et al.* Relationship of circulating tumor cellsto tumor response, progression-free survival, and overall survival in patients with metastatic colorectal cancer. *J. Clin. Oncol.* **26**, 3213–3221 (2008).
112. de Bono, J. S. *et al.* Circulating tumor cells predict survival benefit from treatment in metastatic castration-resistant prostate cancer. *Clin. Cancer Res.* **14**, 6302–6309 (2008).
113. Allard, W. J. *et al.* Tumor cells circulate in the peripheral blood of all major carcinomas but not in healthy subjects or patients with nonmalignant diseases. *Clin. Cancer Res.* **10**, 6897–6904 (2004).
114. Hayes, D. F. *et al.* Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival. *Clin. Cancer Res.* **12**, 4218–4224 (2006).
115. Janni, W. J. *et al.* Pooled analysis of the prognostic relevance of circulating tumor cells in primary breast cancer. *Clin. Cancer Res.* **22**, 2583–2593 (2016).
116. Yu, M. *et al.* Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**, 580–584 (2013).
117. Stott, S. L. *et al.* Isolation and characterization of circulating tumor cells from patients with localized and metastatic prostate cancer. *Sci. Transl Med.* **2**, 25ra23 (2010).
118. Powell, A. A. *et al.* Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines. *PLoS ONE* **7**, e33788 (2012).
119. Pestrin, M. *et al.* Correlation of HER2 status between primary tumors and corresponding circulating tumor cells in advanced breast cancer patients. *Breast Cancer Res. Treat.* **118**, 523–530 (2009).
120. Butt, A. Q. & Mills, K. H. Immunosuppressive networks and checkpoints controlling antitumor immunity and their blockade in the development of cancer immunotherapeutics and vaccines. *Oncogene* **33**, 4623–4631 (2014).
121. Mazel, M. *et al.* Frequent expression of PD-L1 on circulating breast cancer cells. *Mol. Oncol.* **9**, 1773–1782 (2015).
122. Sozzi, G. *et al.* Quantification of free circulating DNA as a diagnostic marker in lung cancer. *J. Clin. Oncol.* **21**, 3902–3908 (2003).
123. Kim, K. *et al.* Circulating cell-free DNA as a promising biomarker in patients with gastric cancer: diagnostic validity and significant reduction of cfDNA after surgical resection. *Ann. Surg. Treat. Res.* **86**, 136–142 (2014).
124. Frattini, M. *et al.* Quantitative and qualitative characterization of plasma DNA identifies primary and recurrent colorectal cancer. *Cancer Lett.* **263**, 170–181 (2008).
125. Chen, X. *et al.* Detecting tumor-related alterations in plasma or serum DNA of patients diagnosed with breast cancer. *Clin. Cancer Res.* **5**, 2297–2303 (1999).
126. Chan, K. C. *et al.* Early detection of nasopharyngeal carcinoma by plasma Epstein-Barr virus DNA analysis in a surveillance program. *Cancer* **119**, 1838–1844 (2013).
127. Gormally, E. *et al. TP53* and *KRAS2* mutations in plasma DNA of healthy subjects and subsequent cancer occurrence: a prospective study. *Cancer Res.* **66**, 6871–6876 (2006).
128. Welch, H. G. & Black, W. C. Overdiagnosis in cancer. *J. Natl Cancer Inst.* **102**, 605–613 (2010).
129. Higgins, M. J. *et al.* Detection of tumor *PIK3CA* status in metastatic breast cancer using peripheral blood. *Clin. Cancer Res.* **18**, 3462–3469 (2012).
130. Dawson, S. J. *et al.* Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368**, 1199–1209 (2013).
131. Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990 (2008).
132. Misale, S. *et al.* Emergence of *KRAS* mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532–536 (2012).
133. Thierry, A. R. *et al.* Clinical validation of the detection of *KRAS* and *BRAF* mutations from circulating tumor DNA. *Nat. Med.* **20**, 430–435 (2014).
134. Newman, A. M. *et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20**, 548–554 (2014).
135. Narayan, A. *et al.* Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing. *Cancer Res.* **72**, 3492–3498 (2012).
136. Takai, E. *et al.* Clinical utility of circulating tumor DNA for molecular assessment in pancreatic cancer. *Sci. Rep.* **5**, 18425 (2015).
137. Liggett, T. *et al.* Differential methylation of cell-free circulating DNA among patients with pancreatic cancer versus chronic pancreatitis. *Cancer* **116**, 1674–1680 (2010).
138. Sturgeon, S. R. *et al.* Detection of promoter methylation of tumor suppressor genes in serum DNA of breast cancer cases and benign breast disease controls. *Epigenetics* **7**, 1258–1267 (2012).
139. Ellinger, J. *et al.* CpG island hypermethylation in cell-free serum DNA identifies patients with localized prostate cancer. *Prostate* **68**, 42–49 (2008).
140. Ellinger, J. *et al.* CpG island hypermethylation of cell-free circulating serum DNA in patients with testicular cancer. *J. Urol.* **182**, 324–329 (2009).
141. Kadam, S. K., Farmen, M. & Brandt, J. T. Quantitative measurement of cell-free plasma DNA and applications for detecting tumor genetic variation and promoter methylation in a clinical setting. *J. Mol. Diagn.* **14**, 346–356 (2012).
142. Kristensen, L. S. & Hansen, L. L. PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin. Chem.* **55**, 1471–1483 (2009).
143. Chan, K. C. *et al.* Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc. Natl Acad. Sci. USA* **110**, 18761–18768 (2013).
144. Balaña, C. *et al.* O^6*-methyl-guanine-DNA methyltransferase methylation* in serum and tumor DNA predicts response to 1,3-bis(2-chloroethyl)-1-nitrosourea but not to temozolamide plus cisplatin in glioblastoma multiforme. *Clin. Cancer Res.* **9**, 1461–1468 (2003).
145. Barault, L. *et al.* Digital PCR quantification of *MGMT* methylation refines prediction of clinical benefit from alkylating agents in glioblastoma and metastatic colorectal cancer. *Ann. Oncol.* **26**, 1994–1999 (2015).
146. Li, M. *et al.* Sensitive digital quantification of DNA methylation in clinical samples. *Nat. Biotechnol.* **27**, 858–863 (2009).
147. Reinert, T. *et al.* Analysis of circulating tumour DNA to monitor disease burden following colorectal cancer surgery. *Gut* **65**, 625–634 (2015).
148. Beaver, J. A. *et al.* Detection of cancer DNA in plasma of early stage breast cancer patients. *Clin. Cancer Res.* **20**, 2643–2650 (2014).
149. Garcia-Murillas, I. *et al.* Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci. Transl Med.* **7**, 302ra133 (2015).
150. Roschewski, M. *et al.* Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. *Lancet Oncol.* **16**, 541–549 (2015).
151. Tie, J. *et al.* Circulating tumor DNA as an early marker of therapeutic response in patients with metastatic colorectal cancer. *Ann. Oncol.* **26**, 1715–1722 (2015).
152. Schiavon, G. *et al.* Analysis of *ESR1* mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Sci. Transl Med.* **7**, 313ra182 (2015).
153. Murtaza, M. *et al.* Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* **497**, 108–112 (2013).
154. Thress, K. S. *et al. EGFR* mutation detection in ctDNA from NSCLC patient plasma: a cross-platform comparison of leading technologies to support the clinical development of AZD9291. *Lung Cancer* **90**, 509–515 (2015).
155. Misale, S. *et al.* Blockade of EGFR and MEK intercepts heterogeneous mechanisms of acquired resistance to anti-EGFR therapies in colorectal cancer. *Sci. Transl Med.* **6**, 224ra26 (2014).
156. Bardelli, A. *et al.* Amplification of the MET receptor drives resistance to anti-EGFR therapies in colorectal cancer. *Cancer Discov.* **3**, 658–673 (2013).
157. Arena, S. *et al.* Emergence of multiple *EGFR* extracellular mutations during cetuximab treatment in colorectal cancer. *Clin. Cancer Res.* **21**, 2157–2166 (2015).
158. Russo, M. *et al.* Acquired resistance to the TRK inhibitor entrectinib in colorectal cancer. *Cancer Discov.* **6**, 36–44 (2016).
159. Hata, A., Katakami, N., Kaji, R., Fujita, S. & Imai, Y. Does T790M disappear? Successful gefitinib rechallenge after T790M disappearance in a patient with *EGFR*-mutant non-small-cell lung cancer. *J. Thorac. Oncol.* **8**, e27–e29 (2013).
160. Hata, A. *et al.* Panitumumab rechallenge in chemorefractory patients with metastatic colorectal cancer. *J. Gastrointest. Cancer* **44**, 456–459 (2013).
161. Hata, A., Katakami, N. & Kitajima, N. Successful cetuximab therapy after failure of panitumumab rechallenge in a patient with metastatic colorectal cancer: restoration of drug sensitivity after anti-EGFR monoclonal antibody-free interval. *J. Gastrointest. Cancer* **45**, 506–507 (2014).
162. Seghers, A. C., Wilgenhof, S., Lebbé, C. & Neyns, B. Successful rechallenge in two patients with BRAF-V600-mutant melanoma who experienced previous progression during treatment with a selective BRAF inhibitor. *Melanoma Res.* **22**, 466–472 (2012).
163. Nakamura, T. *et al.* Application of a highly sensitive detection system for epidermal growth factor receptor mutations in plasma DNA. *J. Thorac. Oncol.* **7**, 1369–1381 (2012).
164. Gremel, G. *et al.* Distinct sub-clonal tumour responses to therapy revealed by circulating cell-free DNA. *Ann. Oncol.* **27**, 1959–1965 (2016).
165. El-Hefnawy, T. *et al.* Characterization of amplifiable, circulating RNA in plasma and its potential as a tool for cancer diagnostics. *Clin. Chem.* **50**, 564–573 (2004).
166. Sørlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869–10874 (2001).

167. Sotiriou, C. & Pusztai, L. Gene-expression signatures in breast cancer. *N. Engl. J. Med.* **360**, 790–800 (2009).
168. Cardoso, F. *et al.* 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N. Engl. J. Med.* **375**, 717–729 (2016).
169. Jordan, N. V. *et al.* HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* **537**, 102–106 (2016).
170. Kimura, H. *et al.* Detection of epidermal growth factor receptor mutations in serum as a predictor of the response to gefitinib in patients with non-small-cell lung cancer. *Clin. Cancer Res.* **12**, 3915–3921 (2006).
171. Han, B. *et al.* Determining the prevalence of *EGFR* mutations in Asian and Russian patients (pts) with advanced non-small-cell lung cancer (aNSCLC) of adenocarcinoma (ADC) and non-ADC histology: IGNITE study [abstract 960]. *Ann. Oncol.* **26** (Suppl.1), 29–44 (2015).
172. Reck, M. *et al.* Investigating the utility of circulating-free tumour-derived DNA (ctDNA) in plasma for the detection of epidermal growth factor receptor (*EGFR*) mutation status in European and Japanese patients (pts) with advanced non-small-cell lung cancer (NSCLC): ASSESS study. [abstract 35O_PR]. *Ann. Oncol.* **26** (Suppl.1), i57–i61 (2015).
173. Qiu, M. *et al.* Circulating tumor DNA is effective for the detection of *EGFR* mutation in non-small cell lung cancer: a meta-analysis. *Cancer Epidemiol. Biomarkers Prev.* **24**, 206–212 (2015).
174. Luo, J., Shen, L. & Zheng, D. Diagnostic value of circulating free DNA for the detection of *EGFR* mutation status in NSCLC: a systematic review and meta-analysis. *Sci. Rep.* **4**, 6269 (2014).
175. Bordi, P., Del Re, M., Danesi, R. & Tiseo, M. Circulating DNA in diagnosis and monitoring *EGFR* gene mutations in advanced non-small cell lung cancer. *Transl Lung Cancer Res.* **4**, 584–597 (2015).
176. Karachaliou, N. *et al.* Association of *EGFR* L858R mutation in circulating free DNA with survival in the EURTAC trial. *JAMA Oncol.* **1**, 149–157 (2015).
177. Zhou, C. *et al.* Erlotinib versus chemotherapy as first-line treatment for patients with advanced *EGFR* mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol.* **12**, 735–742 (2011).
178. Mok, T. *et al.* Detection and dynamic changes of *EGFR* mutations from circulating tumor DNA as a predictor of survival outcomes in NSCLC patients treated with first-line intercalated erlotinib and chemotherapy. *Clin. Cancer Res.* **21**, 3196–3203 (2015).
179. Weber, B. *et al.* Detection of *EGFR* mutations in plasma and biopsies from non-small-cell lung cancer patients by allele-specific PCR assays. *BMC Cancer* **14**, 294 (2014).
180. Cross, D. A. *et al.* AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discov.* **4**, 1046–1061 (2014).
181. Zhou, W. *et al.* Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. *Nature* **462**, 1070–1074 (2009).
182. Oxnard, G. R. *et al.* Association between plasma genotyping and outcomes of treatment with osimertinib (AZD9291) in advanced non-small-cell lung cancer. *J. Clin. Oncol.* **34**, 3375–3382 (2016).
183. Sacher, A. G. *et al.* Prospective validation of rapid plasma genotyping for the detection of *EGFR* and *KRAS* mutations in advanced lung cancer. *JAMA Oncol.* **2**, 1014–1022 (2016).
184. Guardant Health. Lunar. *Guardanthealth.com* http://www.guardanthealth.com/lunar/ (2016).
185. Jiang, P. *et al.* Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl Acad. Sci. USA* **112**, E1317–E1325 (2015).
186. US Food and Drug Administration. Document Number: GEN1500674. *FDA* http://www.fda.gov/downloads/MedicalDevices/ResourcesforYou/Industry/UCM464092.pdf (2015).
187. US Food and Drug Administration. Principles for codevelopment of an *in vitro* companion diagnostic device with a therapeutic product. *FDA* http://www.fda.gov/ucm/groups/fdagov-public/@fdagov-meddev-gen/documents/document/ucm510824.pdf (2016).
188. RNCOS Business Consultancy Services. Global Liquid Biopsy Market Outlook to 2020. *RNCOS* http://www.rncos.com/Market-Analysis-Reports/Global-Liquid-Biopsy-Market-Outlook-to-2020-IM815.htm (2016).
189. Sequist, L. V. *et al.* Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with *EGFR* mutations. *J. Clin. Oncol.* **31**, 3327–3334 (2013).
190. US Food and Drug Administration. therascreen® EGFR RGQ PCR kit instructions for use (handbook). *FDA* http://www.accessdata.fda.gov/cdrh_docs/pdf12/P120022c.pdf (2013).
191. Douillard, J. Y. *et al.* First-line gefitinib in Caucasian *EGFR* mutation-positive NSCLC patients: a phase-IV, open-label, single-arm study. *Br. J. Cancer* **110**, 55–62 (2014).
192. European Medicines Agency. Annex I: summary of product characteristics. EMA http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/001016/WC500036358.pdf (2014).
193. US Food and Drug Administration. PMA P150047: FDA summary of safety and effectiveness data page 1 summary of safety and effectiveness data (SSED). *FDA* http://www.accessdata.fda.gov/cdrh_docs/pdf15/P150047b.pdf (2016).
194. US Food and Drug Administration. cobas EGFR Mutation Test v2. *FDA* http://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm504540.htm (2016).
195. Jenkins, S. *et al.* Plasma ctDNA analysis for detection of *EGFR* T790M mutation in patients (pts) with *EGFR* mutation-positive advanced non-small cell lung cancer (aNSCLC) [abstract 134O_PR]. *J. Thorac. Oncol.* **11** (Suppl. 4), S57–S166. (2016).
196. Sacher, A. G. Prospective validation of rapid plasma genotyping for the detection of *EGFR* and *KRAS* mutations in advanced lung cancer. *JAMA Oncol.* **2**, 1014–1022 (2016).
197. Lanman, R. B. *et al.* Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. *PLoS ONE* **10**, e0140712 (2015).
198. Zill, O. A. *et al.* Somatic genomic landscape of over 15,000 patients with advanced-stage cancer from clinical next-generation sequencing analysis of circulating tumor DNA [abstract]. *J. Clin. Oncol.* **34** (Suppl.), LBA11501 (2016).
199. Devonshire, A. S. *et al.* Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. *Anal. Bioanal. Chem.* **406**, 6499–6512 (2014).
200. Swinkels, D. W., Wiegerinck, E., Steegers, E. A. & de Kok, J. B. Effects of blood-processing protocols on cell-free DNA quantification in plasma. *Clin. Chem.* **49**, 525–526 (2003).
201. Ignatiadis, M. *et al.* International study on inter-reader variability for circulating tumor cells in breast cancer. *Breast Cancer Res.* **16**, R43 (2014).
202. Sorenson, G. D. Detection of mutated *KRAS2* sequences as tumor markers in plasma/serum of patients with gastrointestinal cancer. *Clin. Cancer Res.* **6**, 2129–2137 (2000).
203. Kahlert, C. *et al.* Identification of double-stranded genomic DNA spanning all chromosomes with mutated *KRAS* and *p53* DNA in the serum exosomes of patients with pancreatic cancer. *J. Biol. Chem.* **289**, 3869–3875 (2014).
204. Thakur, B. K. *et al.* Double-stranded DNA in exosomes: a novel biomarker in cancer detection. *Cell Res.* **24**, 766–769 (2014).
205. Pixberg, C. F., Schulz, W. A., Stoecklein, N. H. & Neves, R. P. Characterization of DNA methylation in circulating tumor cells. *Genes (Basel)* **6**, 1053–1075 (2015).
206. Yamamoto, H. *et al.* BARHL2 methylation using gastric wash DNA or gastric juice exosomal DNA is a useful marker for early detection of gastric cancer in an *H. pylori*-independent manner. *Clin. Transl Gastroenterol.* **7**, e184 (2016).
207. Krebs, M. G. *et al.* Evaluation and prognostic significance of circulating tumor cells in patients with non-small-cell lung cancer. *J. Clin. Oncol.* **29**, 1556–1563 (2011).
208. Watanabe, K. *et al.* EGFR mutation analysis of circulating tumor DNA using an improved PNA-LNA PCR clamp method. *Can. Respir. J.* **2016**, 5297329 (2016).
209. Spindler, K. L., Pallisgaard, N., Vogelius, I. & Jakobsen, A. Quantitative cell-free DNA, *KRAS*, and *BRAF* mutations in plasma from patients with metastatic colorectal cancer during treatment with cetuximab and irinotecan. *Clin. Cancer Res.* **18**, 1177–1185 (2012).
210. Milbury, C. A. *et al.* Multiplex amplification coupled with COLD-PCR and high resolution melting enables identification of low-abundance mutations in cancer samples with low DNA content. *J. Mol. Diagn.* **13**, 220–232 (2011).
211. Diehl, F. *et al.* BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat. Methods* **3**, 551–559 (2006).
212. Sanmamed, M. F. *et al.* Quantitative cell-free circulating $BRAF^{V600E}$ mutation analysis by use of droplet digital PCR in the follow-up of patients with melanoma being treated with BRAF inhibitors. *Clin. Chem.* **61**, 297–304 (2015).
213. Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83**, 8604–8610 (2011).
214. Taly, V. *et al.* Multiplex picodroplet digital PCR to detect *KRAS* mutations in circulating DNA from the plasma of colorectal cancer patients. *Clin. Chem.* **59**, 1722–1731 (2013).
215. Mouliere, F. *et al.* Circulating cell-free DNA from colorectal cancer patients may reveal high *KRAS* or *BRAF* mutation load. *Transl Oncol.* **6**, 319–328 (2013).
216. Mouliere, F., El Messaoudi, S., Pang, D., Dritschilo, A. & Thierry, A. R. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol. Oncol.* **8**, 927–941 (2014).
217. Rothé, F. *et al.* Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. *Ann. Oncol.* **25**, 1959–1965 (2014).
218. Forshew, T. *et al.* Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl Med.* **4**, 136ra68 (2012).
219. Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl Acad. Sci. USA* **108**, 9530–9535 (2011).
220. Newman, A. M. *et al.* Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat. Biotechnol.* **34**, 547–555 (2016).
221. Leary, R. J. *et al.* Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci. Transl Med.* **4**, 162ra154 (2012).
222. Diaz, L. A., Sausen, M., Fisher, G. A. & Velculescu, V. E. Insights into therapeutic resistance from whole-genome analyses of circulating tumor DNA. *Oncotarget* **4**, 1856–1857 (2013).
223. Chan, K. C. *et al.* Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin. Chem.* **59**, 211–224 (2013).
224. Leary, R. J. *et al.* Development of personalized tumor biomarkers using massively parallel sequencing. *Sci. Transl Med.* **2**, 20ra14 (2010).
225. McBride, D. J. *et al.* Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes Chromosomes Cancer* **49**, 1062–1069 (2010).

Acknowledgements

We thank Elizabeth Cook, a US-based freelancer graphic artist, for her assistance in drafting the figures for this article, and Beth Van Emburgh and Cosimo Martino of the Candiolo Cancer Institute for their assistance in revising the text. The work of the G.S. and A.B. is supported by the European Community's Seventh Framework Programme under grant agreement no. 602901 MErCuRIC, grant agreement no. 635342–2 MoTriColor, and IMI contract n. 115749 CANCER-ID; AIRC (Associazione Italiana per la Ricerca sul Cancro) 2010 Special Programme Molecular Clinical Oncology 5 per mille, project no. 9970; Fondazione Piemontese per la Ricerca sul Cancro-ONLUS 5 per mille 2010 e 2011 Ministero della Salute; and AIRC Investigator Grants project 16788.

Author contributions

All authors made substantial contributions to researching data for article, discussions of content, and writing, review and editing of manuscript before submission.

Competing interests statement

G.S. is a consultant for Trovagene. A.B. is a member of the scientific advisory board for Biocartis, Horizon Discovery, and Trovagene. S.M. and S.S. declare no competing interests.

SUPPLEMENTARY INFORMATION

See online article: S1 (table)

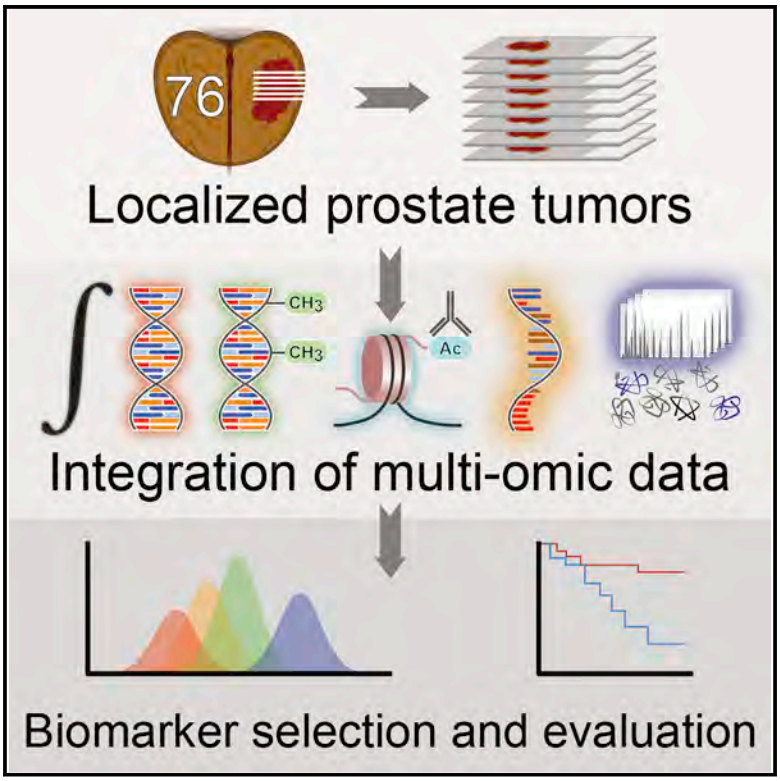
ALL LINKS ARE ACTIVE IN THE ONLINE PDF

Article

Cancer Cell

The Proteogenomic Landscape of Curable Prostate Cancer

Graphical Abstract



Authors

Ankit Sinha, Vincent Huang, Julie Livingstone, ..., Robert G. Bristow, Thomas Kislinger, Paul C. Boutros

Correspondence

thomas.kislinger@utoronto.ca (T.K.), pboutros@mednet.ucla.edu (P.C.B.)

In Brief

Sinha et al. determine the proteogenomic landscape of localized, intermediate-risk prostate cancers and show that the presence of ETS gene fusions has one of the strongest effects on the proteome. Prognostic biomarkers that integrate multi-omics significantly outperform those comprised of a single data type.

Highlights

- A comprehensive proteomic analyses of localized prostate cancers
- Integration of all levels of the central dogma (DNA → RNA → protein)
- ETS fusions have divergent effects on transcriptome and proteome
- Combining genomics and proteomics improves biomarker performance

Sinha et al., 2019, Cancer Cell *35*, 414–427
March 18, 2019
https://doi.org/10.1016/j.ccell.2019.02.005

CellPress

CellPress

Cancer Cell
Article

The Proteogenomic Landscape of Curable Prostate Cancer

Ankit Sinha,[1,16] Vincent Huang,[2,16] Julie Livingstone,[2,16] Jenny Wang,[2,3,7] Natalie S. Fox,[1,2] Natalie Kurganovs,[2,4] Vladimir Ignatchenko,[4] Katharina Fritsch,[1,4] Nilgun Donmez,[5] Lawrence E. Heisler,[2] Yu-Jia Shiah,[2] Cindy Q. Yao,[2] Javier A. Alfaro,[1,2] Stas Volik,[5] Anna Lapuk,[5] Michael Fraser,[2] Ken Kron,[4] Alex Murison,[4] Mathieu Lupien,[1,2,4] Cenk Sahinalp,[5] Colin C. Collins,[5,6] Bernard Tetu,[9] Mehdi Masoomian,[8] David M. Berman,[3,7] Theodorus van der Kwast,[4,8] Robert G. Bristow,[1,4,10] Thomas Kislinger,[1,4,*] and Paul C. Boutros[1,2,11,12,13,14,15,17,*]

[1]Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada
[2]Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada
[3]Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON K7L 3N6, Canada
[4]Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 1L7, Canada
[5]Vancouver Prostate Centre, Vancouver, BC V6H 3Z6, Canada
[6]Department of Urologic Sciences, University of British Columbia, Vancouver, BC V5Z 1M9, Canada
[7]Queen's Cancer Research Institute, Queen's University, Kingston, ON K7L 3N6, Canada
[8]Department of Pathology, Laboratory Medicine Program, University Health Network, Toronto, ON M5G 2C4, Canada
[9]Department of Pathology and Research Centre of CHU de Québec-Université Laval, Québec City, QC G1V 4G2, Canada
[10]Department of Radiation Oncology, University of Toronto, Toronto, ON M5T 1P5, Canada
[11]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada
[12]Department of Human Genetics, University of California, 12-109 CHS, 10833 Le Conte Avenue, Los Angeles, CA 90095, USA
[13]Department of Urology, University of California, Los Angeles, CA 90024, USA
[14]Jonsson Comprehensive Cancer Centre, University of California, Los Angeles, CA 90024, USA
[15]Institute for Precision Health, University of California, Los Angeles, CA 90024, USA
[16]These authors contributed equally
[17]Lead Contact
*Correspondence: thomas.kislinger@utoronto.ca (T.K.), pboutros@mednet.ucla.edu (P.C.B.)
https://doi.org/10.1016/j.ccell.2019.02.005

SUMMARY

DNA sequencing has identified recurrent mutations that drive the aggressiveness of prostate cancers. Surprisingly, the influence of genomic, epigenomic, and transcriptomic dysregulation on the tumor proteome remains poorly understood. We profiled the genomes, epigenomes, transcriptomes, and proteomes of 76 localized, intermediate-risk prostate cancers. We discovered that the genomic subtypes of prostate cancer converge on five proteomic subtypes, with distinct clinical trajectories. ETS fusions, the most common alteration in prostate tumors, affect different genes and pathways in the proteome and transcriptome. Globally, mRNA abundance changes explain only ~10% of protein abundance variability. As a result, prognostic biomarkers combining genomic or epigenomic features with proteomic ones significantly outperform biomarkers comprised of a single data type.

INTRODUCTION

Prostate cancer remains the most common non-skin malignancy in men worldwide. In many regions, its incidence is increasing because of demographic shifts in population structure and increased life expectancy (Canadian Cancer Statistics Advisory Committee, 2017; Center et al., 2012; Torre et al., 2016). Prostate cancer is most often diagnosed while still localized, largely through screening with digital rectal examinations and quantitation of serum levels of prostate specific antigen (PSA). As a result, ~75% of patients receive definitive local treatment with either surgery or radiotherapy. The combination of these morbid treatments and the large number of indolent tumors diagnosed has led to significant overtreatment, creating an urgent

Significance

Our data demonstrate that the prostate cancer proteome is shaped by the complex interplay of genomic, epigenomic, transcriptomic, and post-transcriptional dysregulation. Integration of data along the central dogma enables both a deeper biological insight and the development of multi-omic biomarkers with improved performance.

Table 1. Clinical Characteristics of the Patient Cohort

	BCR	
	Yes (36)	No (39)
Clinical ISUP Group		
1	1	3
2	28	30
3	7	6
Age at Treatment		
40–50 years	2	2
50–65 years	24	25
65–70 years	8	7
≥70 years	2	5
Pre-treatment PSA		
<10 ng/mL	27	27
≥10 ng/mL	9	12
Clinical T Category		
T1	12	21
T2	24	18
ETS fusion		
Present	21	17
Absent	15	22
Data Type		
Whole-genome sequencing	36	38
Copy number aberration	36	37
H3K27Ac	16	19
Methylation	34	38
RNA	25	30
Proteomics	36	39

need for more appropriate prognostic assays (Cooperberg et al., 2009).

The genome and epigenome of prostate cancer have been well studied. Large consortia have cataloged genomic and transcriptomic aberrations, including driver events (Baca et al., 2013; Fraser et al., 2017; Hopkins et al., 2017; The Cancer Genome Atlas Research Network, 2015). Subclonal reconstructions have discovered dramatic intra-tumoral heterogeneity and subclonal selection during disease evolution (Boutros et al., 2015; Cooper et al., 2015; Espiritu et al., 2018; Gundem et al., 2015). The epigenome of localized disease has been analyzed, both for CpG methylation and chromatin marks (Brocks et al., 2014; Kron et al., 2017). Candidate prognostic biomarkers have been developed using copy number and transcriptome data (Blume-Jensen et al., 2015; Cuzick et al., 2011; Den et al., 2015; Klein et al., 2014; Lalonde et al., 2014). Consequently, many mutations are known to drive the tumorigenesis and aggressivity of localized prostate cancer, with alteration of pathways including hypoxia response, androgen signaling, and DNA repair (The Cancer Genome Atlas Research Network, 2015).

However, the ways in which the proteome of localized prostate cancer is shaped by genomic, epigenomic, and transcriptomic aberrations is almost entirely unknown. Recent studies in breast, ovarian, and colorectal cancer have suggested that the transcriptome is a poor proxy for the proteome, with only 10%–20% of variation in protein abundance explained by mRNA abundance (Mertins et al., 2016; Zhang et al., 2014, 2016). Previous studies of the prostate cancer proteome have focused on cohorts of patients with inconsistent clinical features (Drake et al., 2016; Iglesias-Gato et al., 2016, 2018; Latonen et al., 2018; The Cancer Genome Atlas Research Network, 2015), and indeed no tumor type has yet integrated whole-genome, epigenome, and transcriptome data. Improved understanding of the dysregulation of the prostate cancer proteome can enhance interpretation of driver aberrations and facilitate development of rapidly implementable clinical assays based on immunohistochemistry techniques. To fill this gap, we performed a proteogenomic analysis of a richly annotated cohort of localized prostate cancers.

RESULTS

The Proteome of Curable Prostate Cancers

To understand the flow of biological information in prostate tumors, we assembled a clinically homogeneous cohort of 76 patients diagnosed with sporadic, localized, treatment-naive intermediate-risk prostate cancer (Table 1). All patients were treated by radical prostatectomy, with a median follow-up of 6.8 years (Figure 1A; Table S1). The histologically most representative regions (i.e., the index lesion, used for initial diagnosis and treatment; see STAR Methods) were subject to array-based copy number aberration (CNA) profiling and whole-genome sequencing (WGS) to detect genomic rearrangements (GRs), single nucleotide variants (SNVs), chromothripsis, and kataegis. The epigenome was evaluated with methylome profiling and for 35 cases, the *cis*-regulatory element landscape was assessed using histone H3K27Ac chromatin immunoprecipitation sequencing (ChIP-seq) (Kron et al., 2017). The transcriptome was quantified with both RNA sequencing (RNA-seq) and microarrays. Finally, the proteome was quantified via mass spectrometry-based shotgun proteomics, with each sample analyzed in duplicate. Globally, replicate analyses demonstrated high correlation both in terms of detection (>85% detected in both replicates) and quantification (Pearson's correlation >0.95; data not shown). In Data S1, we provide a proteogenomic fingerprint for every analyzed patient tumor, including data for the replicate proteome analyses demonstrating high correlation. Tumors were sequenced to a mean coverage of 79x ± 28x and normal blood reference samples to a mean coverage of 46x ± 17x. To ensure detection of low-abundance transcripts and accurate quantitation of the full dynamic range, ultra-deep RNA-seq was performed (median, 382 ± 138 million reads per tumor).

We detected 7,054 protein groups (Table S2), corresponding to 6,924 protein coding genes. Of these, 3,397 protein groups were detected and quantified in all 76 patients (Figure 1B, top), including those corresponding to classic prostate cancer-associated genes like the PSA gene *KLK3* and the DNA damage repair gene *ATM*, for which both germline polymorphisms and somatic SNVs are associated with patient outcome (Fraser et al., 2017; Pritchard et al., 2016). We separated the 7,054 protein groups into deciles based on their median abundance

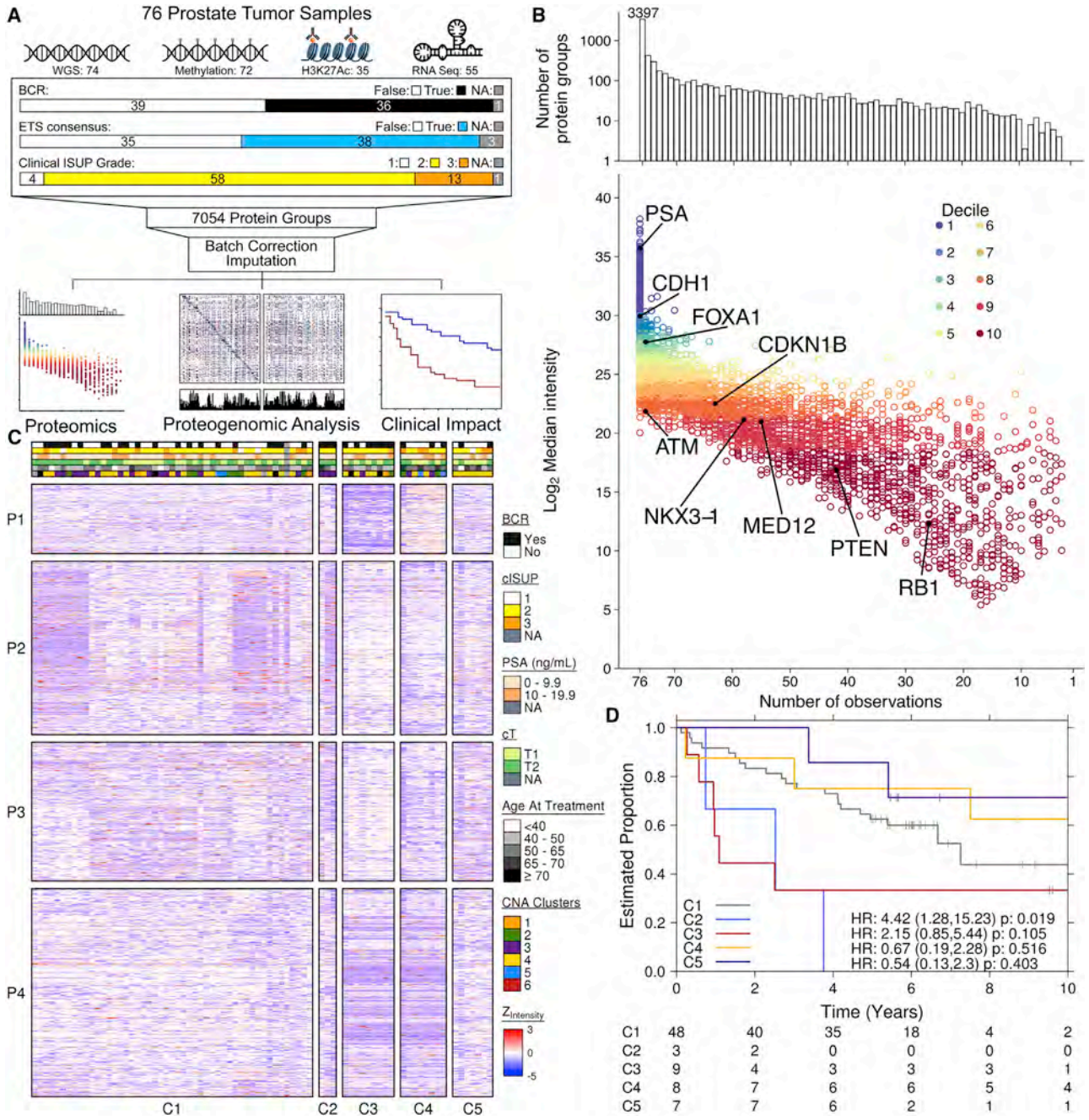


Figure 1. Proteomic Landscape of Curable Prostate Cancer

(A) Study overview showing the clinical characteristics of the cohort (n = 76) and the number of samples with whole-genome sequencing, RNA-seq, methylation data, and ChIP-seq data. Mass spectrometry yielded 7,054 protein groups, whose abundance was corrected for batch effects, and missing values were imputed prior to downstream analyses.

(B) Distribution of protein quantitation measured as median intensity by the number of samples they are detected in. Bar plot on top shows the total counts of proteins quantified in various number of samples. Missing values were omitted when calculating the median.

(C) Consensus clustering of 76 patients (K = 5) using the top 25% most variable genes (n = 1,800, K = 5). Clinical covariates are shown in the heatmap above, indicating for each patient; biochemical relapse (BCR), clinical ISUP grade (cISUP), PSA levels, clinical T category (cT), and age at treatment (years).

(D) Subtypes identified from consensus clustering were evaluated to determine their association with BCR. A Cox PH model was fitted for subtypes C2, C3, C4, and C5 against the baseline group of subtype C1. Hazard ratios and p values are shown with confidence intervals in parentheses.

HR, hazard ratio; ISUP, International Society for Urological Pathology; PSA, prostate specific antigen. See also Figure S1, Tables S1, S2, S3, and S4, and Data S1.

(Figure 1B, bottom). As expected, high-abundance proteins were observed in a larger fraction of samples, replicating previous mass spectrometry results (Kislinger et al., 2006; Liu et al., 2004). Proteins encoded by most prostate cancer driver genes were detected in over 70% of the analyzed tumors, including MED12, FOXA1, NKX3-1, and PTEN, among others.

To understand the global proteomic patterns of primary localized prostate cancer, we performed subtype discovery, identifying four clusters of proteins (i.e., P1, P2, P3, and P4) and five clusters of patients (i.e., C1, C2, C3, C4, and C5; Figure 1C). Protein clusters P1 and P3 are enriched for products of immune-related genes (Sallari et al., 2016), whereas no significantly enriched pathways were detected in P2 and P4 (Table S3). Of the five distinct patient subgroups, C2 and C3 are associated with an increased rate of biochemical recurrence (BCR; Figure 1D). Because CNAs are tightly associated with patient outcome (Hieronymus et al., 2014), we compared these five proteomic clusters with our previously described genomic subtypes and to the TCGA prostate cancer subtypes (Lalonde et al., 2014; The Cancer Genome Atlas Research Network, 2015). Genomic and proteomic subtypes were largely independent (adjusted Rand index [ARI] = −0.004, Figure S1A; ARI = 0.037, Figure S1B). This suggests that nucleotide features are poor proxies for proteomic diversity. Proteomic subtypes were also independent of androgen receptor activity signatures (Stelloo et al., 2015), as expected for treatment-naive, hormone-sensitive tumors (Figure S1C).

While patient clusters were generally not significantly associated with mutational burden in either SNVs or GRs (Table S3), the abundances of specific proteins were associated with clinical phenotypes. Percent genome altered (PGA), a biomarker of aggressive disease (Lalonde et al., 2014), was found to be associated with 421 proteins, where the most significant associations included a protein involved in Wnt signaling, FZD7 (Spearman's $\rho = -0.59$; false discovery rate [FDR] = 6.18×10^{-4}), and a deubiquitinase, USP11 (Spearman's $\rho = -0.58$; FDR = 6.18×10^{-4}) (Figure S1D, Table S4). Tumor size was associated with the abundance of eight proteins, while the presence of the aggressive intraductal carcinoma/cribriform architecture sub-pathology was associated with seven, including *PTEN* ($\Delta_{median\_protein} = -12.96$; FDR = 0.227), recapitulating previous findings (Bhandari et al., 2019; Chua et al., 2017).

ETS Gene Fusions Are Linked to Cell Migration and Lipid Metabolism

One of the strongest effects on the proteome was the presence of ETS gene fusions. These fusions are the most frequent somatic aberration in prostate cancer and are not associated with clinical outcome (Dal Pra et al., 2013; Minner et al., 2011). We focused on 245 mRNAs and 68 proteins significantly associated with ETS gene fusion status (Table S4; mRNA Q < 0.01; protein Q < 0.05; 36 overlapping genes, 277 genes in total). To be conservative, we excluded 22 genes with a high proportion of missing protein abundance measurements. Overall changes in mRNA and protein abundances were well correlated (Spearman's $\rho = 0.72$, $p < 2.2 \times 10^{-16}$; Figure 2A), but protein abundances showed larger dynamic range than mRNA abundances. The median differentially abundant mRNA differed 1.50-fold between ETS fusion-positive and ETS fusion-negative tumors, while the median protein differed 1.66-fold ($p = 4.63 \times 10^{-6}$; paired Mann-Whitney U test). The many genes showing only mRNA or protein abundances associated with ETS fusions, but not both, may be attributed to biological factors like translational and post-translational regulation, as well as to technical factors.

For some individual genes, mRNA and protein abundances diverged dramatically. The transcription factor EB (TFEB) was almost unchanged at the RNA level (1.46-fold higher in tumors with an ETS gene fusion) but was 1,012-fold higher at the protein level. Similarly, lysyl oxidase (LOX) was 1.88-fold higher at the RNA level in tumors with an ETS gene fusion, but had 21,031-fold higher protein abundance. While relative quantification by label-free proteomics is well established, the presence of missing values (i.e., protein not detected and quantified in a sample) are caveats for binary comparisons. More accurate, absolute quantitation via targeted proteomics assays and stable isotope labeled standards, are needed to better understand these divergences.

To better understand the differences in ETS fusion-associated genes in the transcriptome and proteome, we expanded our analysis to methylation, histone status (H3K27Ac), and copy number data (Figure 2B). Only a single gene, *ARHGDIB* was associated with ETS gene fusions at the protein, mRNA, methylation, and acetylation levels. One gene contained in the deletion region between *TMPRSS2* and *ERG* on chromosome 21, *FAM3B*, showed correlated CNAs, methylation changes, and mRNA abundance changes. By contrast, 630 genes showed differential methylation associated with ETS gene fusions, while 124 showed differential H3K27 acetylation. These interactions do not fully explain the modest overlap between ETS-associated proteins and ETS-associated mRNAs, and highlights the importance of post-transcriptional regulatory factors not easily quantified by -omic studies.

To determine if functional inference from RNA and protein data would yield similar conclusions, we performed pathway analysis separately on ETS fusion-associated genes identified at each biological level (i.e., CNAs, methylation, H3K27Ac, RNA abundance, and protein abundance). No pathways were associated with CNAs, and only one with H3K27Ac (Figure 2C), although genes associated with differential H3K27 acetylation were enriched for ETS binding motifs (hypergeometric test; $p = 3.5 \times 10^{-2}$). Genes associated with carboxylic acid metabolism were enriched at the mRNA, protein, and methylation level, corroborating links between *ERG* fusions and lipid metabolism (Hansen et al., 2016; Wu et al., 2014). Genes associated with intra- and extracellular vesicles were enriched in the mRNAs and proteins associated with ETS gene fusions. At the mRNA level, we identified enrichment in cell migration, actin binding, and phospholipid binding, while at the protein level, there was an enrichment in lysosomal genes. These data suggest that a myriad of genomic mechanisms may differentiate the ETS-associated transcriptome from the ETS-associated proteome.

Interestingly, one patient showed ERG overexpression through immunohistochemistry with ERG antibodies, but no ETS gene fusion was detectable by either WGS or RNA-seq (Figure S2A). This tumor exhibited neither the mRNA nor protein signatures of ETS gene fusions (Figure S2B), suggesting not all cases of ERG overexpression validated by immunohistochemistry will have an impact on a tumor's transcriptional and

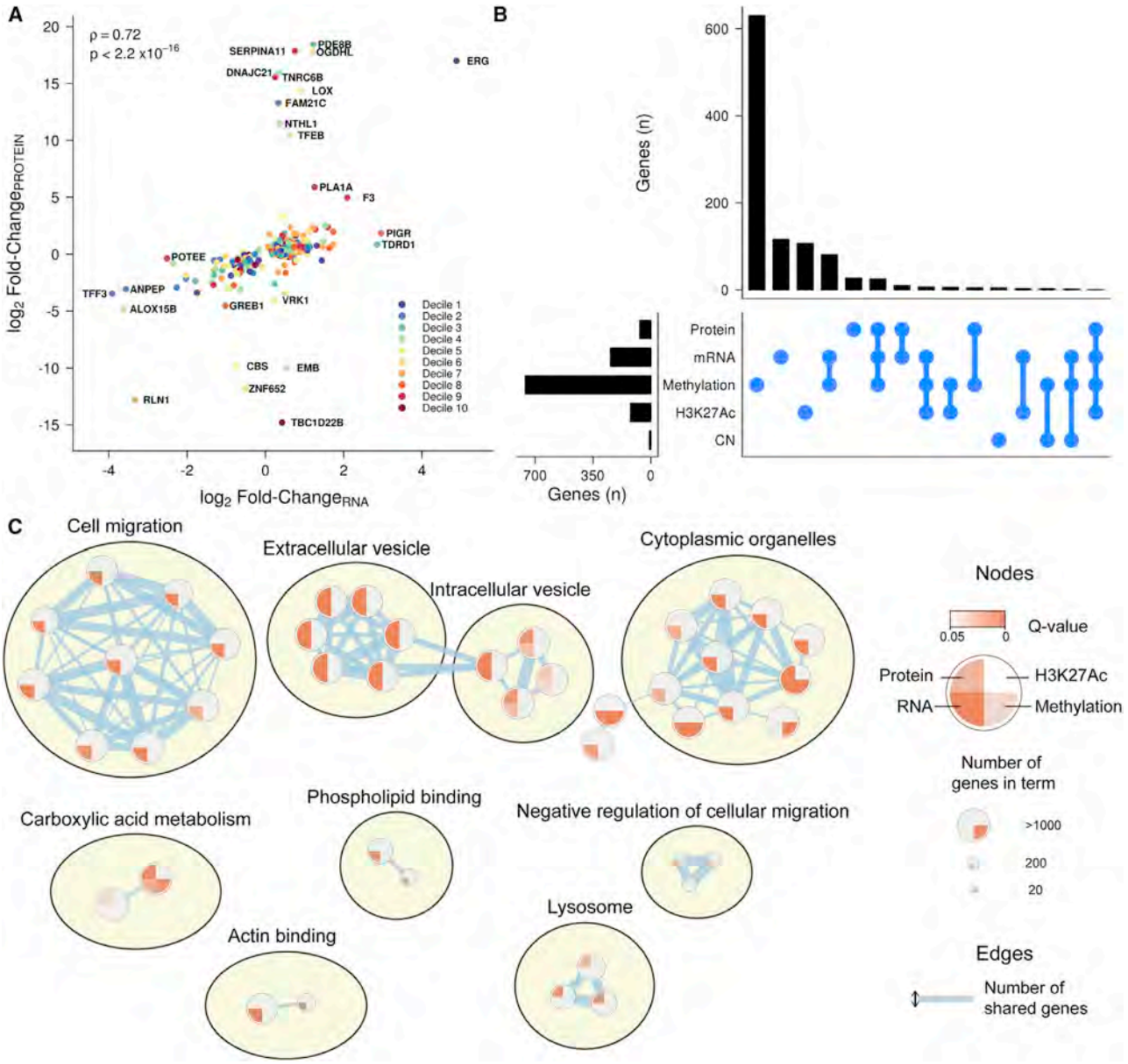


Figure 2. Transcriptomic and Proteomic Consequences of ETS fusions

(A) Comparison of the difference in protein and mRNA abundance observed between samples with an ETS gene fusion and those without. Analysis includes 55 samples with matched RNA-seq and protein data in 255 genes as 22 genes were removed due to a high proportion of missing protein data. Color indicates which protein abundance decile the gene is in, where purple indicates the most abundant. p Value is from Spearman's correlation.

(B) Number of overlapping ETS gene fusion-associated genes between protein, mRNA, methylation, H3K27Ac, and copy number status. Bar plot on the left indicates the total number of associated genes in that data type. Bar plot on top shows the number of genes in the singleton or intersection groups as indicated by the dots below.

(C) Pathway enrichment analysis performed using g:Profiler on the five sets of genes associated with ETS gene fusions in the different data types. Large clusters of similar pathways are outlined in yellow and labeled. Singleton nodes were omitted. No pathway enrichment was detected in copy number changes associated with ETS gene fusions.

See also Figure S2.

proteomic repertoire, or potentially reflecting the large spatial heterogeneity of prostate tumor genomes (Boutros et al., 2015). Critically, divergences in signaling pathway detected by transcriptome and proteome data suggest caution when interpreting the effects of genomic aberrations on the basis of the transcriptome alone. Enhanced study of protein abundances in such analyses is key to fully understand the effects of genomic aberrations.

Quantifying Transcriptome-Proteome Discordance

These large discordances between differences in ETS gene fusion-associated mRNA abundance and protein abundance

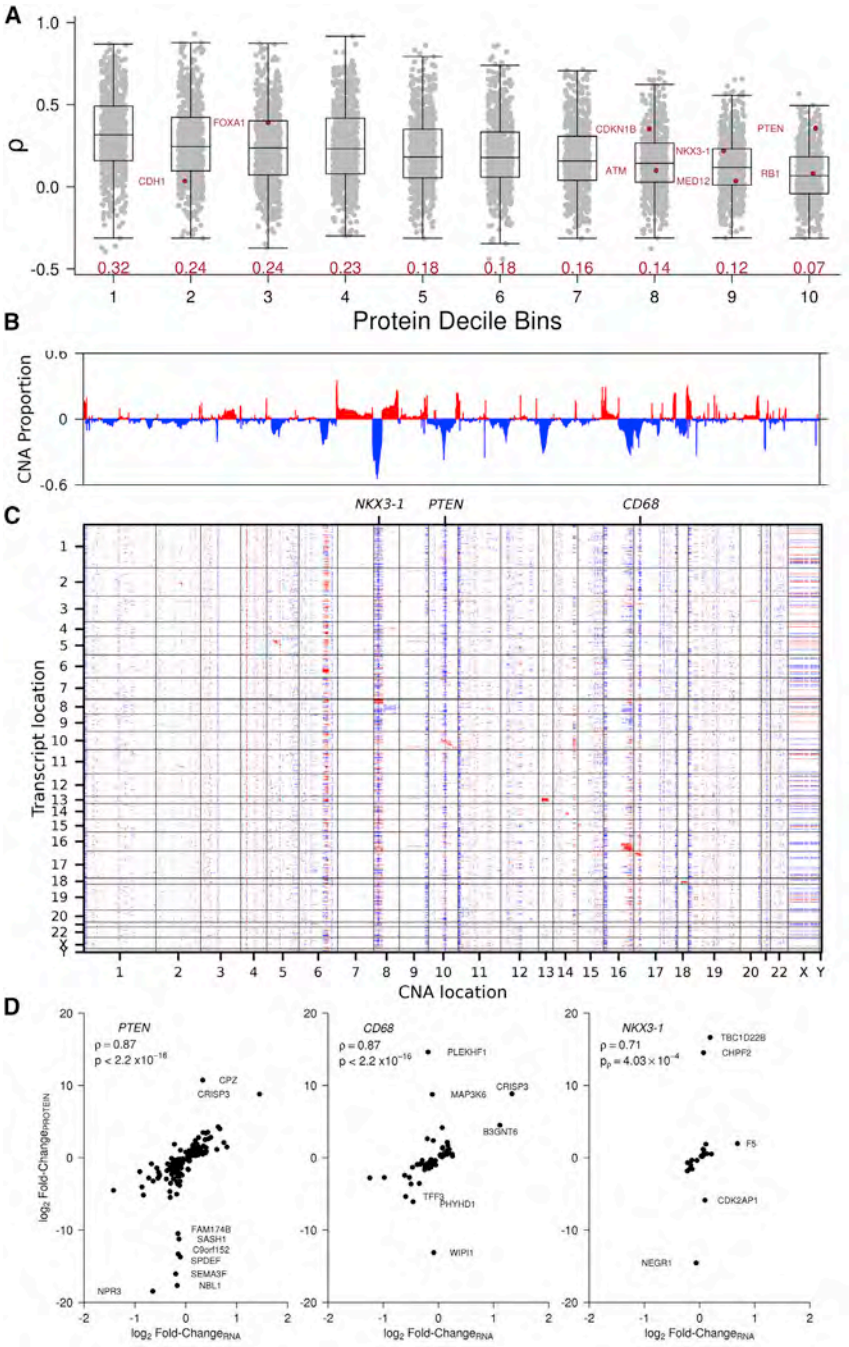


Figure 3. *trans* Proteomic Effects of Somatic CNAs

(A) Distribution of RNA-protein Spearman's ρ in each decile of protein abundance. Median correlations of each decile are indicated in red along the x axis. Known genes of interest are highlighted and labeled in red. Boxplots depict the upper and lower quartiles, with the median shown as a solid line; whiskers indicate 1.5 times the interquartile range (IQR). Data points outside the IQR are shown.

(B) The proportion of samples that contain copy number amplifications (red) and deletions (blue) in 210 samples with mRNA data.

(C) A global overview of the difference in mRNA abundance for each CNA locus comparing abundance from samples with a CNA with those without. Positive fold changes (i.e., higher abundance in samples with an amplification) are shown in red, negative fold changes (i.e., lower abundance in samples with a deletion) are shown in blue (FDR <0.05). The x axis plots 23,068 CNAs and the y axis plots 6,636 mRNA genes. Genes are ordered by chromosome location on both axes.

(D) The fold change in mRNA and protein abundances in 55 matched samples (RNA-seq) comparing abundances in samples with a deletion and those without for *PTEN*, *CD68*, and *NKX3-1*. Only genes that show significant fold changes at the mRNA (Mann-Whitney U test; $p < 0.05$) and protein level (Mann-Whitney U test; $p < 0.05$) are plotted. See also Figures S3 and S4, and Table S5.

led us to systematically quantify their relationship across all genes. Globally, mRNA and protein abundances are weakly correlated (median Spearman's ρ = 0.21), indicating that mRNA abundance is poorly predictive of protein abundance (Figure 3A). RNA-protein correlations varied with protein abundance; the 10% most abundant proteins were much better correlated (median, 0.32) than the bottom 10% (median, 0.07).

One specific example of this phenomenon is ATM, where we detected relatively low protein abundance (8th decile), but higher RNA abundance (3rd decile) and a weak correlation between them (Spearman's ρ = 0.10; Figure S3A). *ATM* SNVs are associated with patient outcome, but there was no significant difference in biochemical relapse rate between patients with and without *ATM* loss (Figure S3B). This poor correlation between transcript and protein abundance may reflect differing rates of mRNA degradation, translation, or protein degradation, and mirrors recent reports in breast, ovarian, and colorectal cancers (Mertins et al., 2016; Zhang et al., 2014, 2016).

RNA-protein correlations were generally less dependent on transcript abundance than on protein abundance (Figure S3C). This may reflect some combination of larger translational regulation or increased measurement error for low-abundance biomolecules (i.e., both transcripts and proteins). Indeed, the 10% most abundant proteins were enriched for localization to membrane-bound organelles and extracellular proteins, and these trends held in an independent, clinically diverse validation cohort (Figure S3D) (Iglesias-Gato et al., 2016).

In extreme cases, either protein or RNA for a gene is detected but the other is not. We examined these cases, focusing on the 10% most abundant transcripts and proteins to minimize the possibility of technical false-negatives. Of the 4,694 most abundant transcripts, 1,342 did not have a detected protein, including 1,070 from known coding genes. By contrast, only 68 of the most

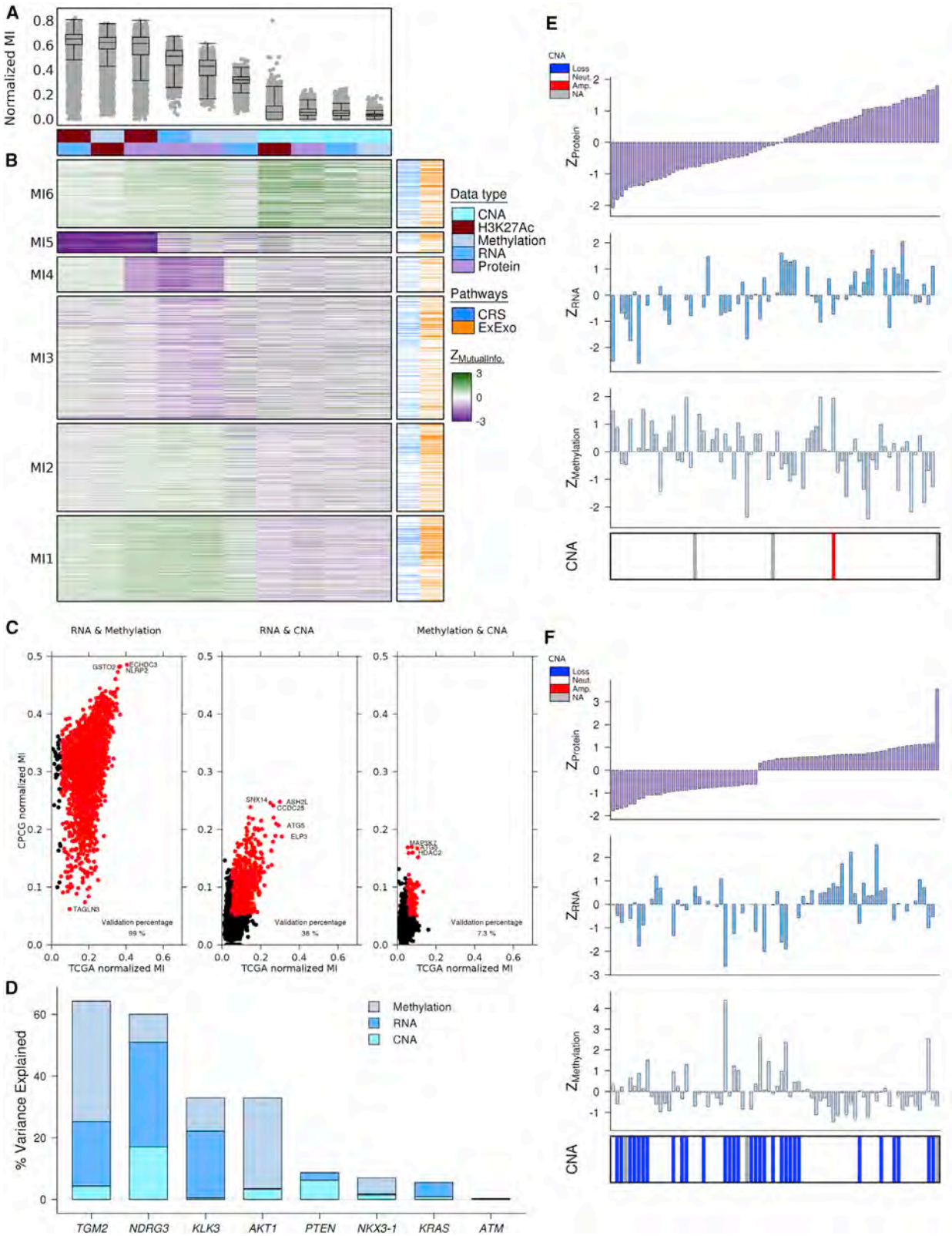


Figure 4. Integrated Clustering of Multi-omics Data

(A) Distribution of the normalized mutual information (MI) for each data-type pair. Boxplots depict the upper and lower quartiles, with the median shown as a solid line; whiskers indicate 1.5 times the interquartile range (IQR). Data points outside the IQR are shown.

(legend continued on next page)

highly abundant proteins had low or undetected transcript abundance (Table S5). Coding transcripts without a detected protein represent a diverse collection of genes preferentially localized to the nucleus, while proteins without detected transcripts are enriched for immune-related genes (Table S5).

cis and *trans* Effects of Genomic and Transcriptomic Changes

Prostate cancer is driven by CNAs more than SNVs: it is a C-class tumor (Fraser et al., 2017). We therefore investigated the role recurrent CNAs play in modulating mRNA and protein abundance, both in *cis* and in *trans* (Figures 3B and 3C). While genes may be lost or gained as part of a larger segment, the effect on mRNA or protein abundance may vary per gene, therefore we investigated differences per gene and not per segment. To increase statistical power, we generated 210 array-based transcriptomes from localized prostate tumors. To identify *cis* effects, for each gene we compared mRNA and protein abundances between tumors with and without a CNA at that gene (6,607 genes with all three types of data available). We detected strong *cis* effects in RNA, which can be seen along the diagonal of Figure 3C as influencing ~10% of all genes (592/6,607; FDR <0.05; t test). These effects were present, but weaker at the protein level, with ~2% of proteins having their abundance associated to CNAs (133/6,607; Figure S4A). We validated this result in TCGA data, which contained 491 samples with matching RNA and CNA data and found that 35% of genes show *cis* effects, highlighting the importance of sample size and, potentially, clinical diversity of the patient cohort, in multi-omic studies (Figure S4B).

Next, to identify *trans* effects where a CNA on one chromosome is associated with mRNA changes of a gene on another chromosome, we repeated our earlier analysis transcriptome- and proteome-wide. For each of the 23,068 genes with copy number information, we identified which of the 6,636 genes had both mRNA and protein abundance data and changed abundance with CNA status. On average, each gene-specific CNA had 593 ± 528 *trans* effects, where it was associated with statistically significant changes in RNA abundance. By contrast, *trans* effects were rarer for protein abundance, influencing 10 ± 31 genes (Figure S4A). For example, deletion of *PTEN* alters abundance in 52% of the genes investigated at the RNA level (3,416/6,607) but only 2.7% at the protein level (179/6,607), all of which showed RNA changes. To be conservative, we removed genes that were themselves frequently copy number altered (>5% of samples) to exclude confounding effects of genomic subtypes. Even after this control, *PTEN* showed *trans* effects on 54% of genes at the RNA level and 2.8% at the protein level (113/4,086; Figure 3D).

Other genes with substantial *trans* effects included *NKX3-1*, a tumor suppressor deleted in almost half of prostate tumors and *CD68*, which mediates macrophage recruitment (Figure 3D). Overall, 694 genes had large *trans* effects: defined as influencing the RNA levels of at least 10% of all genes. Of these, 67.4% (468/694) also exhibited *trans* effects at the protein level (affecting 0.2%–40% of proteins); the smaller effect sizes reflect the small number of samples with protein abundance measurements. Genes with large *trans* effects included *CMAS*, an immune-related gene, *ATAD1*, a gene related to ATPase activity, and *MINPP1*, a gene previously implicated in cancer. Interestingly, these three genes were associated with poor prognosis at both the CNA and RNA levels (Figure S4C). Consistent with our observations of ETS gene fusion-associated genes, protein abundances showed a much higher dynamic range for *trans* effects than transcript abundances (e.g., Spearman's ρ = 0.87 for *PTEN*; Figure 3D).

Not all genes were influenced by CNA *trans* effects at the same frequency. For example *CRISP3*, which plays a role in sperm function and is upregulated in prostate tumors (Ribeiro et al., 2011), shows RNA *trans* effects with 9.2% (2,123/23,068) of genes and protein *trans* effects with 0.4% (85/23,068), in some cases with large magnitudes. *CRISP3* RNA and protein are both more abundant in samples with either *CD68* deletion or *PTEN* deletion (Figure 3D). Thus, a large network of *trans* CNA effects exists, highlighting interconnections between specific somatic mutations and consequent transcriptome and proteome dysregulation.

Multi-layer Information Flow in Prostate Cancer

To better quantify this complex flow of information from the cancer genome to its proteome, we performed an information content analysis. For each gene, we calculated the mutual information (MI) between the five classes of molecular data in our prostate cancer study: CNAs, methylation, histones (H3K27 acetylation), RNA abundance, and protein abundance (Figure 4A). MI measures, in bits, the knowledge of one variable when a second variable becomes known. MI values of zero indicate independent variables: knowing one variable gives no information on the other. MI is related to classic correlations, but lacks some of their assumptions about linearity and ordering, making MI useful for complex relationships. To standardize analyses, we median-normalized MI separately for each dataset.

As expected, different pairs of molecules have varying amounts of redundancy in their information content. For example, CNAs were weakly predictive compared with other molecular data types. CNAs were modestly more tightly associated with protein than with mRNA abundance (median $MI_{CNA\text{-}Protein}$ = 0.055 versus median $MI_{CNA\text{-}RNA}$ = 0.048).

(B) Consensus clustering of normalized mutual information for each data-type pair. Biomolecules are indicated in the covariates along the top. Each row represents a gene (n = 6,484) comparison for which all data types exist. Adjacent plots indicate if genes are known to be associated with the selected pathways. Normalized MI are plotted as Z scores for visualization purposes.

(C) Correlation of normalized mutual information between our cohort and TCGA in genes with MI above 0.05. Red dots indicate genes that had normalized MI about 0.05 in both our dataset and TCGA.

(D) Percent variance explained of protein abundance modeled using copy number status, methylation, and mRNA abundance for a select set of genes known to be associated with prostate cancer.

(E and F) Integrated distribution plots of *KLK3* (E) and *PTEN* (F) showing copy number state and Z scored protein, mRNA, and methylation abundances for each of the 76 samples ordered by increasing protein abundance.

CRS, cellular response to stress; ExExo, extracellular exosomes. See also Figure S4 and Table S6.

Similarly, methylation status was more strongly linked to protein than mRNA abundance (median $MI_{Methylation\text{-}Protein}$ = 0.43 versus median $MI_{Methylation\text{-}RNA}$ = 0.32). Intriguingly, the highest mutual information across genomic regions was between H3K27Ac and RNA (median $MI_{H3K27Ac\text{-}RNA}$ = 0.652), while the lowest was between CNAs and methylation (median $MI_{CNA\text{-}Methylation}$ = 0.032). This may suggest a prominent role for epigenomic features, independent of the frequent subclonal CNAs (Espiritu et al., 2018).

To determine if the regulatory relationships between pairs of biomolecules distinguish specific functional groups of genes, we performed consensus clustering and identified six subgroups (labeled MI1-6; Figure 4B). Individual subgroups were not enriched for MSigDB hallmark gene sets, but rather more specific features (Liberzon et al., 2015) (Table S6). Subgroup MI6 was characterized by genes with higher CNA-H3K27Ac, CNA-protein, CNA-RNA, and CNA-methylation links, and was enriched for genes related to cellular response to stress, suggesting tight regulatory networks (FDR = 0.005; Table S6; Figure 4B). By contrast, MI1 harbored genes with strong H3K27Ac-protein, RNA-protein, and methylation-protein links and are enriched in extracellular exosomes (FDR = 3.87×10^{-15}; Table S6). These results are compelling, but further exploration will be required to fully elucidate the biological mechanisms and implications underlying these links in MI.

To validate our MI findings in an independent dataset, we calculated normalized MI in 245 intermediate-risk TCGA samples. For each pair of molecular data types, we considered genes with significant MI in the discovery cohort (defined as MI > 0.05). We then calculated the MI for these genes in the TCGA cohort, if the same molecular data type was collected within that cohort. MI values validated strongly, with 99% of genes with significant methylation-RNA and 75% of genes with significant CNA-RNA MI validating (Figure 4C). To quantify the validation of the MI analyses, we created a receiver operating characteristic curve for each molecular data type by iteratively increasing the MI threshold used for significance in both datasets (Figure S4D).

We followed by calculating the percent variance explained (PVE) by upstream *cis* information from CNA, methylation, and RNA. This analysis was performed on genes present in most samples with a known link to prostate cancer: *TGM2*, *NDRG3*, *KLK3*, *AKT1*, *PTEN*, *NKX3-1*, *KRAS*, and *ATM* (Figure 4D). A strong association was detected between methylation and protein abundances for *TGM2* and *AKT1* in which almost 40% and 30% of the protein variance, respectively, can be explained by methylation. Both *NDRG3* and *PTEN* show relatively high PVE by CNAs when compared with the other genes examined (17% and 6.2%, respectively), but while 60% of variability *NDRG3* protein abundance can be explained by CNA, methylation, and RNA, less than 10% of variability in *PTEN* abundance was captured by the model. Curiously, despite its high abundance, only 33% of the variance in *KLK3* was explained by RNA (21%) and methylation (11%). *KLK3* is generally copy number neutral (Figure 4E) and protein abundance is univariately correlated with RNA (ρ = 0.48; p = 2.53×10^{-4}) and methylation ($\rho = -0.34$; p = 3.3×10^{-3}). In contrast, *PTEN* is dominated by copy number losses (42%; 31/74, 2 missing; Figure 4F) and its protein abundance was univariately correlated with RNA (ρ = 0.29; p = 2.6×10^{-2}) and methylation ($\rho = -0.29$; p = 1.3×10^{-2}), but its low RNA values ($median_{PTEN}$ = 4.76; $median_{GAPDH}$ = 11.8) and low and narrow methylation values ($Q1,Q3_{PTEN}$ = 0.075,0.086; $Q1,Q3_{KLK3}$ = 0.12,0.13) may explain the low PVE.

These data provided detailed maps of the complex ways in which information flows from the germline and somatic genome, epigenome and transcriptome, and finally to the proteome. By improving quantitation of regulatory data, as by miRNA profiling or histone ChIP-seq, specific functional classes of genes can be delineated.

Protein Abundances May Predict Prostate Cancer Relapse

Finally, to evaluate the potential clinical importance of proteomic profiling of primary prostate tumors, we quantified the association of each gene with disease relapse after definitive local therapy with curative intent. We used time to biochemical relapse (BCR) as our outcome, which reflects rising serum PSA levels, which can trigger administration of salvage therapy. For each gene, we fit Cox proportional hazards (Cox PH) models to patient groups dichotomized by both median protein and median mRNA abundance. Hazard ratios (HRs) from protein abundances were weakly correlated to those from mRNA abundances for the same genes (Spearman's ρ = 0.25; Figure 5A). Thus, some individual genes were associated with aggressive disease at the RNA level, others at the protein level, and a subset of 53 at both. Proteins exhibited a wider dynamic range of HRs (range, 0.22–4.23) than mRNAs (range, 0.33–2.73).

In some cases, mRNA and protein abundances showed divergent associations with patient outcome. For example, increased abundance of *PUS1*, a gene not previously implicated in cancer, was associated with increased risk of BCR, but unexpectedly, increased mRNA abundance was associated with a reduced risk (Figure 5B). In total, six genes showed divergent mRNA-protein associations with patient outcome, which may represent complex regulatory loops, translational dysregulation, post-translational modifications, or post-transcriptional processes that participate in driving aggressive disease. For validation, we focused on the 53 genes whose mRNA and protein abundances were both associated with disease aggressivity. We first considered *ACAD8* (Figure 5C) as it has high protein abundance (3^{rd} decile) and has not been strongly linked to prostate cancer previously. We validated the association of low *ACAD8* protein abundance with poor outcome using immunohistochemistry on a tissue microarray of 73 intermediate-risk prostate tumors (Figures 5D and S5A). Validating these candidate prognostic markers in larger cohorts will be key.

Multi-omic Integration Improves Prediction of Patient Outcome

Clinically used biomarkers are derived from many different classes of biomolecules, with DNA- and RNA-based assays being particularly prominent in prostate cancer (Fraser et al., 2015). It is unknown whether a particular class of biomolecules is generally superior for a given biomarker question. To quantitatively address this question, we again focused on prediction of BCR. We performed a null distribution (information content) analysis, generating 10 million gene sets, each comprising 100 genes randomly selected without replacement (Boutros et al., 2009; Lalonde et al.,

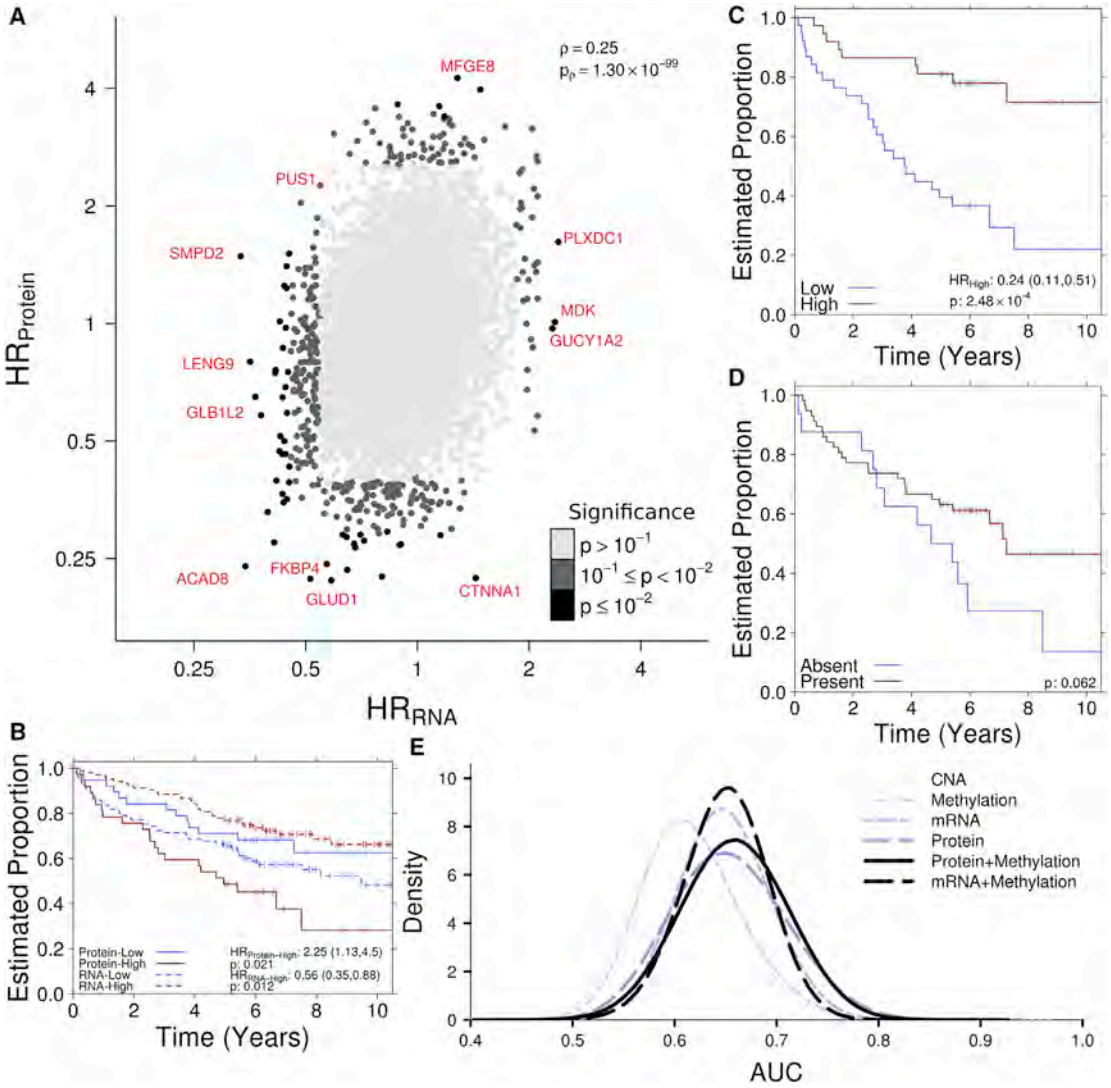


Figure 5. Protein Abundance Robustly Predicts Patient Survival

(A) Hazard ratios were calculated using a Cox model on patient groups determined using median-dichotomized protein and RNA abundances. Shading of dots indicates statistical significance with selected genes labeled in red.

(B) Kaplan-Meier (KM) plot for PUS1 protein (solid lines) and mRNA (dashed lines). A Cox model was fit with patients stratified into high and low abundance of PUS1 protein (75 patients) and mRNA (209 patients).

(C) KM plot showing 10-year biochemical recurrence-free survival of patient groups as dichotomized by high and low protein abundance of ACAD8.

(D) KM plot for ACAD8 in 73 tissue microarrays. Three slides were evaluated per sample, and patients were grouped into low protein abundance if at least two slides reported heterogeneous or faint staining. Significance of association was calculated using a log rank test between high- and low-abundance patient groups.

(E) Null distribution of predictive accuracy for different biomolecules obtained from 10 million replicates of 100 randomly selected genes. For each replicate, a value for the area under the receiver-operator curve (AUC) was calculated using classification results from 4-fold cross-validation in random forest.

See also Figure S5.

2014). This gene-set size was chosen to match that of several validated prognostic biomarkers (Lalonde et al., 2014). For each gene set, we used supervised machine learning (random forests) to train and validate CNA, methylation, RNA, and protein biomarkers, resulting in 40 million trained biomarkers. For each biomarker, we assessed their accuracy via the area under the receiver operating characteristic curve (AUC).

The resulting null distributions showed that random biomarkers generated from CNA or methylation data had similar performance (Figure 5E; blue curves, median AUC = 0.60). By contrast, biomarkers generated from mRNA and protein abundances were significantly superior, improving mean AUC by 0.03 for mRNA and 0.04 for protein ($p < 2.2 \times 10^{-16}$; t test). These results were independent of gene-set size (range, 5–100; Figure S5B). Thus, proteomic features are significantly more informative for BCR prediction than genomic, epigenomic, or transcriptomic ones.

These data provided us with a unique opportunity to consider the synergy of constructing biomarkers from multiple distinct

data types. Pairwise comparison of matched mRNA and protein biomarkers show a classic long-tail distribution, suggestive of potential synergy (Figure S5C). To test this explicitly, we created biomarkers from pairwise combinations of biomolecules and evaluated them using the AUC (Figure S5D). Pairs of data types produced biomarkers significantly better than genomic features only. For example, methylation-protein biomarkers were on average the best combination, with a median AUC of 0.66 (Figure 5E), reflecting their low MI (Figure 4A). Thus, low MI among different biomolecules may facilitate explicit complexity-accuracy trade-offs in the construction of multi-modal biomarkers, but these are incremental in magnitude and require further validation in the future.

DISCUSSION

Cancer is a disease of the genome. The accumulation of point and structural mutations, along with epigenomic changes, which dysregulate the transcriptome, leads to a dysregulated proteome and ultimately to the hallmark phenotypes of cancer cells (Hanahan and Weinberg, 2011). Proteins are the most abundant class of functional molecules in the cell, and the central dogma guides information flow from the genome and epigenome to the proteome. We therefore created a unique cohort of the most commonly treated form of prostate cancer: localized, intermediate-risk disease. Our compendium of whole-genome, epigenome, transcriptome, and proteome profiles reveals patterns of information exchange across levels of the central dogma whose biological implications are uncertain, but intriguing.

Previous studies in other tumor types have shown low correlations between the abundances of specific RNA transcripts and the abundances of the resulting proteoform (Mertins et al., 2016; Zhang et al., 2014, 2016). We confirm that the weak transcriptome-proteome relationships in other tumor types (Shao et al., 2017) also exist in prostate cancer and generalize this to a broad range of genomic and epigenomic features, including *cis*-regulatory elements. These weak correlations are reflected in a large network of *trans* effects across data types, which differentially affect RNA and protein abundances and are correlated to specific functional sets of genes. These networks may provide an avenue for understanding the influences of specific genomic features on specific aspects of the proteome, and subsequently on downstream pathways and cellular and clinical phenotypes.

The proteomic subtypes of prostate tumors are only weakly related to their genomic ones. This suggests strong post-transcriptional regulatory mechanisms that are not easily detected in genomic data. This observation is mirrored by the differences in specific genes and pathways associated with ETS gene fusion status at the RNA and protein levels. The proteomic characteristics of ETS-positive tumors indicate an extensive dysregulation of their metabolic profile (Bose et al., 2017), which is not reflected in the transcriptional changes seen. Indeed, these data highlight the drawbacks of studies that implicitly infer changes in functional protein directly from mRNA abundance data. The assumption that transcriptional profiles are a reliable surrogate for proteomic ones is incorrect for most genes. Only ~10% of variation in protein abundances is explained by changes in the transcriptome, suggesting an urgent need for statistical models that better predict protein abundances from more readily available nucleotide-based data.

The clinical potential of proteomic biomarkers is high; proteins harbored more information on patient relapse than any other data type, and multi-modal biomarkers consistently outperformed those generated from individual data types. Yet multi-modal biomarkers are inherently complex, highlighting the need for improved technologies to accurately profile multiple analytes from individual tumors, especially considering that high-throughput proteomics clinical tests have not been implemented in routine practice to date. Larger patient cohorts and complementary validation studies will be key to reaching the translational potential of multi-modal data. This suggests an opportunity for expansion of existing cancer genomics consortia to pair their high-quality, deeply analyzed genomes and transcriptomes to unbiased proteomic surveys.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Selection of Patient Cohort and Tumor Sections
- METHOD DETAILS
  - Tumor Tissue Preparation for Shotgun Proteomics
  - Shotgun Proteomics
  - Proteomic Data Batch Correction and Missing Value Imputation
  - DNA and RNA-Sequencing
  - mRNA Microarray Data Generation
  - SNP Microarray Data Generation and CNA Calling
  - Somatic Variant Calling
  - Methylation Microarray Data Generation
  - Epigenetic Data Annotation
  - Consensus Clustering of Proteomic Data
  - Clinical Associations and ETS Analysis
  - ETS Fusion Associated Gene Intersection
  - Pathway Enrichment Analysis
  - Correlation Analysis between mRNA and Protein Abundances
  - Identification of Proteins with Undetected Transcripts and *vice versa*
  - Association Analyses of CNAs on mRNA and Protein Abundances
  - Mutual Information Analysis
  - Percent Variance Analysis
  - Univariate Survival Analysis
  - Biomarker Null Distribution Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.ccell.2019.02.005.

ACKNOWLEDGMENTS

The authors thank all members of the Boutros and Kislinger labs for helpful suggestions. P.C.B. and T.K. were funded through a Movember Discovery grant from Prostate Cancer Canada (D2013-21), an operating grant from the National Cancer Institute Early Detection Research Network (1U01CA214194-01), a CIHR Project Grant (PJT 156357), and a Collaborative Personalized Cancer Medicine Team Grant from the Princess Margaret Cancer Centre. P.C.B. was supported by a Terry Fox Research Institute New Investigator Award, a CIHR New Investigator Award and a CIHR Project Grant. A.S. was supported by a CIHR Banting and Best doctoral research scholarship. This research was funded in part by the Ontario Ministry of Health and Long Term Care. This work was supported by the Discovery Frontiers: Advancing Big Data Science in Genomics Research program, which is jointly funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) of Canada, the Canadian Institutes of Health Research (CIHR), Genome Canada, and the Canada Foundation for Innovation (CFI). This work was supported by Prostate Cancer Canada and is proudly funded by the Movember Foundation (grant no. RS2014-01). This work was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-125). This research is funded by the Canadian Cancer Society (grant no. 705649).

AUTHOR CONTRIBUTIONS

Sample preparation and LC-MS analyses: A.S., K.F., M.F., K.K., and A.M. Pathology analyses: M.M., B.T., and T.v.d.K. Statistical and bioinformatics analyses: A.S., V.H., J.L., J.W., N.S.F., N.K., V.I., N.D., L.E.H., Y.J.S., C.Q.Y., J.A.A., S.V., and A.L. Wrote the first draft of the manuscript: V.H., J.L., A.S., T.K., and P.C.B. Initiated the project: T.v.d.K., R.G.B., T.K., and P.C.B. Supervised research: C.S., C.C.C., D.M.B., M.L., T.v.d.K., R.G.H., T.K., and P.C.B. Approved the manuscript: all authors.

DECLARATION OF INTERESTS

All authors declare that they have no conflicts of interest.

Received: August 28, 2018
Revised: November 26, 2018
Accepted: February 14, 2019
Published: March 18, 2019

REFERENCES

Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369.

Baca, S.C., Prandi, D., Lawrence, M.S., Mosquera, J.M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T.Y., Ghandi, M., et al. (2013). Punctuated evolution of prostate cancer genomes. Cell *153*, 666–677.

Bhandari, V., Hoey, C., Liu, L.Y., Lalonde, E., Ray, J., Livingstone, J., Lesurf, R., Shiah, Y.-J., Vujcic, T., Huang, X., et al. (2019). Molecular landmarks of tumor hypoxia across cancer types. Nat. Genet. *51*, 308–318.

Blume-Jensen, P., Berman, D.M., Rimm, D.L., Shipitsin, M., Putzi, M., Nifong, T.P., Small, C., Choudhury, S., Capela, T., Coupal, L., et al. (2015). Development and clinical validation of an in situ biopsy-based multimarker assay for risk stratification in prostate cancer. Clin. Cancer Res. *21*, 2591–2600.

Bose, R., Karthaus, W.R., Armenia, J., Abida, W., Iaquinta, P.J., Zhang, Z., Wongvipat, J., Wasmuth, E.V., Shah, N., Sullivan, P.S., et al. (2017). ERF mutations reveal a balance of ETS factors controlling prostate oncogenesis. Nature *546*, 671–675.

Boutros, P.C., Lau, S.K., Pintilie, M., Liu, N., Shepherd, F.A., Der, S.D., Tsao, M.-S., Penn, L.Z., and Jurisica, I. (2009). Prognostic gene signatures for non-small-cell lung cancer. Proc. Natl. Acad. Sci. U S A *106*, 2824–2828.

Boutros, P.C., Fraser, M., Harding, N.J., de Borja, R., Trudel, D., Lalonde, E., Meng, A., Hennings-Yeomans, P.H., McPherson, A., Sabelnykova, V.Y., et al. (2015). Spatial genomic heterogeneity within localized, multifocal prostate cancer. Nat. Genet. *47*, 736–745.

Broad Institute TCGA Genome Data Analysis Center (2016). Correlation between mRNA Expression and DNA Methylation (Broad Institute MIT Harvard).

Brocks, D., Assenov, Y., Minner, S., Bogatyrova, O., Simon, R., Koop, C., Oakes, C., Zucknick, M., Lipka, D., Weischenfeldt, J., et al. (2014). Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. Cell Rep. *8*, 798–806.

Canadian Cancer Statistics Advisory Committee (2017). Canadian cancer statistics. http://www.cancer.ca/~/media/cancer.ca/CW/cancer%20information/cancer%20101/Canadian%20cancer%20statistics/Canadian-Cancer-Statistics-2018-EN.pdf?la=en.

Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. Bioinformatics *26*, 2363–2367.

Center, M.M., Jemal, A., Lortet-Tieulent, J., Ward, E., Ferlay, J., Brawley, O., and Bray, F. (2012). International variation in prostate cancer incidence and mortality rates. Eur. Urol. *61*, 1079–1092.

Chen, H., and Boutros, P.C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics *12*, 35.

Chua, M.L.K., Lo, W., Pintilie, M., Murgic, J., Lalonde, E., Bhandari, V., Mahamud, O., Gopalan, A., Kweldam, C.F., van Leenders, G.J.L.H., et al. (2017). A prostate cancer "Nimbosus": genomic instability and SChLAP1 dysregulation underpin aggression of intraductal and cribriform subpathologies. Eur. Urol. *71*, 183–192.

Cooper, C.S., Eeles, R., Wedge, D.C., Van Loo, P., Gundem, G., Alexandrov, L.B., Kremeyer, B., Butler, A., Lynch, A.G., Camacho, N., et al. (2015). Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. Nat. Genet. *47*, 367–372.

Cooperberg, M.R., Broering, J.M., and Carroll, P.R. (2009). Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. J. Natl. Cancer Inst. *101*, 878–887.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol. Cell. Proteomics *13*, 2513–2526.

Cuzick, J., Swanson, G.P., Fisher, G., Brothman, A.R., Berney, D.M., Reid, J.E., Mesher, D., Speights, V., Stankiewicz, E., Foster, C.S., et al. (2011). Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. Lancet Oncol. *12*, 245–255.

Dal Pra, A., Lalonde, E., Sykes, J., Warde, F., Ishkanian, A., Meng, A., Maloff, C., Srigley, J., Joshua, A.M., Petrovics, G., et al. (2013). TMPRSS2-ERG status is not prognostic following prostate cancer radiotherapy: implications for fusion status and DSB repair. Clin. Cancer Res. *19*, 5202–5209.

Den, R.B., Yousefi, K., Trabulsi, E.J., Abdollah, F., Choeurng, V., Feng, F.Y., Dicker, A.P., Lallas, C.D., Gomella, L.G., Davicioni, E., et al. (2015). Genomic classifier identifies men with adverse pathology after radical prostatectomy who benefit from adjuvant radiation therapy. J. Clin. Oncol. *33*, 944–951.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Drake, J.M., Paull, E.O., Graham, N.A., Lee, J.K., Smith, B.A., Titz, B., Stoyanova, T., Faltermeier, C.M., Uzunangelov, V., Carlin, D.E., et al. (2016). Phosphoproteome integration reveals patient-specific networks in prostate cancer. Cell *166*, 1041–1054.

Espiritu, S.M.G., Liu, L.Y., Rubanova, Y., Bhandari, V., Holgersen, E.M., Szyca, L.M., Fox, N.S., Chua, M.L.K., Yamaguchi, T.N., Heisler, L.E., et al. (2018). The

evolutionary landscape of localized prostate cancers drives clinical aggression. Cell *173*, 1003–1013.e15.

Fraser, M., Berlin, A., Bristow, R.G., and van der Kwast, T. (2015). Genomic, pathological, and clinical heterogeneity as drivers of personalized medicine in prostate cancer. Urol. Oncol. *33*, 85–94.

Fraser, M., Sabelnykova, V.Y., Yamaguchi, T.N., Heisler, L.E., Livingstone, J., Huang, V., Shiah, Y.-J., Yousif, F., Lin, X., Masella, A.P., et al. (2017). Genomic hallmarks of localized, non-indolent prostate cancer. Nature *541*, 359–364.

Govind, S.K., Zia, A., Hennings-Yeomans, P.H., Watson, J.D., Fraser, M., Anghel, C., Wyatt, A.W., van der Kwast, T., Collins, C.C., McPherson, J.D., et al. (2014). ShatterProof: operational detection and quantification of chromothripsis. BMC Bioinformatics *15*, 78.

Gundem, G., Van Loo, P., Kremeyer, B., Alexandrov, L.B., Tubio, J.M.C., Papaemmanuil, E., Brewer, D.S., Kallio, H.M.L., Högnäs, G., Annala, M., et al. (2015). The evolutionary history of lethal metastatic prostate cancer. Nature *520*, 353–357.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hansen, A.F., Sandsmark, E., Rye, M.B., Wright, A.J., Bertilsson, H., Richardsen, E., Viset, T., Bofin, A.M., Angelsen, A., Selnæs, K.M., et al. (2016). Presence of TMPRSS2-ERG is associated with alterations of the metabolic profile in human prostate cancer. Oncotarget *7*, 42071–42085.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for the ENCODE project. Genome Res. *22*, 1760–1774.

Hausser, J., and Strimmer, K. (2008). Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. ArXiv.

Hieronymus, H., Schultz, N., Gopalan, A., Carver, B.S., Chang, M.T., Xiao, Y., Heguy, A., Huberman, K., Bernstein, M., Assel, M., et al. (2014). Copy number alteration burden predicts prostate cancer relapse. Proc. Natl. Acad. Sci. U S A *111*, 11139–11144.

Hopkins, J.F.J., Sabelnykova, V.Y.V., Weischenfeldt, J., Simon, R., Aguiar, J.A.J., Alkallas, R., Heisler, L.E., Zhang, J., Watson, J.D.J., Chua, M.L.K.M., et al. (2017). Mitochondrial mutations drive prostate cancer aggression. Nat. Commun. *8*, 656.

Huang, B.F.F., and Boutros, P.C. (2016). The parameter sensitivity of random forests. BMC Bioinformatics *17*, 331.

Iglesias-Gato, D., Wikström, P., Tyanova, S., Lavallee, C., Thysell, E., Carlsson, J., Hägglöf, C., Cox, J., Andrén, O., Stattin, P., et al. (2016). The proteome of primary prostate cancer. Eur. Urol. *69*, 942–952.

Iglesias-Gato, D., Thysell, E., Tyanova, S., Crnalic, S., Santos, A., Lima, T.S., Geiger, T., Cox, J., Widmark, A., Bergh, A., et al. (2018). The proteome of prostate cancer bone metastasis reveals heterogeneity with prognostic implications. Clin. Cancer Res. *24*, 5433–5444.

Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127.

Kislinger, T., Cox, B., Kannan, A., Chung, C., Hu, P., Ignatchenko, A., Scott, M.S., Gramolini, A.O., Morris, Q., Hallett, M.T., et al. (2006). Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. Cell *125*, 173–186.

Klein, E.A., Cooperberg, M.R., Magi-Galluzzi, C., Simko, J.P., Falzarano, S.M., Maddala, T., Chan, J.M., Li, J., Cowan, J.E., Tsiatis, A.C., et al. (2014). A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. Eur. Urol. *66*, 550–560.

Kron, K.J., Murison, A., Zhou, S., Huang, V., Yamaguchi, T.N., Shiah, Y.-J., Fraser, M., van der Kwast, T., Boutros, P.C., Bristow, R.G., et al. (2017). TMPRSS2–ERG fusion co-opts master transcription factors and activates NOTCH signaling in primary prostate cancer. Nat. Genet. *49*, 1336–1345.

Lalonde, E., Ishkanian, A.S., Sykes, J., Fraser, M., Ross-Adams, H., Erho, N., Dunning, M.J., Halim, S., Lamb, A.D., Moon, N.C., et al. (2014). Tumour genomic and microenvironmental heterogeneity for integrated prediction of 5-year biochemical recurrence of prostate cancer: a retrospective cohort study. Lancet Oncol. *15*, 1521–1532.

Larson, D.E., Harris, C.C., Chen, K., Koboldt, D.C., Abbott, T.E., Dooling, D.J., Ley, T.J., Mardis, E.R., Wilson, R.K., and Ding, L. (2012). SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics *28*, 311–317.

Latonen, L., Afyounian, E., Jylhä, A., Nättinen, J., Aapola, U., Annala, M., Kivinummi, K.K., Tammela, T.T.L., Beuerman, R.W., Uusitalo, H., et al. (2018). Integrative proteomics in prostate cancer uncovers robustness against genomic and transcriptomic aberrations during disease progression. Nat. Commun. *9*, 1176.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. Cell Syst. *1*, 417–425.

Liu, H., Sadygov, R.G., and Yates, J.R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal. Chem. *76*, 4193–4201.

Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One *5*, e13984.

Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature *534*, 55–62.

Michalski, A., Damoc, E., Hauschild, J.-P., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. (2011). Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. Mol. Cell. Proteomics *10*, M111.011015.

Minner, S., Enodien, M., Sirma, H., Luebke, A.M., Krohn, A., Mayer, P.S., Simon, R., Tennstedt, P., Müller, J., Scholz, L., et al. (2011). ERG status is unrelated to PSA recurrence in radically operated prostate cancer in the absence of antihormonal therapy. Clin. Cancer Res. *17*, 5878–5888.

P'ng, C., Green, J., Chong, L.C., Waggott, D., Prokopec, S.D., Shamsi, M., Nguyen, F., Mak, D.Y.F., Lam, F., Albuquerque, M.A., et al. (2019). BPG: seamless, automated and interactive visualization of scientific data. BMC Bioinformatics *20*, 42.

Pidsley, R., Y Wong, C.C., Volta, M., Lunnon, K., Mill, J., and Schalkwyk, L.C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics *14*, 293.

Pritchard, C.C., Mateo, J., Walsh, M.F., De Sarkar, N., Abida, W., Beltran, H., Garofalo, A., Gulati, R., Carreira, S., Eeles, R., et al. (2016). Inherited DNA-repair gene mutations in men with metastatic prostate cancer. N. Engl. J. Med. *375*, 443–453.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339.

Reimand, J., Arak, T., and Vilo, J. (2011). G: profiler - A web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res. *39*, 307–315.

Ribeiro, F.R., Paulo, P., Costa, V.L., Barros-Silva, J.D., Ramalho-Carvalho, J., Jerónimo, C., Henrique, R., Lind, G.E., Skotheim, R.I., Lothe, R.A., et al. (2011). Cysteine-rich secretory protein-3 (CRISP3) is strongly up-regulated in prostate carcinomas with the TMPRSS2-ERG fusion gene. PLoS One *6*, e22317.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Sallari, R.C., Sinnott-Armstrong, N.A., French, J.D., Kron, K.J., Ho, J., Moore, J.H., Stambolic, V., Edwards, S.L., Lupien, M., and Kellis, M. (2016). Convergence of dispersed regulatory mutations predicts driver genes in prostate cancer. BioRxiv. https://doi.org/10.1101/097451.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Shao, W., Guo, T., Toussaint, N.C., Wagner, U., Li, L., Charmpi, K., Zhu, Y., Beyer, A., Ratsch, G., Wild, P., et al. (2017). Comparative analysis of mRNA degradation and protein degradation in 68 pairs of adjacent prostate tumor tissue samples. In Proceedings of the 65th ASMS Conference on Mass Spectrometry and Allied Topics, (Indianapolis, IN), p. MP 114.

Sinha, A., Ignatchenko, V., Ignatchenko, A., Mejia-Guerrero, S., and Kislinger, T. (2014). In-depth proteomic analyses of ovarian cancer cell line exosomes reveals differential enrichment of functional categories compared to the NCI 60 proteome. Biochem. Biophys. Res. Commun. *445*, 694–701.

Song, S., Nones, K., Miller, D., Harliwong, I., Kassahn, K.S., Pinese, M., Pajic, M., Gill, A.J., Johns, A.L., Anderson, M., et al. (2012). qpure: a tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. PLoS One *7*, e45835.

Stelloo, S., Nevedomskaya, E., van der Poel, H.G., de Jong, J., van Leenders, G.J.L.H., Jenster, G., Wessels, L.F.A., Bergman, A.M., and Zwart, W. (2015). Androgen receptor profiling predicts prostate cancer outcome. EMBO Mol. Med. *7*, 1450–1464.

Svetnik, V., Liaw, A., Tong, C., Christopher Culberson, J., Sheridan, R.P., and Feuston, B.P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. J. Chem. Inf. Comput. Sci. *43*, 1947–1958.

The Cancer Genome Atlas Research Network. (2015). The molecular taxonomy of primary prostate cancer. Cell *163*, 1011–1025.

Torre, L.A., Siegel, R.L., Ward, E.M., and Jemal, A. (2016). Global cancer incidence and mortality rates and trends—an update. Cancer Epidemiol. Biomarkers. Prev. *25*, 16–27.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

Woo, Y.H., and Li, W.-H. (2012). Evolutionary conservation of histone modifications in mammals. Mol. Biol. Evol. *29*, 1757–1767.

Wu, X., Daniels, G., Lee, P., and Monaco, M.E. (2014). Lipid metabolism in prostate cancer. Am. J. Clin. Exp. Urol. *2*, 111–120.

Yousif, F., Prokopec, S., Sun, R.X., Fan, F., Lalansingh, C.M., Park, D.H., Szyca, L., Network, P., and Boutros, P.C. (2018). The origins and consequences of localized and global somatic hypermutation. BioRxiv. https://doi.org/10.1101/287839.

Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383.

Zhang, B., Wang, J., Wang, X., Zhu, J., Liu, Q., Shi, Z., Chambers, M.C., Zimmerman, L.J., Shaddox, K.F., Kim, S., et al. (2014). Proteogenomic characterization of human colon and rectal cancer. Nature *513*, 382–387.

Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.-Y., Petyuk, V.A., Chen, L., Ray, D., et al. (2016). Integrated proteogenomic characterization of human high-grade serous ovarian cancer. Cell *166*, 755–765.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-ACAD8	Sigma-Aldrich	Cat#HPA040689; RRID: AB_2677080; Lot# A114184
Deposited Data		
Mass Spectrometry Data - CPC-Gene	This study	ftp://massive.ucsd.edu/MSV000081552; MassIVE: MSV000081552
WGS and RNA-seq Data - CPC-Gene	Fraser et al., 2017, Espiritu et al., 2018	EGA: EGAS00001000900
RNA Microarray Data - CPC-Gene	This study	GEO: GSE107299
H3K27Ac Data - CPC-Gene	Kron et al., 2017	GEO: GSE96652
Methylation - CPC-Gene	Fraser et al., 2017, This study	GEO: GSE107298
Processed CNA, mRNA and Methylation - TCGA PRAD	TCGA, 2015	https://portal.gdc.cancer.gov/projects/TCGA-PRAD
Gencode Annotations (v24lift37)	Gencode	https://www.gencodegenes.org/human/
Software and Algorithms		
Cytoscape (v3.6.1)	Shannon et al., 2003	https://cytoscape.org
MaxQuant (v1.6.1.0)	Cox and Mann, 2008	https://www.biochem.mpg.de/5111795/maxquant
ComBat (V3.20.0)	Johnson et al., 2007	https://www.bu.edu/jlab/wp-assets/ComBat/Abstract.html
Limma (v3.28.21)	Ritchie et al., 2015	https://doi.org/doi:10.18129/B9.bioc.limma
sva (v3.20.0)	Leek et al., 2012	https://doi.org/doi:10.18129/B9.bioc.sva
STAR aligner (v2.5.3a)	Dobin et al., 2013	https://github.com/alexdobin/STAR
RSEM (v1.2.29)	Li and Dewey, 2011	https://deweylab.github.io/RSEM/
oligo (v1.32.0)	Carvalho and Irizarry, 2010	https://doi.org/doi:10.18129/B9.bioc.oligo
Nexus Express	ThermoFisher	https://www.thermofisher.com/ca/en/home/life-science/microarray-analysis/microarray-analysis-instruments-software-services/microarray-analysis-software/nexus-express-software-for-oncoscan-ffpe-assay-kit.html
BEDTools (v2.17.0)	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
BWA (v0.5.7)	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
SomaticSniper (v1.0.2)	Larson et al., 2012	http://gmt.genome.wustl.edu/packages/somatic-sniper/
ANNOVAR (v2015-06-17)	Wang et al., 2010	http://annovar.openbioinformatics.org/en/latest/
Delly (v0.5.5)	Rausch et al., 2012	https://github.com/dellytools/delly
SeqKat (v0.0.1)	Yousif et al., 2018	https://CRAN.R-project.org/package=SeqKat
Shatterproof (v0.14)	Govind et al., 2014	https://metacpan.org/pod/Shatterproof
wateRmelon (v1.15.1)	Pidsley et al., 2013	https://doi.org/doi:10.18129/B9.bioc.wateRmelon
minifi (v1.22.1)	Aryee et al., 2014	https://doi.org/doi:10.18129/B9.bioc.minfi
ChIPseeker (v1.12.1)	Yu et al., 2015	https://doi.org/doi:10.18129/B9.bioc.ChIPseeker
Survival (v2.40-3)	NA	https://CRAN.R-project.org/package=survival
ConsensusClusterPlus (v1.38.0)	Wilkerson and Hayes, 2010	https://doi.org/doi:10.18129/B9.bioc.ConsensusClusterPlus
VennDiagram (v1.6.19)	Chen and Boutros, 2011	https://CRAN.R-project.org/package=VennDiagram
BoutrosLab.plotting.general (v5.9.2)	P'ng et al., 2019	https://cran.r-project.org/package=BoutrosLab.plotting.general
g:Profiler (v r1732_e88_eg35)	Reimand et al., 2011	https://biit.cs.ut.ee/gprofiler/
Entropy (v1.2.1)	Hausser and Strimmer, 2008	https://CRAN.R-project.org/package=entropy
randomForest (v4.6-12)	Svetnik et al., 2003	https://cran.r-project.org/package=randomForest

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Paul C. Boutros (pboutros@mednet.ucla.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Selection of Patient Cohort and Tumor Sections

Patient selection, tissue collection and sample processing was performed as previously described (Fraser et al., 2017). Informed consent, consistent with local Research Ethics Board (REB) and International Cancer Genome Consortium (ICGC) guidelines, was obtained at the time of clinical follow-up. Previously collected tumor tissues were used, following University Health Network REB-approved study protocols (UHN 06-0822-CE, UHN 11-0024-CE, CHUQ 2012-913:H12-03-192). All patients were treated surgically via radical prostatectomy (RadP). Primary treatment failure was defined as PSA levels at or above 0.2 ng/mL three months after surgery; no patients in this cohort experienced primary treatment failure. Biochemical recurrence (BCR) after RadP was defined as two consecutive measurements of PSA > 0.2 ng/mL or the administration of salvage radiotherapy. All patients were National Comprehensive Cancer Network (NCCN) intermediate-risk based on pre-surgical parameters. ISUP Scores and tumor cellularity were evaluated by two genitourinary pathologists (T.v.d.K., and B.T.) on scanned haematoxylin- and eosin-stained slides as described previously (Fraser et al., 2017). Adjacent 10 serial sections (10 μm thickness of each section) from each patient tumor were used for acquiring each of the multi-omic dataset. All sections were pathologically inspected by our co-author (T.v.d.K.) and tumor tissue was macro-dissected to reach ~70% cellularity. As a validation, cellularity was calculated *in silico* from OncoScan array data using qpure (v1.1) (Song et al., 2012). Additionally, for all multi-omic analyses, we analyzed the index lesion - the lesion that led to the initial diagnosis and treatment of the patient. Given our focus on biomarkers, this allowed a consistent analysis that avoids conflating information available post-operatively with that available during initial staging and management.

METHOD DETAILS

Tumor Tissue Preparation for Shotgun Proteomics

Fresh-frozen RadP specimens were obtained from the University Health Network (UHN) Pathology BioBank or from the Genito-Urinary BioBank of the Centre Hospitalier Universitaire de Québec (CHUQ). Ten pathologically inspected, optimal cutting temperature compound (OCT) embedded tissue sections (10 μm each) were processed for each prostate cancer sample (i.e. sections are in close proximity to previously used sections for genomics, epigenomics and transcriptomics, but not identical). Tissues were scraped from the glass slides and transferred to a 1.5 mL conical tube. Removal of the OCT compound was performed using various dilutions of ethanol with water as follows. Initially, 1 mL of 70% (v/v) ethanol was added to each conical tube, followed by 30 s of vortex at high speed. The tubes were centrifuged at 14,000-rcf for 3 min, and the supernatant was discarded. Subsequently, tissue pellets were treated with 100% ethanol, 70% ethanol, 85% ethanol and lastly in 100% ethanol with an additional 5 min incubation at room temperature and rigorous high-speed vortexing. The obtained tissue pellet was then processed for shotgun proteomics.

Shotgun Proteomics

Tissue pellets were resuspended in 100 μL of 50% (v/v) 2,2,2-Trifluoroethanol in phosphate buffered saline (pH 7.4) and incubated for one hr at 60°C. Subsequently disulphide bonds were reduced by incubation for 30 min at 60°C with 5-mM dithiothreitol. Afterward, alkylation of reduced disulphide-bridges was performed using 25 mM iodoacetamide for 30 min at room temperature in the dark. Samples were diluted 1:5 using 100 mM ammonium bicarbonate with 2 mM CaCl (pH 8.0). Proteins were digested with 5 μg of trypsin at 37°C overnight. Peptides were desalted using C18-based solid phase extraction. Subsequently, eluted peptides were lyophilized and solubilized in mass spectrometry-grade water with 0.1% formic acid. Peptide concentration was quantified using a NanoDrop Lite (at 280 nm) and a constant aliquot of 2 μg of peptides were injected onto the column for chromatography and proteomics analysis. LC-MS/MS data was acquired using an Easy nLC 1000 (Thermo) nano-flow liquid chromatography system with a 50 cm EasySpray (Thermo) column coupled to a Q Exactive (Thermo) tandem mass spectrometer. Data was acquired using a four hr chromatographic gradient with the mass spectrometer operating in data dependent mode. MS^1 data was acquired at resolution of 70,000, while MS^2 data was acquired at resolution of 17,500 with a top 15 method (Michalski et al., 2011; Sinha et al., 2014). The acquired data was searched using MaxQuant (v1.6.1.0) (Cox and Mann, 2008) and a UniProt protein sequences database (complete human proteome; v1-27-2015, number of sequences 42,041). Searches were performed with a maximum of two missed cleavages, cabamidomethylation of cysteine as fixed modification and oxidation of methionine as variable modification. False discovery rate (FDR) was set to 1% for peptide spectral matches and protein identification using a target-decoy strategy (Kislinger et al., 2006). The ProteinGroup.txt file was used for all subsequent analysis. Proteins identified with two or more peptides were carried forward. Relative quantification was performed using MS^1 signal intensity for label-free quantification, following a standard MaxQuant analysis strategy MaxLFQ (Cox et al., 2014), which uses an aggregate of all MS^1 peptide intensities of a reported protein.

Proteomic Data Batch Correction and Missing Value Imputation

Four batch correction methods were evaluated: ComBat (v3.20.0) (Johnson et al., 2007), limma (v3.28.21) (Ritchie et al., 2015) and the removal of one and two surrogate variables using sva (v3.20.0) (Leek et al., 2012). ComBat batch correction was performed using the null model. Correction with limma was performed using the removeBatchEffect command with no additional covariates included. For surrogate variable analysis, biochemical recurrence was used as the endpoint of the model to preserve associated variance. The number of nuisance variables was automatically estimated using the num.sv command, and a data matrix was regenerated following removal of one or two nuisance variables. Metrics used to evaluate the correction methods included examining the variance of *GAPDH*, *SDHA* and *GPI* post correction, fitting a linear model between batch and the protein abundance for each gene as the response and calculating the 90th percentile of percent variance explained by the batch term, and by calculating Spearman's correlation between a duplicated sample in two of the batches. We ranked each method based on these criteria, calculated the rank product and arrived at ComBat as the highest ranked batch correction method for this dataset. Missing values were imputed from the lower half of a Gaussian distribution around a mean of the protein intensities from the 0.01th percentile of all protein intensities. The imputation of missing values can potentially lead to an overestimation for binary comparisons (i.e. ratios).

DNA and RNA-Sequencing

Whole genome sequencing of DNA was performed as previously described (Fraser et al., 2017). RNA samples were sent to BGI Americas and underwent QC and DNAse treatment. For each sample, 200 ng of total RNA was used to construct a TruSeq strand specific library with the Ribo-Zero protocol (Illumina), and all samples were sequenced on a HiSeq2000v3 to a minimal target of 180 million paired-end reads. Reads were mapped using the STAR aligner (v2.5.3a) (Dobin et al., 2013) to GRCh37 with Gencode v24lift37 (Harrow et al., 2012). RSEM (v1.2.29) was used to quantify gene abundance (Li and Dewey, 2011).

mRNA Microarray Data Generation

Total RNA was extracted using the mirVana miRNA Isolation Kit (Life Technologies), according to the manufacturer's instructions and assayed on Affymetrix transcriptome arrays as previously described (Fraser et al., 2017). All mRNA analysis was performed using R (v3.2.1). Background correction, normalization algorithms and annotation were implemented in the oligo (v1.32.0) package (Carvalho and Irizarry, 2010) from the BioConductor (v3.0) open-source project. The Robust multichip average (RMA) algorithm was applied to the raw intensity data. Annotations were performed using hugene20sttranscriptcluster.db (v2.13.0) and hta20sttranscriptcluster.db (v8.3.1). The sva package (v3.14.0) was used to correct for batch effects between different arrays. Annotated data from HuGene 2.0 ST and HTA 2.0 were combined into one data set based on Entrez Gene IDs. The mRNA abundance values were averaged amongst duplicated Entrez Gene IDs.

SNP Microarray Data Generation and CNA Calling

SNP microarrays were performed with 200 ng of DNA on Affymetrix OncoScan FFPE Express 3.0 arrays as previously described (Fraser et al., 2017). BioDiscovery's Nexus Express^{TM} for OncoScan 3 Software was used to call CNAs using the SNP-FASST2 algorithm with default parameters except that the minimum number of probes per segment was changed from 3 to 20. When necessary, samples were re-centred using the Nexus Express^{TM} software, choosing regions that showed diploid $\log_2$ ratios and B allele frequency profiles. Gene level CNAs for each patient were identified by overlapping copy number segments, with RefGene (2014-07-15) annotation, using BEDTools (v2.17.0) (Quinlan and Hall, 2010). To account for technical noise, a gene level CNV blacklist was created from matched normal blood samples. Genes were added to the blacklist if they were seen in at least 75% of normal samples and filtered from downstream analyses. Percent genome altered (PGA) was calculated for each sample by dividing the number of base pairs that were involved in all copy number segments by the total length of the genome.

Somatic Variant Calling

Single nucleotide variants (SNVs) and genomic rearrangements (GRs) were called using pipelines that have been described in detail elsewhere (Fraser et al., 2017). Briefly, lane-level WGS reads for blood normal and tumor samples were aligned against human reference build hg19 with BWA (v0.5.7) (Li and Durbin, 2009) before being merged. SNVs were called using SomaticSniper (v1.0.2) (Larson et al., 2012) and annotated using ANNOVAR (v2015-06-17) (Wang et al., 2010) with the RefGene database. Somatic GRs were called using Delly (v0.5.5) (Rausch et al., 2012) and filtered for mapping quality (>20) or pair-end evidence (>4 reads) before being filtered against their corresponding normal sample and a consolidated set of normal calls. Kataegis was called using the SeqKat (v0.0.1) (Yousif et al., 2018) R package. Chromothriptic regions were identified using Shatterproof (v0.14) (Govind et al., 2014) with default settings.

Methylation Microarray Data Generation

Illumina Infinium HumanMethylation 450k BeadChip kits were used to assess global methylation, using 500 ng of input genomic DNA, at McGill University and the Genome Quebec Innovation Centre (Montreal, QC). All samples used in this study (n = 72) were processed from fresh-frozen prostate cancer tissue and can be found on GEO under the accession GSE107298. Methylation pre-processing were performed in R statistical environment (v3.4.0). The IDAT files were loaded and converted to raw intensity values with the use of wateRmelon package (v1.15.1) (Pidsley et al., 2013). Quality control was conducted using the minfi package (v1.22.1) (Aryee et al., 2014). No outlier samples were detected. Raw methylation intensity levels were then pre-processed using Dasen

(Pidsley et al., 2013). Probe filtering was conducted after the normalization, as previously described (Fraser et al., 2017). Annotation to chromosome location, probe position, and gene symbol was conducted using the IlluminaHumanMethylation450kanno.ilmn12.hg19 package (v0.6.0).

Epigenetic Data Annotation

H3K27Ac peaks were annotated to the closest genes using the *annotatePeak* function from the ChIPseeker (v1.12.1) (Yu et al., 2015) R package. The *tssRegion* was set as: -5000 to +5000) as proximal promoters can be up to 5 kbp away from transcription start site (Woo and Li, 2012).

Rather than using the full complement of CpG methylation sites, the probes with the greatest negative correlation to their corresponding mRNA abundance from the TCGA prostate cancer methylation dataset (Broad Institute TCGA Genome Data Analysis Center, 2016) were used, irrespective of their proximity to the gene. If there were no correlated probes associated with the gene, but the gene had annotated probes, a probe with the greatest variance was selected with the following priority: proximity to transcription start site, 5' UTR, 3' UTR and gene body. If a probe was not annotated to the gene, we retrieved the closest probe within 10 kbp. Genes that were not assigned a probe were denoted as missing (NA). Gene names from each biomolecule type were intersected to identify genes present in all data types.

Consensus Clustering of Proteomic Data

Consensus clustering (*max_k* = 20; Spearman's ρ as the similarity metric; *pItem* = 0.8, *pFeature* = 0.8; *seed* = 17; *reps* = 1000; ConsensusClusterPlus v1.38.0) (Wilkerson and Hayes, 2010) was performed using a divisive algorithm on the 25% most variable proteins to cluster the patients and the proteins. Adjusted Rand Index (ARI) was calculated on patient classification using the protein subtypes and subtypes defined by copy number aberrations (Fraser et al., 2017) to determine if there is an overlap. Associations between patient subgroups and mutation burden were performed using a Mann-Whitney U test. Survival analysis was performed on the protein subtypes with the R package Survival (v2.40-3) by fitting a Cox PH model between patients in C2, C3, C4 and C5 against C1 as the baseline, which was the largest group with BCR as the end point. Proportional hazards assumptions were evaluated using the *cox.zph* function ($p < 0.1$). Androgen receptor signature scores were created by identifying the top 100 genes that are positively correlated to *AR* (Spearman's ρ), converting their abundances to z-score before taking the mean. For the signature from the literature, the abundances of the genes used in the signature were retrieved, converted to z-scores, and then averaged. ANOVA was used to test for an association between the subtypes and the scores.

Clinical Associations and ETS Analysis

To identify proteins that may be associated with clinical features, univariate association tests (Spearman's ρ for continuous values, Mann-Whitney U test for binary values) were performed with each protein group against the following clinical covariates: percent genome altered (PGA), ETS gene fusion status, clinical T-category, presence of intraductal carcinoma or cribriform architecture, biochemical recurrence (BCR), age at treatment, pre-treatment prostate specific antigen levels, kataegis score, chromothripsis score and ISUP scores (dichotomized by ISUP 1 and 2 vs. ISUP 3). P values were adjusted for multiple comparisons using FDR. Mann-Whitney U tests were performed on mRNA abundances and ETS gene fusion status to identify mRNAs that are associated with the presence of ETS gene fusions and p values were corrected using FDR. Spearman's ρ was calculated on the difference in fold-change between mRNA and protein abundances in genes that were significantly associated with ETS gene fusion status at either the mRNA or protein level.

ETS Fusion Associated Gene Intersection

Mann-Whitney U tests were performed to identify H3K27Ac peaks, methylation and copy number aberrations associated with ETS gene fusions. P values were corrected using FDR. The VennDiagram (v1.6.19) (Chen and Boutros, 2011) package was used to identify the genes and number of genes found at all possible intersections amongst the genes significantly associated with ETS gene fusions in the protein, mRNA, methylation, H3K27 acetylation, and copy number data.

Pathway Enrichment Analysis

Gene sets of interest were processed using g:Profiler (Reimand et al., 2011) (v r1732_e88_eg35; significant only; query ordered by significance when applicable; the list of all proteins detected as the background; significance threshold set to FDR; output set to generic enrichment map; gene ontology and REACTOME databases), which was subsequently visualized in Cytoscape (v3.6.1) (Shannon et al., 2003) using the Enrichment Map App (Merico et al., 2010). For the ETS associated genes (at FDR < 0.05) g:Profiler was ran on all gene sets separately, but visualized in the same instance to better show potential overlaps in pathways.

Correlation Analysis between mRNA and Protein Abundances

To determine the strength of the correlation between mRNA and protein abundances for each gene, overlapping genes (n = 6,946) were identified between the two data types using 55 matched samples. Correlation between the mRNA and protein abundance values for each of these gene was determined using Spearman's ρ.

Identification of Proteins with Undetected Transcripts and *vice versa*

High abundance mRNA transcripts were identified by filtering out transcripts that appeared in two or fewer samples and then selecting transcripts with a median abundance in the top 10%. The high abundance transcripts were filtered against the detected proteins to identify RNAs without a protein counterpart. Similarly, to identify proteins without an RNA transcript, proteins were intersected with transcripts that have a median abundance of zero in the cohort. Pathway enrichment analysis was performed on both sets of genes using g:Profiler. To quantify number of coding and non-coding transcripts, the transcript sets were annotated using information from Gencode v24lift37 (Harrow et al., 2012).

Association Analyses of CNAs on mRNA and Protein Abundances

FDR adjusted p values from a two-sided t -test and fold changes were calculated for each gene (n = 6,607, microarray) by CNA locus (n = 23,068) testing the difference in mean mRNA abundance between samples with a copy number aberration against those without (n = 210 samples). CNA status was quantified from OncoScan SNP arrays and mRNA abundance data was measured from Affymetrix transcriptome arrays. The same analysis was performed using protein abundance data (n = 55).

Mutual Information Analysis

For each pairwise combination of mRNA abundance, protein abundance, CNA state, methylation β value and H3K27Ac score, mutual information was calculated in bits for each gene using 21 bins, and the entropy function from the Entropy R package (v1.2.1) (Hausser and Strimmer, 2008). $I(X;Y) = H(X) + H(Y) - H(X,Y)$, where H(X) and H(Y) are the marginal entropies and H(X,Y) is the joint entropy. MI was normalized over the mean entropy of the two input vectors. Consensus clustering was performed on the z-scored normalized MI using a maxK of 15, Spearman's correlation as the similarity metric and with 1,000 replicates. Hallmark enrichment analysis was performed using a hypergeometric test. P values were FDR-adjusted to control for multiple comparisons. Agreement of normalized MI between our dataset and TCGA was assessed using the area under the receiver operating characteristic curve. For each gene, normalized MI was binarized within each dataset based on whether it was above or below a threshold. True positive and false positive rates were calculated using whether normalized MI was greater than the threshold in our dataset as a True Positive.

Percent Variance Analysis

Sum of squares were extracted from an ANOVA on a linear model of protein abundances to CNA, RNA, and methylation R for the following genes: *TGM2*, *NDRG3*, *KLK3*, *AKT1*, *PTEN*, *NKX3-1*, *KRAS* and *ATM*. Percent variance explained was calculated as the sum of squares for each of the input variables divided by the total sum of squares. For *KLK3* and *PTEN*, Spearman ρ was used to determine if protein abundances are correlated with their corresponding mRNA abundances and methylation. For PTEN, a Mann-Whitney U test was used to determine if the CNAs are associated with protein abundances.

Univariate Survival Analysis

Survival analysis was performed on the top 25% of proteins whose abundance had the highest variance, and their mRNA counterparts, to determine if the protein abundance or the mRNA abundance of each gene were univariately associated with biochemical recurrence. Hazard ratios were calculated by fitting Cox PH models to patient groups dichotomized using the median of the protein intensity or mRNA abundance against BCR as the endpoint. Assumptions for the Cox PH models were tested using the *cox.zph* function in the R Survival package (v2.41-3). Genes were considered to have divergent association with outcome if they were significantly associated with BCR, but $\log_2$ hazard ratios had opposite signs. For genes that failed the assumptions for a Cox PH model ($p < 0.1$), a log-rank test was used and p values were adjusted for multiple comparisons using FDR. Spearman's ρ was used to calculate the relationship of hazard ratios between mRNA and proteins. For proteins that were detected in 15% to 85% of the 75 samples (one sample was removed due to lack of clinical information), patients were dichotomized based on protein presence and absence before being fitted with a Cox PH model to determine if presence or loss of that protein group was associated with biochemical recurrence. P values were adjusted for multiple comparisons using FDR. Kaplan-Meier curves were generated for specific genes of interest. Validation of prognostic proteins was performed through immunohistochemistry on 79 additional prostate cancer samples in tissue microarrays. Antibodies for ACAD8 were obtained from Sigma-Aldrich (Prestige Antibodies - HPA040689 for ACAD8; Lot #A114184). A pathologist (M.M.) scored each core in comparison with the internal positive control using a semi-quantitative scoring system based on the intensity of the staining (cytoplasmic for ACAD8): 0 (No reactivity), 1 (mild intensity), 2 (moderate) and 3 (high intensity, equivalent to positive control cells). Staining intensity was recorded per TMA core. A Cox PH model for the validation set was fit to patients' groups, dichotomized by whether they have any cores scored with a '3 - high intensity staining'.

Biomarker Null Distribution Analysis

To assess the performance of biomolecules at predicting patients with BCR at 10 years, a null distribution analysis was performed using 10 million areas under the receiver-operating characteristic curves (AUC) for each biomolecule (40 million AUCs total). To calculate each AUC, 100 genes were randomly selected without replacement from the intersection of the genes present in each biomolecule data matrix. A random forest classification model with 4-fold cross validation (randomForest v4.6-12) (Svetnik et al., 2003) was then built for each gene set in each biomolecule. The hyper-parameters *mtry* and *sampsize* were tuned through a grid search based on lowest out-of-bag errors while *nTrees* was set to 10,000 to reduce grid search time since having too few trees will negatively impact model performance, but having more trees only incurs more computational time (Huang and Boutros, 2016). For protein,

mRNA and CNA, there were 7,042 matched protein-groups to genes. Methylation data was set to one probe per gene as described above. For both the RNA-seq and protein, random gene sets of 5, 10, 25, 50 and 100 were used for the random forest model to determine if gene set numbers will change the conclusions, whereas the CNA and methylation null distributions was generated using a gene set size of 100. To evaluate the predictive power of using two biomolecules at the same time, 100 genes were randomly selected and values for both biomolecules were used as features in the same random forest model.

QUANTIFICATION AND STATISTICAL ANALYSIS

The specific statistical tests used are indicated in the figure legends or appropriate methods section and were performed within the R statistical environment (v3.3.1). Visualization in R was performed through the BoutrosLab.Plotting.General package (v5.9.2) (P'ng et al., 2019). Pathway network graphs were generated using Cytoscape (v3.6.1) with the Enrichment Map App (Merico et al., 2010). Study outline was produced with Inkscape (v0.48) for Ubuntu.

DATA AND SOFTWARE AVAILABILITY

MS data was deposited in UCSD's MASSive database under the accession MassIVE: MSV000081552 at ftp://massive.ucsd.edu/MSV000081552. Oncoscan CNA microarray data can be found in the European Genome-Phenome Archive (EGA) at https://www.ebi.ac.uk/ega/studies/EGAS00001000900. Whole genome DNA sequencing and RNA-seq data can also be found on EGA, under accession EGA: EGAS00001000900. H3K27Ac CHiP-Seq data are in the Gene Expression Omnibus under GEO: GSE96652. RNA microarray data is available under the accession GEO: GSE107299. Methylation data is available under the accession GEO: GSE107298.