NEOPLASIA
www.neoplasia.com

Volume 17 Number 4 April 2015 pp. 385–399 385

Development and Validation of a Scalable Next-Generation Sequencing System for Assessing Relevant Somatic Variants in Solid Tumors[1,2]

Daniel H. Hovelson[*,3], Andrew S. McDaniel[†,3], Andi K. Cani[†,3], Bryan Johnson[‡‡], Kate Rhodes[‡‡], Paul D. Williams[‡‡], Santhoshi Bandla[‡‡], Geoffrey Bien[‡‡], Paul Choppa[‡‡], Fiona Hyland[‡‡], Rajesh Gottimukkala[‡‡], Guoying Liu[‡‡], Manimozhi Manivannan[‡‡], Jeoffrey Schageman[‡‡], Efren Ballesteros-Villagrana[‡‡], Catherine S. Grasso[§§,4], Michael J. Quist[§§], Venkata Yadati[†], Anmol Amin[†], Javed Siddiqui[†], Bryan L. Betz[†], Karen E. Knudsen[¶¶,##,***], Kathleen A. Cooney[‡,¶,**], Felix Y. Feng[§,**], Michael H. Roh[†], Peter S. Nelson[†††,‡‡‡], Chia-Jen Liu[†], David G. Beer[#,**], Peter Wyngaard[‡‡], Arul M. Chinnaiyan[*,†,¶,**,††], Seth Sadis[‡‡], Daniel R. Rhodes[†,‡‡] and Scott A. Tomlins[†,¶,**]

*Michigan Center for Translational Pathology, Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA; [†]Department of Pathology, University of Michigan Medical School, Ann Arbor, MI, USA; [‡]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI, USA; [§]Department of Radiation Oncology, University of Michigan Medical School, Ann Arbor, MI, USA; [¶]Department of Urology, University of Michigan Medical School, Ann Arbor, MI, USA; [#]Department of Surgery, University of Michigan Medical School, Ann Arbor, MI, USA; **Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA; [††]Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI, USA; [‡‡]Thermo Fisher Scientific, Ann Arbor, MI, USA; [§§]Department of Pathology, Oregon Health and Sciences University, Portland, OR, USA; [¶¶]Department of Cancer Biology, Thomas Jefferson University, Philadelphia, PA, USA; [##]Department of Urology, Thomas Jefferson University, Philadelphia, PA, USA; ***Department of Radiation Oncology, Thomas Jefferson University, Philadelphia, PA, USA; [†††]Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [‡‡‡]Division of Clinical Research, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Abbreviations: AOHC, AcroMetrix Oncology Hotspot Control; CNAs, copy number alterations; FFPE, formalin-fixed paraffin-embedded; GoF, gain-of-function; indels, insertions/deletions; LoF, loss-of-function; LU, lung cohort; MCR, minimal common region; MO, molecular cohort; NCCN, National Comprehensive Cancer Network; NGS, next-generation sequencing; OCP, Oncomine Comprehensive Panel; PGM, Personal Genome Machine; PR, prostate cohort; QMRS, Quantitative Multiplex Reference Standard; SCC, small cell carcinoma; TCGA, The Cancer Genome Atlas

Address all correspondence to: Scott A. Tomlins, MD, PhD, University of Michigan Medical School, 1524 BSRB, 109 Zina Pitcher Place, Ann Arbor, MI 48109-2200, USA.
E-mail: tomlinss@umich.edu

[1] This article refers to supplementary materials, which are designated by Tables S1 to S16 and Figures S1 to S8 and are available online at www.neoplasia.com.

[2] Support: This work was supported in part by the Evans Foundation/Prostate Cancer Foundation (to F.Y.F., K.E.K., and S.A.T), by the National Institutes of Health (R01 CA183857 to S.A.T., R01 CA181605 to P.S.N. and S.A.T., R01 CA159945 to K.E.K., R01 CA154365 to D.G.B., and UM1 HG006508 to A.M.C.), by the Department of Defense (PC120464 to K.A.C.), and by a sponsored research agreement with Life Technologies, now part of Thermo Fisher Scientific. F.Y.F. and S.A.T. were supported by the University of Michigan Prostate SPORE Career Development Awards. F.Y.F., P.S.N., A.M.C., and S.A.T. are supported by a Stand Up To Cancer—Prostate Cancer Foundation Prostate Dream Team Translational Cancer Research Grant. Stand Up To Cancer is a program of the Entertainment Industry Foundation administered by the American Association for Cancer Research (SU2C-AACR-DT0712). F.Y.F., A.M.C., and S.A.T. are supported by the A. Alfred Taubman Medical Research Institute. Conflicts of interest: The University of Michigan has been issued a patent on the detection of ETS gene fusions in prostate cancer, on which A.M.C. and S.A.T are listed as co-inventors. The University of Michigan licensed the diagnostic field of use to Gen-Probe, Inc, who has sublicensed some rights to Ventana/Roche. S.A.T. serves as a consultant to, and has received honoraria from, Ventana/Roche. B.J., K.R., P.D.W., S.B., G.B., P.C., F.H., R.G., G.L., M.M., J.S., E.B.-V., P.W., S.S., and D.R.R. are employees of Thermo Fisher Scientific. S.S. and D.R.R. own stock in Thermo Fisher Scientific. The remaining authors have no competing interests to disclose. S.A.T. has a sponsored research agreement with Thermo Fisher Scientific that supported this work. Thermo Fisher was involved in data collection, interpretation, and analysis and participated in the study design, drafting/revision of the manuscript, and the decision to submit for publication. D.H.H. and S.A.T. had access to all data and were responsible for primary data analysis.

[3] These authors contributed equally.

[4] Present address: Epidemiology Program, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA.

Received 26 March 2015; Accepted 27 March 2015

1476-5586/15
http://dx.doi.org/10.1016/j.neo.2015.03.004

Abstract

Next-generation sequencing (NGS) has enabled genome-wide personalized oncology efforts at centers and companies with the specialty expertise and infrastructure required to identify and prioritize actionable variants. Such approaches are not scalable, preventing widespread adoption. Likewise, most targeted NGS approaches fail to assess key relevant genomic alteration classes. To address these challenges, we predefined the catalog of relevant solid tumor somatic genome variants (gain-of-function or loss-of-function mutations, high-level copy number alterations, and gene fusions) through comprehensive bioinformatics analysis of >700,000 samples. To detect these variants, we developed the Oncomine Comprehensive Panel (OCP), an integrative NGS-based assay [compatible with <20 ng of DNA/RNA from formalin-fixed paraffin-embedded (FFPE) tissues], coupled with an informatics pipeline to specifically identify relevant predefined variants and created a knowledge base of related potential treatments, current practice guidelines, and open clinical trials. We validated OCP using molecular standards and more than 300 FFPE tumor samples, achieving >95% accuracy for *KRAS, epidermal growth factor receptor*, and *BRAF* mutation detection as well as for *ALK* and *TMPRSS2:ERG* gene fusions. Associating positive variants with potential targeted treatments demonstrated that 6% to 42% of profiled samples (depending on cancer type) harbored alterations beyond routine molecular testing that were associated with approved or guideline-referenced therapies. As a translational research tool, OCP identified adaptive *CTNNB1* amplifications/mutations in treated prostate cancers. Through predefining somatic variants in solid tumors and compiling associated potential treatment strategies, OCP represents a simplified, broadly applicable targeted NGS system with the potential to advance precision oncology efforts.

Neoplasia (2015) 17, 385–399

Introduction

Precision medicine approaches, where patients are treated with therapies directed against the specific molecular alterations driving their tumors, have revolutionized oncology [1–4]. Such approaches require identification of driving molecular alterations (which may occur only in a subset of a given histologic cancer type or in cancers arising from diverse organs), development of targeted therapies, and diagnostic tests to identify appropriate patient populations for clinical trials and eventual implementation [5–7]. The early successes of trastuzumab (a monoclonal antibody against ERBB2) in the subset of breast adenocarcinomas with *ERBB2* amplifications [8], and imatinib (an ABL kinase inhibitor) in the subset of leukemia driven by BCR-ABL gene fusions (chronic myeloid leukemia) [9], have been replicated in numerous cancers [1–4]. For example, multiplexed assessment of driving somatic alterations in lung cancer has been shown to aid in physician selection of therapy, and patients with drivers receiving a matched therapy lived significantly longer than those not receiving a matched therapy [10].

Recent advances in genome sciences, including next-generation sequencing (NGS), have led to the identification of hundreds of recurrent somatically altered genes through the analysis of tens of thousands of cancer samples from individual investigators and large consortia, such as The Cancer Genome Atlas (TCGA) [11–15]. These technological advances are also changing routine molecular pathology practice from single gene–based tests (i.e., Sanger sequencing to assess *(epidermal growth factor receptor) EGFR* mutations in lung adenocarcinoma) to multiplexed NGS assays. Several NGS approaches have been successfully clinically implemented in oncology, including multiplexed polymerase chain reaction (PCR)–based panels assessing tens to hundreds of genes, hybrid capture–based panels targeting hundreds of genes, as well as comprehensive exome/genome/transcriptome sequencing [16–26]. These approaches vary in sample requirements, nucleic acids assessed, cost, throughput, genes and alteration types assessed, and performance. For example, most clinically implemented multiplexed PCR-based approaches fail to assess copy number alterations (CNAs) and/or gene fusions [16,19,22,25], which guide current treatment selection for several cancers.

The primary challenge with comprehensive NGS approaches, however, is the specialty infrastructure and expertise needed to interpret the results and convey treatment strategies to clinicians. Several centers using comprehensive NGS-based oncology approaches require NGS-based tumor boards [21] to guide interpretation and inform clinical decision-making. Large companies have also been established with the goal of providing comprehensive NGS-based precision oncology services [18]; however, interpretation of results and prioritizing treatment strategies may still be outsourced. Scalability limitations hinder widespread adoption of such initiatives in reference laboratories.

To enable precision medicine approaches for all patients with cancer, rapid, inexpensive, scalable NGS solutions capable of assessing all classes of current and near-term clinically relevant targets [point mutations, short insertions/deletions (indels), CNAs, and gene fusions] from routine formalin-fixed paraffin-embedded (FFPE) tissues are required. Such a technical solution must be coupled with

a dynamic, scalable, analytical approach capable of prioritizing treatment options. To begin to address these challenges, we report the development and validation of the Oncomine Comprehensive Panel (OCP), a multiplexed PCR-based NGS assay and analytical system to identify and prioritize potential treatment strategies from predefined somatic solid tumor genome variants. The OCP is compatible with 20 ng of FFPE isolated DNA and 15 ng of FFPE isolated RNA and benchtop Ion Torrent sequencers. Demonstrating the potential for a scalable solution to enable widespread precision medicine oncology applications, the OCP will be used in the National Cancer Institute (NCI) Match Trial to assess 3000 cancer samples for trial selection in a multi-arm umbrella study with sequencing conducted at multiple sites.

Materials and Methods

Analysis of Relevant Somatic Variants in Solid Tumors

The OCP was designed to interrogate somatic mutations, CNAs, and gene fusions involving oncogenes and tumor suppressors recurrently altered in solid tumors with the potential for near-term clinical relevance. To define OCP content, we used evidence-based analysis of genomic alterations present in Oncomine, a resource composed of mutation, copy number, and gene fusion data from >700,000 cell line, xenograft, and clinical cancer samples as of December 2013 [27–29]. Candidate genes with somatic driver mutations were derived from gain-of-function (GoF) and loss-of-function (LoF) analyses performed on 686,530 tumor samples with mutation data in Oncomine. Candidate driver CNA events were identified by performing a minimal common region (MCR) assessment on a pan-cancer subset of 10,249 tumor samples in Oncomine [28]. In addition, single cancer–type assessments were performed to identify private candidate copy number drivers. Candidate driver gene fusions were identified from the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (http://cgap.nci.nih.gov/Chromosomes/Mitelman) as well as from analyzing 6438 primary tumor sample RNA-seq profiles contained within Oncomine. Complete details on somatic variant analysis to define candidates are provided in the Supplementary Materials. All candidate driver GoF, LoF, and CNA genes, as well as gene fusions, were then assessed for evidence of near-term potential clinical relevance as defined in the Supplementary Materials.

OCP NGS Assay Design

We developed multiplexed PCR (Ion AmpliSeq) NGS panels to characterize DNA (mutations and CNAs) and RNA (gene fusions) based alterations. For GoF alterations, amplicons were designed to assess recurrent hotspots as defined above. For LoF alterations, amplicons were designed to tile the gene's entire coding sequence. For CNAs, sufficient amplicons (n = 3 to 38) were designed from coding and noncoding regions to facilitate copy number profiling. For the RNA-based panel, primers were included to detect known gene fusion junctions, assess the 5′ and 3′ regions of *RET*, *ROS1*, and *ALK* (to enable testing for fusions involving novel 5′ fusion partners through 3′/5′ expression imbalance), and quantify housekeeping/positive expression genes (*HMBS*, *ITGB7*, *LMNA*, *MYC*, and *TBP*); a small subset of the prostate cancer samples (n = 12) was sequenced using a version of the RNA panel that did not contain the 3′/5′ expression imbalance assays. AmpliSeq panels were designed using AmpliSeq Designer, and multiplexed pools were obtained from Ion Torrent. Two versions of both the DNA- and RNA-based panels were assessed herein during iterative optimization, and complete information on the panels is provided in Table S1.

Molecular Standards

We used two commercially available molecular standards to assess performance of the DNA component of the OCP. The AcroMetrix Oncology Hotspot Control (AOHC; Life Technologies, Foster City, CA) was designed to assess somatic mutation detection performance by NGS assays. The custom version used herein contained 365 applicable single/multiple nucleotide variants (SNVs/MNVs) and 33 indels each at an estimated allele frequency of 0.20 on the GM24385 cell line genomic background. AOHC DNA was used directly for library preparation.

The Quantitative Multiplex Reference Standard (QMRS; Horizon Diagnostics, Cambridge, United Kingdom) consists of one to three FFPE tissue sections from multiplexed FFPE cell lines with a known set of 30 engineered and endogenous mutations present at specific variant allele frequencies quantified by droplet digital PCR. Of the 30 mutations, 16 (all 11 primary engineered mutations and 5 of 19 secondary endogenous mutations) were targeted by the OCP and were used for evaluation. QMRS tissue was processed as for the remaining tissue cohorts for DNA isolation and library preparation.

Tissue Cohorts

We used three cohorts of routine FFPE tissues for OCP evaluation [molecular (MO), lung (LU), and prostate (PR)]. All FFPE specimens were obtained from the University of Michigan Department of Pathology Tissue Archive with Institutional Review Board (IRB) approval. Diagnostic hematoxylin and eosin–stained slides were reviewed by board-certified anatomic pathologists (A.S.M. and S.A.T.).

The MO cohort consisted of all cancer specimens (including biopsy, resection, and cell block specimens) sent during a 5-month period to the Clinical Laboratories Improvement Amendments–certified University of Michigan Molecular Oncology/Genetics Laboratory for 1) *EGFR*, *BRAF*, or *KRAS* mutation testing or 2) *ALK* rearrangement testing. Complete details of the MO cohort and clinicopathologic information for all cases are provided in the Supplementary Materials and Table S2. The LU and PR cohorts consisted of 104 and 118 retrospectively identified FFPE tissue specimens, respectively. A subset of the PR samples (n = 37) has been previously assessed by a combined capture-based NGS (Agilent Haloplex followed by Ion Torrent NGS) and TaqMan low-density array quantitative reverse transcription–PCR (qRT-PCR) panel [30]. Clinicopathologic information for all included cases in the LU and PR cohorts is provided in Tables S3 and S4. Targeted NGS of all tumor tissues was performed with IRB approval.

Nucleic Acid Isolation

For each specimen, 3 × 10 to 10 × 10 μm FFPE sections were cut from a single representative block per case, using macrodissection with a scalpel as needed to enrich the tumor content. DNA and RNA were isolated using the Qiagen AllPrep FFPE DNA/RNA Kit (Qiagen, Valencia, CA) as described [31]. DNA and RNA were quantified using the Qubit 2.0 fluorometer (Life Technologies).

DNA/RNA Libraries

DNA/RNA libraries were generated essentially as described [31,32]. DNA libraries were generated from 20 ng of DNA per sample using the Ion AmpliSeq Library Kit 2.0 (Life Technologies)

and the OCP AmpliSeq panel according to the manufacturer's instructions with barcode incorporation. RNA libraries were generated from 15 ng of RNA per sample using the Ion AmpliSeq RNA Library Kit. OCP AmpliSeq libraries were quantified using the Ion Library Quantification Kit according to the manufacturer's instructions.

Template Generation and Sequencing

Templates for DNA and RNA libraries were prepared using the Ion Personal Genome Machine (PGM) Template OT2 200 Kit (Life Technologies) on the Ion One Touch 2 according to the manufacturer's instructions. Sequencing of multiplexed templates was performed using the Ion Torrent PGM on Ion 318 chips using the Ion PGM Sequencing 200 Kit v2 (Life Technologies) according to the manufacturer's instructions. For the LU and PR cohorts, a single DNA template and four to eight RNA templates were assessed separately on a single 318 chip. For the MO cohort, a single DNA template was combined with a single RNA template in a 4:1 ratio and assessed on a single 318 chip. For experiments with molecular standards, single DNA templates were assessed on one 318 chip.

Data Analysis

Data analysis was performed using Torrent Suite (4.2.0) and the Coverage Analysis (or Coverage Analysis RNA) plug-ins (both v4.0-r73765), along with the Ion Reporter (4.2.0) Fusion analysis workflow essentially as described [31,32]. For DNA sequencing (DNA-seq), alignment was performed using TMAP with default parameters, and variant calling was performed using the Torrent Variant Caller (TVC) plug-in (version 4.2-8-r87740) using default low-stringency somatic variant settings. Somatic variant identification was performed essentially as described [31,33] using read and base level filtering, which we have previously confirmed to identify variants that pass Sanger sequencing validation with >95% accuracy. Copy number analysis from total amplicon read counts provided by the Coverage Analysis plug-in was performed essentially as described [19,31,32]. As an estimate of data quality, we determined the SD of the amplicon-level copy number estimates relative to the gene-level estimate for each gene per sample (Figure S1). Gene fusion analysis was performed within the Ion Reporter (4.2.0) Fusion analysis workflow, with reads from the RNA AmpliSeq panel aligned using TMAP to a gene reference of targeted chimeric fusion transcripts as well as reference sequences for expression imbalance and expression control gene targets. Complete description of all data analyses is provided in the Supplementary Materials.

Alteration Prioritization and Potential Actionability Assessment

Somatic SNVs/indels passing filtering in a GoF gene were considered GoF if occurring at the predefined hotspot residue targeted in OCP. Somatic variants in an LoF gene were considered LoF if deleterious (nonsense or frame shifting) or occurring at a predefined hotspot residue. Somatic CNAs were considered for potential actionability analysis if they were concordant with the predicted alteration (amplification or deletion) from Oncomine analysis as described above. Somatic gene fusions were considered for actionability analysis if they represented known gene fusions from the Mitelman database or Oncomine analysis or involved known 3′ or 5′ drivers with novel partners (i.e., *ERC1-BRAF* fusion in MO-17, with recurrent fusions involving *BRAF* as a 3′ partner reported previously [34]).

These prioritized variants were then associated with potential actionability using the Oncomine database. Briefly, for each patient, the "most actionable" alteration was identified by prioritizing 1) variants referenced in Food and Drug Administration (FDA) drug labels, 2) variants referenced in National Comprehensive Cancer Network (NCCN) treatment guidelines in the patient's cancer type, 3) variants referenced in an NCCN guideline in another cancer type, and 4) variants referenced as inclusion criteria in a clinical trial. Actionable variants were identified by manual curation of FDA labels and NCCN guidelines and by keyword searches and manual curation of clinical trial records in the TrialTrove database. Alterations associated with specific treatments are shown in Table S5.

qRT-PCR and Immunohistochemistry Validation

Details of qRT-PCR validation of *ERC1:BRAF* and *TPR:NTRK1* fusions as well as ERBB2 immunohistochemistry (IHC) to confirm copy number gains are provided in the Supplementary Materials and Table S6.

Statistical Tests. All statistical tests were performed in R (3.1.0) using two-sided tests. P values $<.05$ were considered statistically significant.

Results

OCP Development

To define relevant somatic cancer genome variants based on near-term potential actionability, we first interrogated data from >700,000 tumor samples in the Oncomine database (including >8000 exomes, >7000 transcriptomes, and >30,000 copy number profiles in addition to tumors studied by single gene/targeted approaches) to identify pan-cancer, recurrently altered oncogenes (enriched in GoF hotspot mutations), tumor suppressors (enriched in LoF deleterious mutations), genes targeted by high-level amplifications or deletions, and driving gene fusions (Figure 1*A*). Genes with these variants were then filtered based on near-term potential actionability (see Materials and Methods section). The distribution of these variants across >7000 TCGA samples from 23 cancer types is shown in Figure 1*B*.

To translate the relevant somatic cancer genome to an NGS assay capable of detecting mutations, CNAs, and gene fusions (including multiple splice isoforms) but compatible with limited amounts of routine FFPE tissues, we developed custom Ion Torrent multiplexed PCR-based DNA-seq and RNA sequencing (RNA-seq) panels, together comprising the OCP, as shown in Figure 1*C*. In total, the final OCP version assessed herein (v0.9b) interrogates 143 unique cancer genes including 73 oncogenes, 49 CNA genes, 26 tumor suppressor genes, and 22 fusion driver genes. The targeted DNA-seq panel includes 2530 amplicons covering 260,717 base pairs in 130 different genes. To minimize panel size and focus on predefined relevant alterations, only GoF mutations were targeted in oncogenes, while high-level CNA genes were targeted by 3 to 38 probes to facilitate copy number profiling [19] and the entire coding sequence of tumor suppressors were targeted to identify LoF mutations and GoF mutations. The targeted RNA-seq panel included a total of 154 primer pairs targeting known gene fusion isoforms (n = 148) as well as 5′ and 3′ expression assays for *RET*, *ROS1*, and *ALK* to enable novel fusion discovery through 3′/5′ expression imbalance ratios. To enable appropriate normalization in downstream analyses, the targeted RNA-seq panel also includes five additional primer pairs targeting a pre-determined set of housekeeping/positive expression genes (*HMBS*, *ITGB7*, *LMNA*, *MYC*, and *TBP*). Details of the two versions of the OCP panel (v0.9a and v0.9b) validated and applied herein are presented in Table S1.

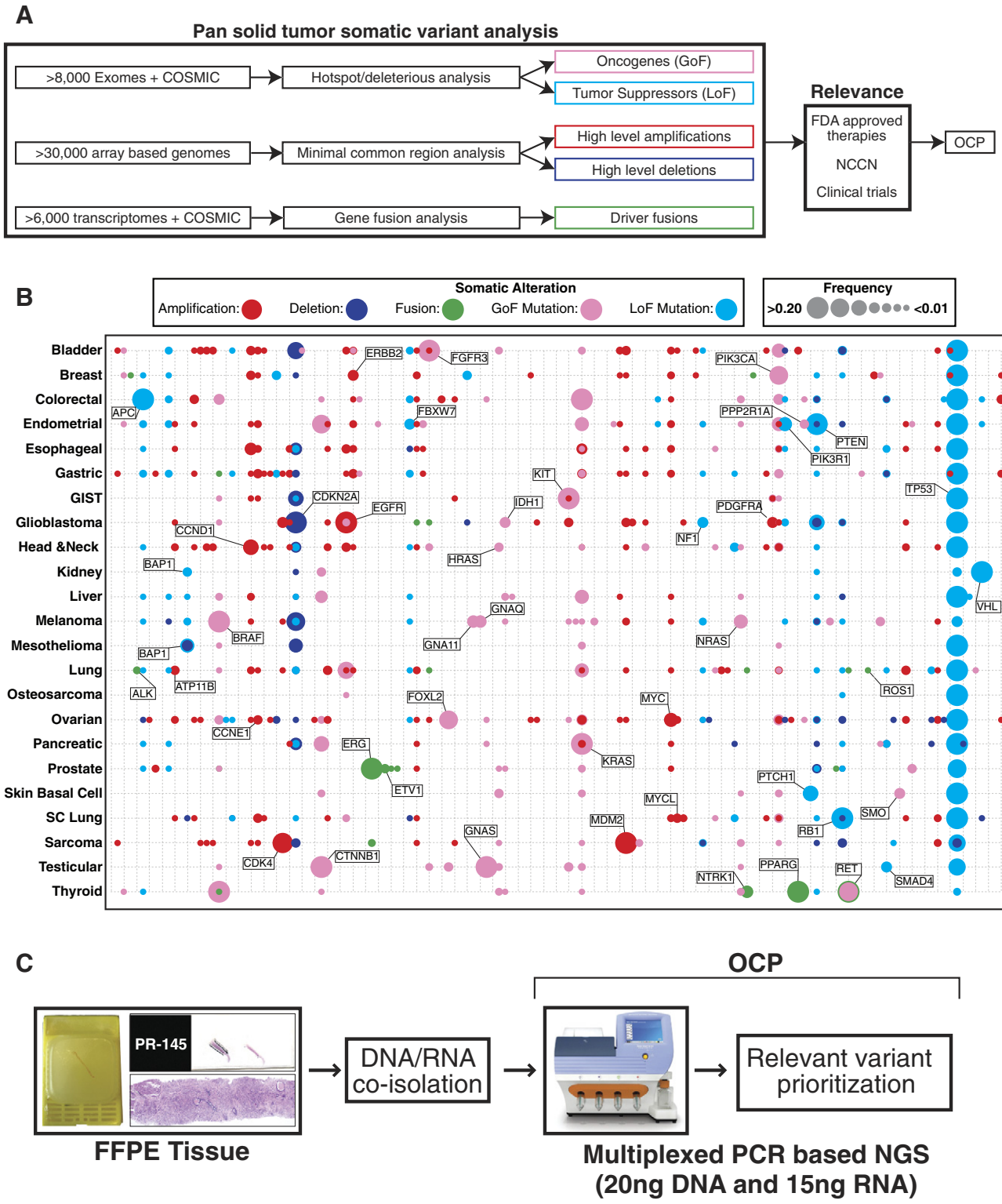
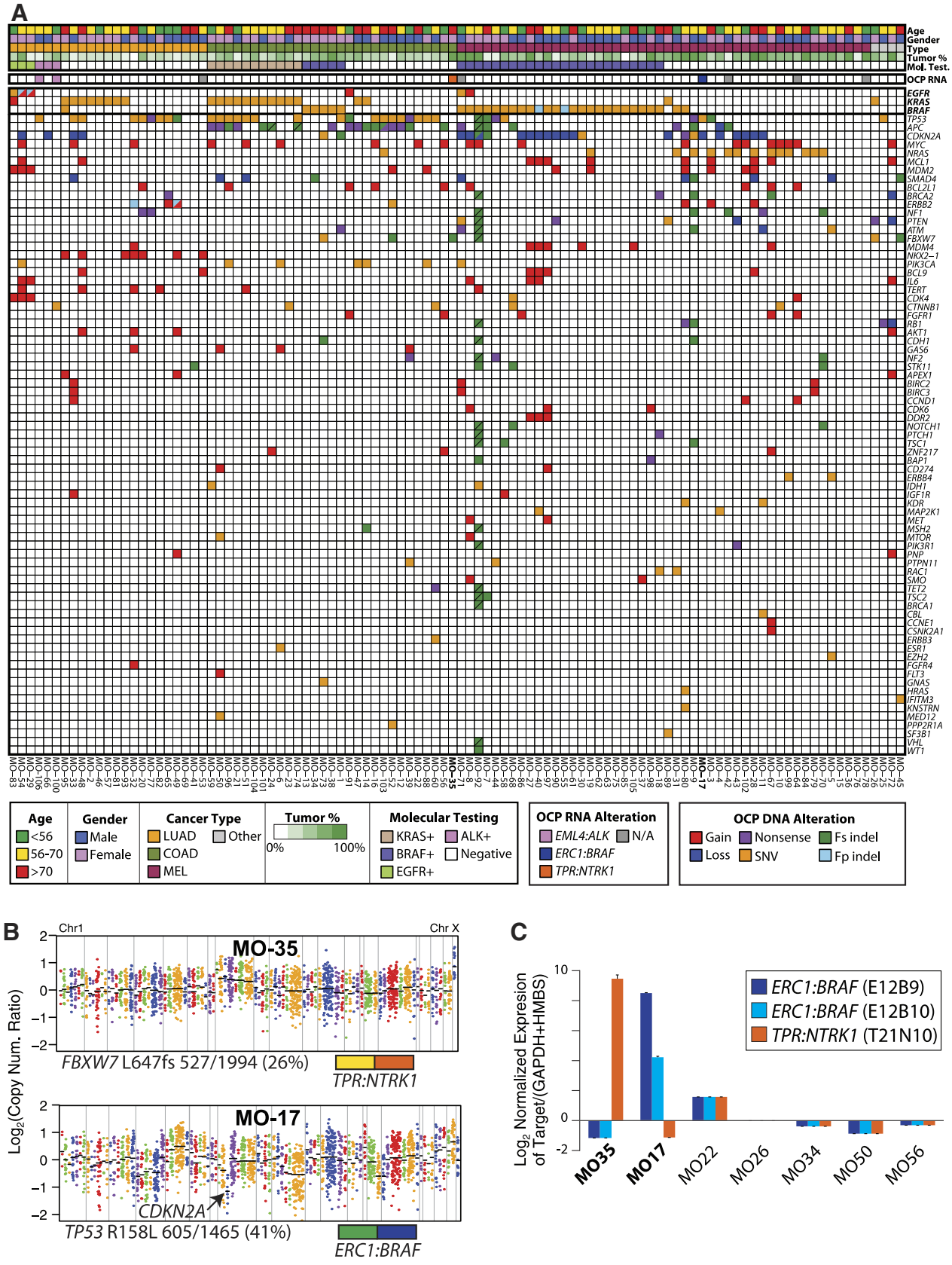


Figure 1. Pan-solid tumor cancer somatic alteration analysis to identify relevant variants. (A) Using the Oncomine database supplemented with data from COSMIC, more than 700,000 tumor samples (including >8000 cancer exomes) were used to assess genes for overrepresentation of hotspot (GoF) and deleterious (LoF) mutations to identify oncogenes and tumor suppressors, respectively. Array-based copy number profiles from >30,000 tumors were assessed by MCR analysis to identify targets of focal, high-level amplifications or deletions. Transcriptomes from >7000 cancers were similarly assessed for driver gene fusions. Prioritized genes were further filtered to include only near-term relevant alterations for inclusion into the OCP. (B) Frequency of somatic alterations (type according to color in the legend) in OCP included genes across publically available TCGA data. For each gene per cancer type, alteration frequency (<0.01 to >0.20) is indicated by the size of the circle according to the legend. Selected genes of interest are highlighted. (C) The OCP was designed for compatibility with routine FFPE tissues, with co-isolation of DNA/RNA from FFPE tissues used in our validation. The OCP consists of multiplexed PCR (AmpliSeq) panels compatible with 20 ng of DNA and 15 ng of RNA, which can be combined after library generation for NGS on Ion Torrent benchtop sequencers. By predefining relevant somatic variants, identified variants can be linked to potential treatment strategies.



Molecular Standard Validation

To validate OCP performance, we first assessed the DNA component using the AOHC molecular standard, a cell line DNA sample engineered to contain 398 OCP targeted variants at 0.20 expected variant allele frequencies. OCP detected 364 of 365 (99.7%) targeted SNVs/MNVs, with a median variant allele frequency of 0.24 (interquartile range 0.21-0.28) as shown in Figure S2 and Table S7. Of the 33 OCP targeted indels, we detected 25 (75.8%) at a median

variant allele frequency of 0.22 (interquartile range 0.18-0.28; Figure S2 and Table S7). Of the eight indels that were not detected, three were over 10 bases in length (12, 30, 41 bases) and five were single nucleotide insertions or deletions occurring within two to seven base homopolymer runs. Accurate indel identification in homopolymer runs is a known challenge with current Ion Torrent sequencing technology [16].

We also profiled DNA isolated from commercially available FFPE sections containing a cell line mixture (QMRS cell line) with engineered and endogenous mutations at precise variant frequencies. In total, 16 known mutations in the QMRS cell line (11 primary induced and 5 endogenous mutations; median variant allele frequency 0.10, range 0.01-0.33) are targeted by the OCP. To prioritize high-quality somatic variants, we applied our standard filtering approach (which includes filters at <5% or <10% depending on alteration type; see Materials and Methods section) to default variant calls. Ten of 16 (63%) OCP-targeted mutations (including 8/11 induced and 2/5 endogenous) were called by the TVC using our standard approach at variant allele frequencies highly correlated with those expected (r^2 = 0.99; Table S8). Five of the remaining six (83%) OCP-targeted mutations were detectable at close-to-expected frequencies—including indels and point mutations at <1% to 5% variant allele frequencies—through automated variant calling (Table S8). The only variant not detected by OCP was a secondary *NF1* frameshift deletion at the start of a 6-bp homopolymer run (expected frequency of 7.5%); hence, in total, 15 of 16 (94%) of OCP-targeted known mutations in the QMRS cell line were detected through our automated variant calling procedures (Table S8). Highly concordant results were observed with both OCP versions, as well as separate Ion Torrent PGM template preparation and sequencing runs performed at two locations (Ann Arbor, MI and Carlsbad, CA) from aliquots of the same DNA library (Table S8).

OCP Performance in FFPE Tumor Tissue Cohorts

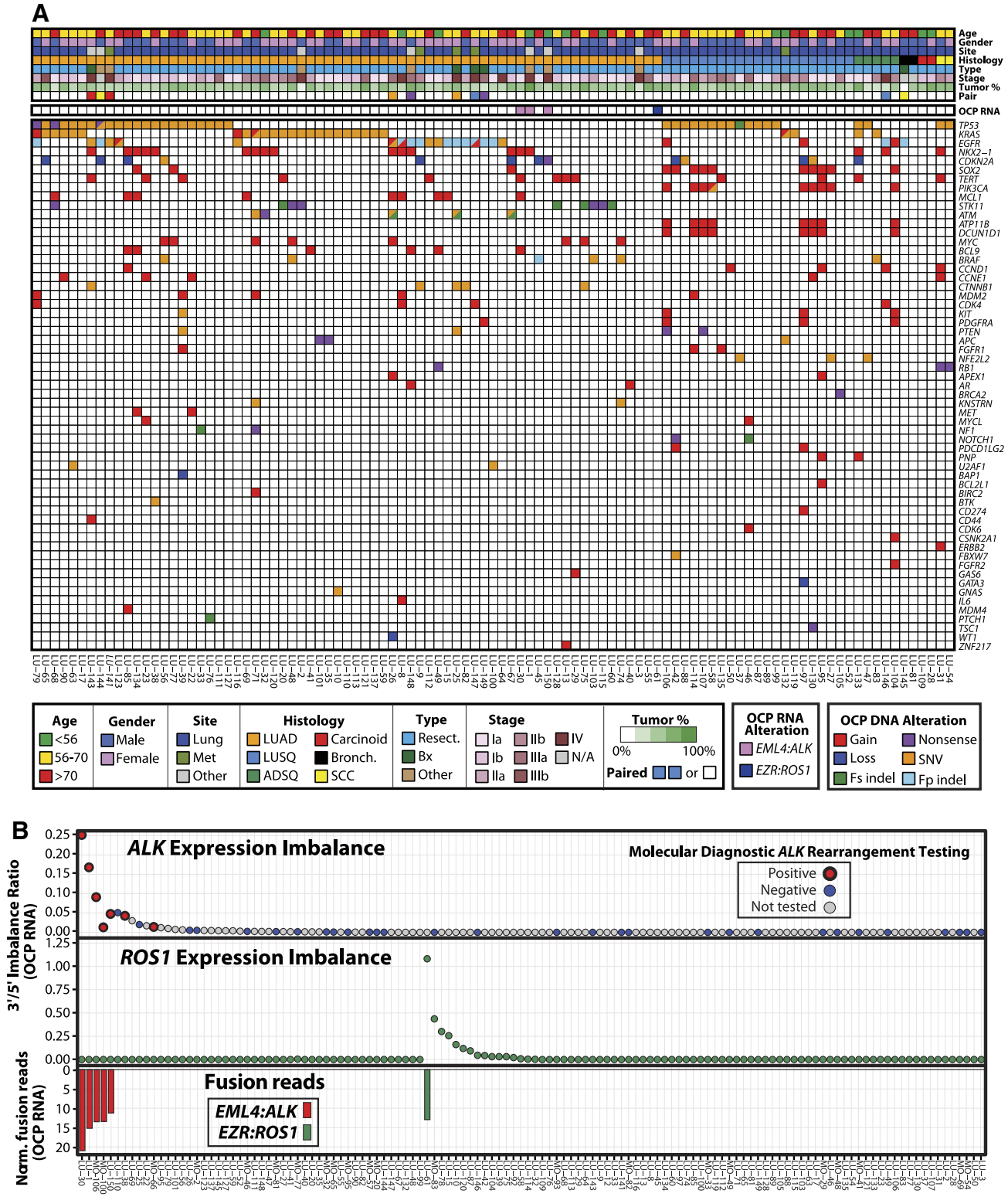
To validate performance and demonstrate applicability, we applied the OCP to three cohorts of routine FFPE tissue specimens: a cohort composed of tumor samples sent for routine molecular diagnostics (MO cohort, n = 105 samples) and retrospective lung cancer (LU, n = 104 samples) and prostate cancer (PR, n = 118 samples) cohorts. For each cohort, 3 × 10 to 10 × 10 μm FFPE sections were used for DNA/RNA co-isolation after macrodissection, with an overall average of 52% estimated post-dissection tumor content per sample (range 5%-90%), as assessed by histology (Tables S2–S4). Across the MO, LU, and PR cohorts, we isolated an average of 1.3/2.8 μg, 2.0/3.3 μg, and 2.2/6.7 μg of DNA/RNA per sample, respectively. Overall, 32% of the FFPE specimens were at least 3 years old (average 2 years, range <1 to 10 years).

Multiplexed PCR-based DNA and RNA libraries were generated from each sample for template preparation and NGS on the Ion Torrent PGM using Ion 318 chips. We excluded DNA and RNA libraries from 1/105 MO, 3/104 LU, and 2/118 PR samples due to low-quality libraries, resulting in a total of 321/327 (98%) informative samples. An additional two samples (MO-46 and LU-141) were excluded from CNA analysis due to excessively noisy copy number profiles (Figure S1). Across the three cohorts, using the DNA panel, we achieved an average of 5,142,690 mapped reads (97% on-target), 1941× coverage across targeted bases, 93.6% of targeted bases covered by at least 20 reads, and 202 called variants per informative sample. Using the RNA panel, we achieved an average of 306,872 total mapped reads (including 210,712 reads mapped to the five housekeeping/positive expression genes) per sample. Complete sequencing statistics are provided in Tables S2 to S4, with all called DNA variants, including splice site alterations, given in Table S9. All high-level OCP-prioritized (see Materials and Methods sections) CNAs and gene fusions across the cohorts are given in Tables S10 and S11, respectively.

OCP Validation in a Clinical Molecular Diagnostics Cohort

To validate the performance of the OCP and identify additional relevant variants beyond current routine practice, we assessed a cohort of 105 FFPE cancer samples sent for molecular testing for *EGFR*, *BRAF*, *KRAS*, and *ALK* alterations in a Clinical Laboratories Improvement Amendments/College of American Pathologists–certified molecular diagnostics laboratory. The 104 informative MO samples from 104 patients were composed of colorectal adenocarcinomas (n = 29), lung adenocarcinomas (n = 23), melanomas (n = 48), and four other cancers (see Table S2). After filtering to the predefined Oncomine variants, we identified an average of 1.7, 0.8, and 1.7 relevant somatic point mutations, indels, and high-level CNAs, respectively, per sample. Genes most frequently harboring relevant alterations across the MO cohort were *TP53* (33%), *BRAF* (31%),

Figure 2. Validation of the OCP using an oncology cohort undergoing molecular diagnostics testing. (A) We applied the OCP to a prospectively identified cohort of FFPE cancer samples undergoing molecular diagnostics testing for somatic mutations in *BRAF*, *KRAS* or *EGFR*, or *ALK* rearrangements (MO cohort). All OCP-defined relevant alterations from the RNA (in the header) and DNA components of the OCP for the 104 informative samples are shown in the heat map. Specific alteration types are indicated according to the legend (Nonsyn. SNV = nonsynonymous SNV; Fs and Fp indel = frame-shifting and frame-preserving indels, respectively). Slashed boxes indicate two alterations. Samples not sequenced in OCP RNA analysis are indicated as in the legend. Samples excluded from copy number analysis due to noisy profiles are shown in italics. Clinicopathologic information is given in the header according to the legend (LUAD, lung adenocarcinoma; COAD, colon adenocarcinoma; MEL, melanoma); 100% concordance with molecular testing was observed for mutations (see Table W10). Detailed OCP RNA-seq results, including 3′/5′ expression imbalance, for the *ALK* rearrangement–positive lung cancers are shown in Figure 3*B*. (B) Integrative OCP results from two cases, MO-17 and MO-25 (names bolded in A) harboring relevant gene fusions. Copy number plots show $\log_2$ copy number ratios (compared to a composite normal sample) per amplicon, with each individual amplicon represented by a single dot and individual genes indicated by different colors. Gene-level copy number estimates are shown as black bars. By OCP, MO-17 (top), a *BRAF* wild-type melanoma by clinical testing, harbored *CDKN2A* high-level copy number loss, *TP53* R158L mutation, and a novel *ERC1:BRAF* gene fusion. OCP profiling of MO-35, a *KRAS*/*BRAF* wild-type colon adenocarcinoma by clinical testing, identified an *FBXW7* L647fs mutation and a *TPR:NTRK1* gene fusion. For mutations, variant allele containing reads/total reads and the variant allele frequency are shown. (C) Validation of OCP identified gene fusions using qRT-PCR for *ERC1:BRAF* [*ERC1* exon 12 fused to *BRAF* exon 9 (E12B9, blue) or 10 (E12B10, cyan)] and *TPR:NTRK1* [*TPR* exon 21 fused to *NTRK1* exon 10 (T21N10, orange)]. qRT-PCR was performed on MO-17, MO-35, and five control MO samples without OCP-detected gene fusions. Mean $\log_2$ expression (normalized to the arithmetic mean of *glyceraldehyde-3-phosphate dehydrogenase (GAPDH) + hydroxymethylbilane synthase (HMBS)* calibrated to the mean of the MO control samples) + SD of triplicate qPCRs is plotted. No detectable expression of *ERC1:BRAF* or *TPR:NTRK1* was present in any sample other than that identified by OCP.



and *APC* (24%). An integrative heat map of prioritized alterations across the MO cohort is shown in Figure 2*A*, and copy number profiles for all samples are shown in Figure S3.

A total of four prioritized gene fusions was identified across the cohort: *EML4:ALK* in two lung cancer samples positive for *ALK* rearrangement by molecular testing (MO-100 and MO-106), *ERC1:BRAF* in a melanoma sample (MO-17), and *TPR:NTRK1* in a colon cancer sample (MO-35; Figure 2, *A* and *B*). Importantly, multiple isoforms of the *ERC1:BRAF* fusion were identified in MO-17 due to combinatorial priming/amplification, including fusions of *ERC1* and *BRAF* exons 17 to 8 (designated as E17B8), 12 to 9 (E12B9), and 12 to 10 (E12B10), respectively. qRT-PCR confirmed expression of *ERC1:BRAF* and *TPR:NTRK1* fusions in MO-17 and MO-35, respectively (Figure 2*C*).

The 104 informative MO samples underwent 129 total molecular diagnostic tests for *EGFR*, *BRAF*, *KRAS*, and *ALK* alterations. The

DNA component of the OCP demonstrated 100% sensitivity (44 of 44, 100%) and specificity (61 of 61, 100%) for detecting the clinically identified *EGFR*, *BRAF*, and *KRAS* mutations (Figure 2*A* and Table S12). Likewise, as described above, the RNA component of the OCP detected gene fusions involving *ALK* in two of the three (66%) samples with *ALK* rearrangements by fluorescence in situ hybridization (FISH) testing in the molecular diagnostics laboratory. In MO-66, a lung adenocarcinoma with an *ALK* rearrangement by molecular testing, OCP profiling identified only nine *EML4:ALK* fusion reads, which was below our threshold for calling a gene fusion present; however, as described below, we observed 3′/5′ *ALK* expression imbalance in this case (see Figure 3*B*). In total, considering MO-66 as failing to detect the *ALK* rearrangement, the 129 molecular tests performed across the MO cohort involving integrative DNA/RNA profiling by OCP showed 99.2% accuracy compared to molecular diagnostic testing. Additional findings from the MO cohort, including identification of relevant alterations not assessed by molecular testing, are described below.

OCP Application in a Lung Cancer Cohort

We also applied the OCP to a retrospectively identified cohort of 104 primary lung tumors given the assessment of somatic variants in lung cancer management [10]. The 101 informative samples from 96 individuals, which were chosen to represent the pathologic/histologic spectrum, consisted of 69 adenocarcinomas, 21 squamous cell carcinomas, 5 adenosquamous carcinomas, 2 bronchioloalveolar carcinomas (1 adenocarcinoma *in situ* and 1 well-differentiated lepidic predominant adenocarcinoma), 2 pulmonary small cell carcinomas (SCCs), and 2 carcinoid tumors (Figure 3*A* and Table S3). After filtering to the predefined Oncomine variants, we identified an average of 1.2, 0.3, and 1.9 relevant somatic point mutations, indels, and high-level CNAs, respectively, per sample. *TP53* (38%), *KRAS* (28%), and *EGFR* (24%) were the genes most frequent relevant alterations across the LU cohort. Alteration frequencies varied between histologic subtypes as expected. For example, high-level CNAs in *NKX2-1*, which represent the most significant focal gain in lung adenocarcinoma [35], were observed in 15 of 69 (22%) adenocarcinomas in our LU cohort but were not observed in the 21 squamous cell carcinomas ($P = .0083$, two-sided Fisher exact test). Of note, both SCCs harbored nonsense *RB1* mutations, while both carcinoid tumors lacked prioritized alterations. An integrative heat map of prioritized alterations across the LU cohort is shown in Figure 3*A*, and copy number profiles for all LU samples are shown in Figure S4.

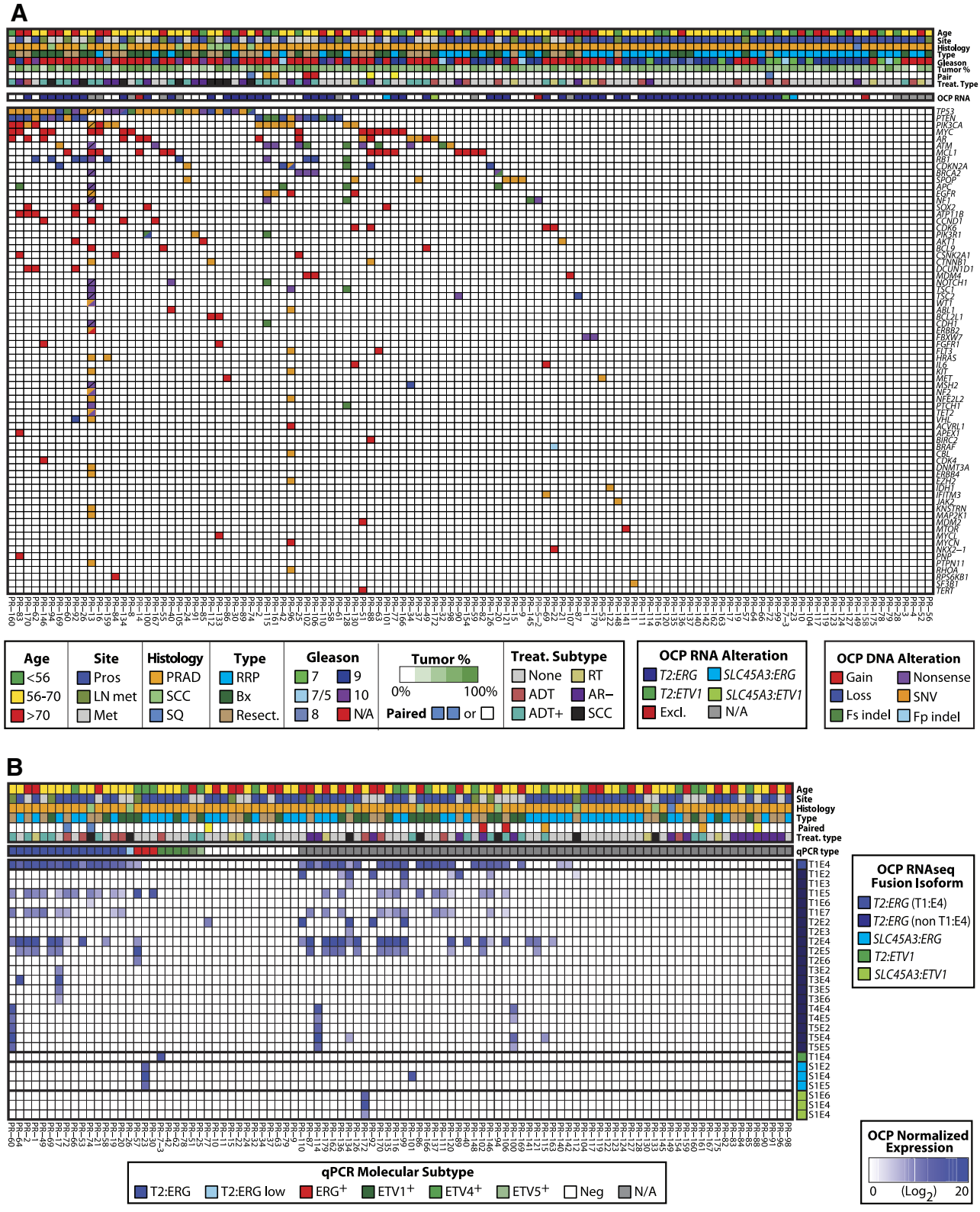
Fourteen samples in the LU cohort underwent successful diagnostic molecular testing (as in the MO cohort) for *EGFR* and/or *ALK* alterations (27 total tests). OCP demonstrated 100% sensitivity and specificity for *EGFR* alterations in these samples; in LU-49, two somatic *EGFR* hotspot GoF mutations that were not assayed for through molecular testing were identified by OCP (p.S768I, 37% variant allele frequency; p.G719C, 35% variant allele frequency). All three samples with *EML4:ALK* fusions by OCP (LU-1, LU-30, and LU-150) harbored *ALK* rearrangements by FISH. In addition, in LU-61, an adenocarcinoma lacking other actionable alterations, we identified an *EZR:ROS1* fusion (exon 10 of *EZR* fused to exon 34 of *ROS1*). As shown in Figure 3*B*, all LU samples with detectable *ALK* and *ROS1* fusions by targeted RNA-seq, as well as MO-100 and MO-106 (*EML4:ALK* fusion positive as described above), showed 3′/5′ expression imbalance in the involved 3′ partner. In total, across the LU and MO cohorts, six of the eight (75%) samples with the greatest 3′/5′ *ALK* expression imbalance by OCP harbored *ALK* rearrangements by FISH (Figure 3*B*), supporting the complementary information provided by this approach. Additional assessment of OCP performance in cases with known gene fusions is provided in the PR cohort below.

OCP Validation and Application in a Prostate Cancer Cohort

Lastly, we applied the OCP to a cohort of 118 retrospectively identified prostate cancers for validation and application. The PR cohort was selected to enrich for samples poorly represented in standard frozen tissue cohorts, with the 116 informative samples (from 114 patients) including 35 diagnostic biopsy samples, 20 samples from individuals ≤ 55 years of age, and 50 previously treated samples (Figure 4*A* and Table S4). After filtering to the predefined Oncomine variants, we identified an average of 1.0, 0.2, and 1.2 relevant somatic point mutations, indels, and high-level CNAs, respectively, per sample. Besides *T2:ERG* gene fusions (see below), the genes most frequently harboring relevant alterations in the PR cohort were *TP53* (27%), *PTEN* (18%), and *ATM* (11%). An integrative heat map of prioritized alterations across the PR cohort is shown in Figure 4*A*, and copy number profiles for all PR samples are shown in Figure S5.

Approximately 40% to 60% of prostate cancers harbor recurrent gene fusions, typically involving 5′ androgen–regulated genes fused to 3′ ETS transcription factor family members, with the most common fusion being *TMPRSS2*(*T2*):*ERG* [36,37]. The RNA component of the OCP is designed to detect recurrent gene fusions in prostate cancer through inclusion of forward primers in known 5′ fusion partners (including *TMPRSS2*, *SLC45A3*, and *C150RF21*) and reverse primers in known 3′ fusion partners (including *ERG*, *ETV1*, *ETV4*, and *BRAF*). Across the PR samples, OCP detected ETS gene fusions in 58 of 100 (58%) samples, as shown in Figure 4*B*. Of note, among the 54 *T2*:*ERG* fusion–positive samples, we identified a median of three unique fusion isoforms (range 1-9) due to combinatorial priming allowed by targeted RNA-seq, consistent

Figure 3. OCP identified relevant somatic alterations, including gene fusions, in a lung cancer cohort. (A) We applied the OCP to a retrospective cohort of FFPE lung tumors selected to represent diverse pathology (LU cohort). All OCP defined relevant alterations from the RNA (in header) and DNA components of the OCP for the 101 informative samples are shown in the heat map. Clinicopathologic information is given in the header according to the legend [Met, metastasis; LUSQ, squamous cell carcinoma; ADSQ, adenosquamous carcinoma; BAC, bronchioloalveolar carcinoma (adenocarcinoma *in situ* or well-differentiated lepidic predominant adenocarcinoma); SCC, small cell carcinoma; Resect., resection; Bx, biopsy]. All 101 informative lung samples were included in OCP RNA analysis. Samples excluded from copy number analysis due to noisy profiles are shown in italics. (B) In addition to primers for pan-cancer prioritized 5′ and 3′ gene fusion partners, OCP includes 5′ and 3′ amplicons for *ALK*, *ROS1*, and *RET* to identify 3′/5′ expression imbalance indicative of gene fusions. For all lung tumors (including those from the MO cohort), normalized OCP RNA-seq expression of gene fusions involving *ALK* (red) and *ROS1* (green) is plotted. No fusions involving *RET* were detected. Corresponding normalized 3′/5′ expression imbalance for *ALK* (top panel) and *ROS1* (middle panel) for each sample is plotted. *ALK* rearrangement positive (bolded red), negative (blue), or untested (gray) samples by molecular testing are indicated.



with the known expression of multiple *T2:ERG* splice variants in fusion-positive tumors, including those reported to drive aggressive disease [38].

Thirty-seven informative PR samples were previously assessed using an integrative DNA/RNA molecular profiling assay (MiPC) based on Haloplex target capture and Ion Torrent NGS coupled with qRT-PCR [30], providing an opportunity for additional OCP validation. Using automated variant calling and filtering, OCP profiling demonstrated 97% sensitivity (29 of 30) for detecting commonly targeted somatic variants (from 37 samples assessed by both approaches) with highly concordant observed variant allele frequencies (Table S13). High-level CNAs in 34 genes targeted by both OCP and MiPC were also strongly correlated (Pearson r: 0.95; $P < .001$, [30]).

The 37 PR samples assessed by the OCP were also assessed by the RNA component of MiPC, which used a validated TaqMan qRT-PCR assay for *T2:ERG* (exon 1 of *TMPRSS2* fused to exon 4 of *ERG*, designated as T1E4) and 3′ TaqMan expression assays for *ERG*, *ETV1*, *ETV4*, and *ETV5* expression [30], with outlier expression of these genes indicative of gene fusions. We observed 100% concordance for *T2:ERG* isoform T1E4 expression by OCP and MiPC (Figure 4*B*). Importantly, in the three cases identified as *ERG* expression outliers (without *T2:ERG* isoform T1E4 expression) by MiPC, we identified an *SLC45A3:ERG* gene fusion in PR-23 (three detected fusion isoforms) and expression of non-T1E4 *T2:ERG* isoforms in PR-30 and PR-57 (PR-57 had *T2:ERG* T1E4 fusion reads detectable at <1/10,000th of non-T1E4 reads). Additionally, by OCP, we detected a *TMPRSS2:ETV1* gene fusion (supported by three fusion isoforms) in one of the four samples with *ETV1* outlier expression by MiPC (PR-7-3). No fusions were detected involving *ETV4* or *ETV5* in the PR cohort, although 2 of the 37 samples profiled previously by MiPC harbored *ETV4* or *ETV5* outlier expression, consistent with fusions involving 5′ partners not targeted by the OCP. Taken together, with results from the MO and LU cohorts described above, these results support the ability of targeted RNA-seq to identify isoform-specific gene fusions through combinatorial priming and suggest that inclusion of 5′/3′ expression amplicons (as for lung fusions) may improve detection of fusions involving novel 5′ partners.

The inclusion of a large number of treated samples in the PR cohort enabled comparisons related to treatment status and unique histology/immunophenotype post-treatment (i.e., prostatic neuroendocrine/SCC and samples with no/low canonical ardrogen receptor (AR) signaling by IHC [AR$^-$]). For example, although *TP53* was the most frequently altered gene (besides *ERG*) in the PR cohort, *TP53* alteration frequency varied significantly across sample types, from 8.4% (6 of 71) of untreated or single modality treated samples [androgen deprivation (ADT) or radiation therapy (XRT)] to 100% of prostatic SCC (Table S14, $P < .001$). Likewise, *ATM* alteration frequency varied across treatment subtypes, with 7 of 22 (32%) of samples treated with ADT + XRT and/or chemotherapy [ADT+] harboring *ATM* alterations compared to 0 of 8 (0%) of SCCs (Table S15, $P = .14$). Robust prostate cancer molecular subtypes have been identified, including those defined by ETS gene fusions, *SPOP* hotspot mutations, and rare alterations (i.e., fibroblast growth factor receptor [FGFR] or RAF family fusions) [36]. Of interest, PR-122 harbored an *IDH1* R132 hotspot mutation (at 18% variant allele frequency) but lacked ETS gene fusions, *SPOP* mutations, or other prioritized alterations (Figures 4*A* and S6*A*). Assessment of the current PR cohort combined with 353 prostate cancer samples in the cBioPortal database identified *IDH1* R132 mutations in 6/453 (1%) prostate cancers, all of which lacked ETS gene fusions or *SPOP* mutations ($P = .004$, Fisher exact test; Figure S6*B* and Table S16), supporting *IDH1* mutations as defining a unique prostate cancer molecular subtype.

Lastly, OCP allowed us to assess paired samples that can inform on molecular correlates of disease progression, which is particularly challenging in prostate cancer given the long follow-up typically required to obtain sequential progressive specimens and the lack of routine biopsy confirmation of metastatic disease. In the PR cohort, PR-77 represents a primary, untreated Gleason score 9, pT3b N0 prostatectomy sample, while PR-88 is a paired urinary bladder tumor resected 4 years later after ADT, XRT, and docetaxel chemotherapy with AR$^-$ phenotype. Both samples showed focal prioritized *MCL1* and *MYC* amplifications (and non-prioritized high-level *BRCA1* amplification), consistent with clonality; however, a *TMPRSS2:ERG* fusion (exon T2E2) was identified by the OCP RNA-seq panel exclusively in PR-77, consistent with the AR$^-$ phenotype in PR-88 (Figure S7*A*). In contrast, PR-88, the AR$^-$ metastasis, uniquely harbored prioritized *AR* amplification and *CDKN2A* deletion, as well as a *CTNNB1* (*β-catenin*) GoF mutation (S37C, variant allele frequency 10%). Of note, no read support for *CTNNB1* S37C was present in PR-77, despite >5000 covering reads. Likewise, PR-160, a post-therapy (ADT + chemotherapy) epidural metastasis resected after rapid progression in a man who presented with metastatic disease at the age of 49, harbored a focal, prioritized *CTNNB1* amplification, which was not present in a pre-treatment, diagnostic prostate biopsy specimen that shared other clonal alterations with PR-160 (Figure S7*B*). These results demonstrate the utility of OCP for identifying alterations associated with treatment resistance through profiling pre-treatment/post-treatment limiting FFPE specimens.

Actionability Assessment

An important component of the OCP is a knowledge base of therapies and clinical trials associated with the predefined potential actionable variants targeted by the NGS assay. Potential therapeutic strategy prioritization for each OCP assessed sample is based on histologic cancer type and level of evidence associated with the potential actionability of each variant (FDA-approved agent, within cancer-type NCCN guideline, outside cancer-type NCCN guideline, and biomarker directed/informed clinical trials; see Table S5). In cases with multiple potential actionable variants, potential treatment strategies are prioritized, including consideration of detected variants that preclude treatment strategies based on other identified variants (i.e., *KRAS* mutations and potential EGFR inhibitor–based treatment in colorectal adenocarcinoma).

Figure 4. Application of OCP to a prostate cancer cohort identifies variable alterations across histologic and treatment subtypes and confirms isoform-specific gene fusion detection. (A) We applied the OCP to a retrospective cohort of aggressive FFPE prostate cancers. All OCP-defined relevant alterations from the RNA (in the header) and DNA components of the OCP for the 116 informative samples are shown in the heat map. Clinicopathologic information is given in the header according to the legend (Met, metastasis; Pros., prostate; LN met, lymph node metastasis; PRAD, prostatic adenocarcinoma; SCC, small cell carcinoma; SQ, squamous differentiation; RRP, radical prostatectomy). For treatment subtype, ADT = prior androgen deprivation therapy, XRT = radiation therapy, ADT+ = ADT plus XRT and/or chemotherapy, AR− = no (or reduced) AR signaling as indicated by no/focal prostate specific antigen (PSA) staining. Samples excluded from or not sequenced in OCP RNA analysis are indicated as in the legend. (B) The RNA component of the OCP contains forward primers in known 5′ fusion partners and reverse primers in known 3′ fusion partners for recurrent gene fusions in prostate cancer. Normalized $\log_2$ read counts for indicated gene fusion isoforms are indicated in each cell according to the color scale, with individual fusions indicated by the color blocks (right) and fusion isoforms named by the exon junctions of the involved genes (e.g., *T2:ERG* T1E4 indicates a fusion junction of *TMPRSS2* exon 1 and *ERG* exon 4). qRT-PCR was previously performed on a subset of these cases, as indicated in qPCR type. *T2:ERG* T1E4 status (including low expression), and *ERG* outlier expression without T1E4 isoform detection (ERG$^+$), *ETV1* (ETV1$^+$), *ETV4* (ETV4$^+$), or *ETV5* (ETV5$^+$) are indicated in the header. Samples without any of these alterations (Neg) or not tested (N/A) by qPCR are indicated.

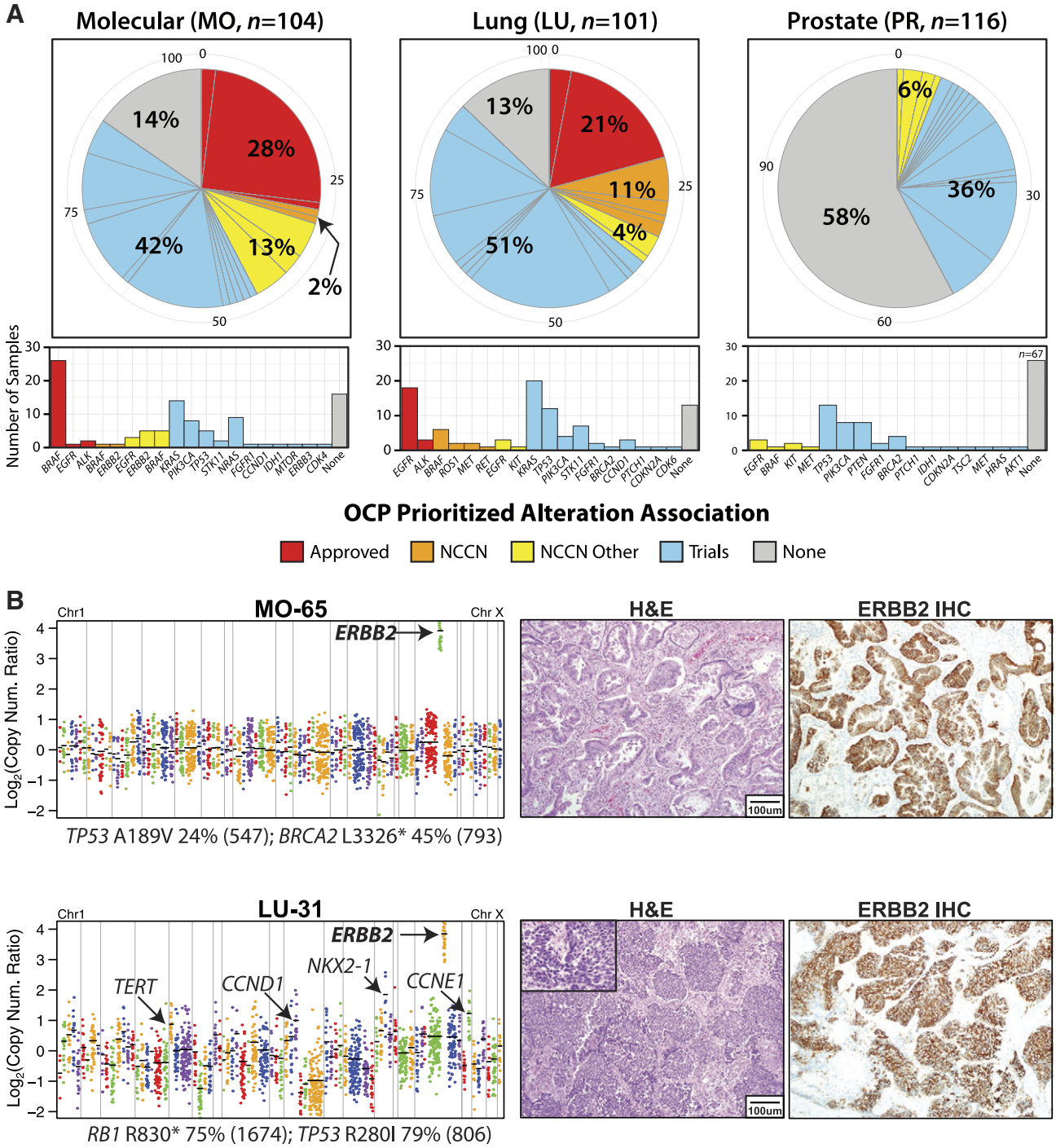


Figure 5. Automated treatment prioritization by OCP identifies relevant alterations beyond routine molecular testing. (A) For each OCP assessed cohort, the breakdown of the highest prioritized alteration per sample is shown, according to whether the alteration is associated with 1) FDA-approved therapies (red), 2) therapies within NCCN indications (orange), 3) therapies outside that specific cancer type's NCCN indication (yellow), or 4) clinical trial entry requirements (blue). This assessment incorporates variants precluding treatment strategies based on other identified variants but does not prioritize variants that only exclude approved agents. Individual prioritized alterations are indicated as slices of each pie and are shown in the histogram. (B) Integrative OCP profiling prioritized high-level *ERBB2* copy gains in two lung carcinomas. Integrative OCP results are shown as in Figure 2*B* (gene fusions were not identified in either sample). OCP profiling prioritized high-level *ERBB2* copy number gains in MO-65 (top), an *EGFR*/*ALK* wild-type lung adenocarcinoma by diagnostic molecular testing, and LU-31 (bottom), a lung SCC with no previous molecular diagnostic testing. Morphology by hematoxylin and eosin staining is shown (inset of LU-31 shows typical small cell morphology). Diffuse 3+ ERBB2 protein expression was confirmed by IHC.

To assess the potential utility of the OCP in identifying treatment options, we identified the highest priority alteration for each sample assessed herein, as shown in Figure 5*A*. These analyses only include positively associated variants (i.e., *KRAS* mutations in colorectal cancer excluding EGFR inhibitors are not prioritized). In the MO cohort, OCP confirmed the presence of *BRAF*, *EGFR*, and *ALK*

alterations in 29 samples (28%), each associated with FDA-approved indications. In an additional 15 MO samples (14%), OCP identified an actionable variant that is not routinely tested for in that cancer type but which is associated with the same (n = 2) or other cancer type (n = 13) approved therapies referenced in NCCN clinical guidelines (e.g., *ERBB2*, *BRAF*, and *EGFR* alterations). These findings are especially important because emerging evidence supports benefit, in some cases substantial, to an available targeted therapy. For example, responses to the BRAF inhibitor dabrafenib in lung cancer patients with *BRAF* mutations in a phase II trial led to Breakthrough Therapy designation by the FDA, and combination trials with the mitogen activated protein kinase kinase (MEK) inhibitor trametinib—which proved superior to single-agent BRAF therapy in melanoma [39,40]—are enrolling. An additional 44 samples (42%) harbored alterations in a gene that is a positive eligibility criterion for a clinical trial involving a targeted therapy (e.g., *PIK3CA*, *NRAS*, and so on).

Likewise, in the LU cohort, OCP identified alterations associated with FDA-approved therapies, NCCN guidelines, and clinical trial eligibility in 21 (21%), 15 (15%; n = 11 same cancer; n = 4 other cancer), and 52 (51%) samples, respectively. Lastly, in the PR cohort, OCP identified alterations associated with FDA-approved therapies, NCCN guidelines, and clinical trial eligibility in 0 (0%), 7 (6%; all other cancers), and 42 (36%) samples, respectively, demonstrating that "actionable" alterations occur with variable frequency across cancers from different organs. As an example of a highly actionable alteration that is not routinely tested for in the specific cancer type (lung cancer) nor assessed by targeted NGS approaches that do not assess CNAs, OCP prioritized high-level gains in *ERBB2* in MO-86 and LU-31 (lung adenocarcinoma and lung SCC, respectively), with overexpression in both cases confirmed by IHC (Figure 5*B*).

Discussion

Here, we report the development, validation, and assessment of a highly scalable, FFPE-compatible, targeted NGS-based system to prioritize potential treatment strategies from predefined relevant somatic variants in solid tumors. To identify candidate driving somatic alterations for inclusion in the OCP, we queried genomic data from more than 700,000 tumor samples to define pan-solid tumor, recurrent driving somatic alterations through defining GoF mutations in oncogenes, LoF (and GoF) mutations in tumor suppressors, CNAs through MCR analysis, and recurrent gene fusions. These alterations were combined with a comprehensive knowledge base of currently available oncology therapeutics and clinical trials to define variants with immediate or near-term relevance. We then developed a targeted multiplexed PCR-based NGS panel compatible with limited amounts of routine FFPE tissue samples (20 ng of DNA/15 ng of RNA) to detect these variants. These nucleic acid requirements are 2- to 50-fold less than those for comprehensive capture-based precision oncology approaches [18,21]. To balance OCP panel size and clinical relevance, we excluded genes without near-term clinical actionability and only currently identified hotspots are targeted. Hence, additional genes/amplicons may be included in future OCP versions or supplemental panels to target novel relevant alterations, including treatment resistance hotspots poorly represented in most publically available profiling studies.

We validated OCP performance using more than 300 FFPE tumor specimens, including a prospective cohort of 104 samples undergoing concurrent molecular diagnostics testing for *BRAF*, *KRAS*, or *EGFR* point mutations and indels, achieving a sensitivity and specificity of 100%. We and others have previously validated the utility of multiplexed PCR-based Ion Torrent sequencing for CNA assessment [19,25,31–33] and herein confirm high-level *ERBB2* CNAs identified by OCP using IHC. OCP identified mutations and high-level CNAs were also highly concordant with results from Haloplex capture-based NGS in a subset of the PR cohort profiled by both technologies. Taken together, these results demonstrate the ability of OCP to identify these relevant classes of alterations. Frame-shifting indels in long homopolymer runs are challenging to detect with current Ion Torrent approaches (and are excluded using our filtering criteria) and multiplexed PCR approaches cannot detect large structural rearrangements; however, these alterations predominantly result in LoF alterations in tumor suppressors [41], which represent a minority of current therapeutic targets. We anticipate that our cohort and additional OCP profiled samples will enable the development of panel- and laboratory-specific error models to improve performance in homopolymer regions.

The RNA component of the OCP is designed to identify known recurrent gene fusions (through primers spanning known exon junctions) as well as fusions of *RET*, *ROS1*, and *ALK* with novel 5′ partners (or novel fusion isoforms) through 3′/5′ expression imbalance. We confirmed 100% concordance for *T2:ERG* gene fusion isoform-specific detection between OCP and a validated quantitative PCR (qPCR) assay in a subset of our PR cohort profiled by both methods, with multiple splice variants detected in the majority of fusion-positive cases. Likewise, in seven lung cancers known to harbor *ALK* rearrangements across our cohorts, OCP profiling identified *EML4:ALK* fusions in five (71%), with these five samples also showing 3′/5′ expression imbalance by OCP. MO-66 (the known *ALK* rearranged sample with fusion read support below our threshold criteria) and LU-38 (known *ALK* rearrangement without fusion read support) also showed 3′/5′ expression imbalance. Of note, in our MO cohort, we identified two additional relevant fusions [*ERC1:BRAF* in a melanoma sample negative for *BRAF* mutation (MO-17) and *TPR:NTRK1* in a colon cancer (MO-35)], validating both fusions by qPCR. Of note, *ERC1:BRAF* was not directly targeted in the OCP RNA panel design, as *ERC1* had previously only been reported as a fusion partner with *RET* [42], highlighting the utility of the combinatorial nature of targeted multiplexed PCR-based RNA-seq.

Taken together, our results validate the multiplexed PCR-based RNA-seq approach for detecting targeted gene fusions. Characterization of additional cohorts will be required to determine performance and optimal 3′/5′ expression imbalance cutoffs for *ALK*, *RET*, and *ROS1* fusions in lung cancer (and other cancer types) involving unknown partners. Likewise, we anticipate that inclusion of additional 3′/5′ expression imbalance amplicons will improve fusion detection involving other genes. Lastly, splice variant detection of non-gene fusion events, such as *AR* splice variants in prostate cancer [43,44] or alternatively spliced tyrosine kinases (e.g., *MET*) in other cancers [45,46], may also be assessed in OCP through inclusion of additional amplicons. Importantly, although comprehensive capture-based NGS approaches assessing only DNA can identify gene fusions through sequencing introns of involved genes [18], such approaches cannot detect or quantify potentially relevant splice variants.

As a demonstration of the utility of OCP for translational research, we applied this approach to a cohort of 116 prostate cancers, including 50 previously treated samples. We recapitulated known molecular subtypes and alterations with specific histology, including

the high prevalence of *TP53* alterations in prostatic SCC [47–52]. We also identified a high burden of *ATM* alterations in heavily treated patients, which can be investigated in future efforts characterizing this understudied population. Of note, through integration with previous profiling studies, we identify *IDH1* R132 mutant prostate cancer as a novel molecular subtype that lacks other subtype defining lesions. This finding is especially important as IDH1 inhibitors are now in early phase clinical trials. Lastly, two pairs of pre-treatment and post-treatment samples each demonstrated *AR* amplifications (a known adaptive response to ADT [53]) and *CTNNB1* GoF mutation/amplification exclusively in the post-treatment sample. Although activation of the WNT/CTNNB1 pathway has been identified in ADT-treated prostate cancer [29,54,55], our report is the first to demonstrate that ADT and/or subsequent chemotherapy specifically induces (or selects) for *CTNNB1* amplification/activating mutation, supporting a functional role in treatment resistance.

The OCP is compatible with routine Ion Torrent workflows, and the DNA/RNA components of the OCP can be combined for template preparation and concurrent PGM sequencing on a single Ion Torrent 318 chip in a standard ~4 hour PGM run, with the potential for higher throughput using the Ion Torrent Proton. Although complete analytic validation will need to be performed in individual laboratories, we demonstrate highly concordant results with typical specimens sent for molecular diagnostic testing as well as molecular standards (performance with downsampled reads is shown in Figure S8). Hence, this approach provides a rapid, highly scalable approach requiring small amounts of routine tissue specimens with performance comparable to previous multiplexed PCR-based Ion Torrent panels assessing DNA alterations [16,19,22], capture-based approaches[18,56], and anchored multiplexed PCR-based NGS [24]. A critical component of the OCP is a highly automated analysis pipeline that links to a knowledge base of potential treatment options, facilitated by predefining the actionable cancer genome before panel development. As shown through our actionability assessment, a significant number of samples currently harbor relevant alterations that are identifiable using our approach. As clinical sequencing efforts and expertise become more prevalent, a key advantage of the OCP is the potential for integration into multiple independent institutions (rather than a single centralized testing center), enabling valuable direct involvement from molecular biologists, pathologists, and oncologists. Taken together, the highly scalable assay and framework described herein may have utility in future oncology precision medicine approaches, such as the NCI Match Trial, where multiple sites will sequence 3000 cancer samples using the OCP.

Acknowledgements

The authors thank Mandy Davis and Angela Fullen for technical assistance.

Appendix A. Supplementary Materials

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neo.2015.03.004.

References

[1] Garraway LA, Verweij J, and Ballman KV (2013). Precision oncology: an overview. *J Clin Oncol* **31**, 1803–1805.

[2] Mendelsohn J (2013). Personalizing oncology: perspectives and prospects. *J Clin Oncol* **31**, 1904–1911.

[3] Arteaga CL and Baselga J (2012). Impact of genomics on personalized cancer medicine. *Clin Cancer Res* **18**, 612–618.

[4] McDermott U, Downing JR, and Stratton MR (2011). Genomics and the continuum of cancer care. *N Engl J Med* **364**, 340–350.

[5] Pant S, Weiner R, and Marton MJ (2014). Navigating the rapids: the development of regulated next-generation sequencing-based clinical trial assays and companion diagnostics. *Front Oncol* **4**, 78.

[6] Olsen D and Jorgensen JT (2014). Companion diagnostics for targeted cancer drugs—clinical and regulatory aspects. *Front Oncol* **4**, 105.

[7] Parkinson DR, Johnson BE, and Sledge GW (2012). Making personalized cancer medicine a reality: challenges and opportunities in the development of biomarkers and companion diagnostics. *Clin Cancer Res* **18**, 619–624.

[8] Mass RD, Press MF, Anderson S, Cobleigh MA, Vogel CL, Dybdal N, Leiberman G, and Slamon DJ (2005). Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. *Clin Breast Cancer* **6**, 240–246.

[9] Druker BJ (2008). Translation of the Philadelphia chromosome into therapy for CML. *Blood* **112**, 4808–4817.

[10] Kris MG, Johnson BE, Berry LD, Kwiatkowski DJ, Iafrate AJ, Wistuba II, Varella-Garcia M, Franklin WA, Aronson SL, and Su PF, et al (2014). Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. *JAMA* **311**, 1998–2006.

[11] Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, and Getz G (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501.

[12] Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, and Roberts SA, et al (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218.

[13] Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz Jr LA, and Kinzler KW (2013). Cancer genome landscapes. *Science* **339**, 1546–1558.

[14] Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang CZ, Wala J, and Mermel CH, et al (2013). Pan-cancer patterns of somatic copy number alteration. *Nat Genet* **45**, 1134–1140.

[15] Tomlins SA, Wyngaard P, Khazanov N, Williams P, Bowden E, Sadis S, and Rhodes DR (2012). Analysis of 2,700 cancer exomes to identify novel cancer drivers and therapeutic opportunities. *Eur J Cancer* **48**, 134.

[16] Beadling C, Neff TL, Heinrich MC, Rhodes K, Thornton M, Leamon J, Andersen M, and Corless CL (2013). Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping. *J Mol Diagn* **15**, 171–176.

[17] Borad MJ, Champion MD, Egan JB, Liang WS, Fonseca R, Bryce AH, McCullough AE, Barrett MT, Hunt K, and Patel MD, et al (2014). Integrated genomic characterization reveals novel, therapeutically relevant drug targets in FGFR and EGFR pathways in sporadic intrahepatic cholangiocarcinoma. *PLoS Genet* **10**, e1004135.

[18] Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, Schnall-Levin M, White J, Sanford EM, and An P, et al (2013). Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol* **31**, 1023–1031.

[19] Grasso C, Butler T, Rhodes K, Quist M, Neff TL, Moore S, Tomlins SA, Reinig E, Beadling C, and Andersen M, et al (2015). Assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. *J Mol Diagn* **17**, 53–63.

[20] Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, Marko AC, Sanford T, Buddavarapu K, Krosting J, and Garmire L, et al (2013). Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. *J Mol Diagn* **15**, 234–247.

[21] Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, Cao X, Kalyana-Sundaram S, Sam L, Balbin OA, and Quist MJ, et al (2011). Personalized oncology through integrative high-throughput sequencing: a pilot study. *Sci Transl Med* **3**, 111ra121.

[22] Singh RR, Patel KP, Routbort MJ, Reddy NG, Barkoh BA, Handal B, Kanagal-Shamanna R, Greaves WO, Medeiros LJ, and Aldape KD, et al (2013). Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. *J Mol Diagn* **15**, 607–622.

[23] Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, Marlow S, Jane-Valbuena J, Friedrich DC, Kryukov G, and Carter SL, et al (2014). Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med* **20**.

[24] Zheng Z, Liebers M, Zhelyazkova B, Cao Y, Panditi D, Lynch KD, Chen J, Robinson HE, Shim HS, and Chmielecki J, et al (2014). Anchored multiplex PCR for targeted next-generation sequencing. *Nat Med* **20**, 1479–1484.

[25] Hoogstraat M, Hinrichs JW, Besselink NJ, Radersma-van Loon JH, de Voijs CM, Peeters T, Nijman IJ, de Weger RA, Voest EE, and Willems SM, et al (2015). Simultaneous detection of clinically relevant mutations and amplifications for routine cancer pathology. *J Mol Diagn* **17**, 10–18.

[26] Samorodnitsky E, Datta J, Jewell BM, Hagopian R, Miya J, Wing MR, Damodaran S, Lippus JM, Reeser JW, and Bhatt D, et al (2015). Comparison of custom capture for targeted next-generation DNA sequencing. *J Mol Diagn* **17**, 64–75.

[27] Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, and Ward S, et al (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805–D811.

[28] Rhodes DR, Kalyana-Sundaram S, Mahavisno V, Varambally R, Yu J, Briggs BB, Barrette TR, Anstet MJ, Kincead-Beal C, and Kulkarni P, et al (2007). Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166–180.

[29] Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, Quist MJ, Jing X, Lonigro RJ, and Brenner JC, et al (2012). The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243.

[30] Grasso CS, Cani AK, Hovelson DH, Quist MJ, Douville NJ, Yadati V, Amin AM, Nelson PS, Betz BL, and Liu CJ, et al (2015). Integrative molecular profiling of routine clinical prostate cancer specimens. *Ann Oncol* [Epub ahead of print].

[31] Warrick JI, Hovelson DH, Amin A, Liu C-J, Cani AK, McDaniel AS, Yadati V, Quist MJ, Weizer AZ, and Brenner CJ, et al (2015). Tumor evolution and progression in multifocal and paired non-invasive/invasive urothelial carcinoma. *Virchows Arch* **466**, 297–311.

[32] Cani AK, Hovelson DH, McDaniel AS, Sadis S, Haller MJ, Yadati V, Amin AM, Bratley J, Bandla S, and Williams PD, et al (2015). Next-Gen sequencing exposes frequent MED12 mutations and actionable therapeutic targets in phyllodes tumors. *Mol Cancer Res* [Epub ahead of print].

[33] McDaniel AS, Zhai Y, Cho KR, Dhanasekaran SM, Montgomery JS, Palapattu G, Siddiqui J, Morgan T, Alva A, and Weizer A, et al (2014). HRAS mutations are frequent in inverted urothelial neoplasms. *Hum Pathol* **45**, 1957–1965.

[34] Palanisamy N, Ateeq B, Kalyana-Sundaram S, Pflueger D, Ramnarayanan K, Shankar S, Han B, Cao Q, Cao X, and Suleman K, et al (2010). Rearrangements of the RAF kinase pathway in prostate cancer, gastric cancer and melanoma. *Nat Med* **16**, 793–798.

[35] Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, Lin WM, Province MA, Kraja A, and Johnson LA, et al (2007). Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**, 893–898.

[36] Tomlins SA (2013). Molecular clues assist in the cancer clinic. *Sci Transl Med* **5**, 193fs126.

[37] Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, and Kuefer R, et al (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648.

[38] Wang J, Cai Y, Ren C, and Ittmann M (2006). Expression of variant TMPRSS2/ERG fusion messenger RNAs is associated with aggressive prostate cancer. *Cancer Res* **66**, 8347–8351.

[39] Robert C, Karaszewska B, Schachter J, Rutkowski P, Mackiewicz A, Stroiakovski D, Lichinitser M, Dummer R, Grange F, and Mortier L, et al (2015). Improved overall survival in melanoma with combined dabrafenib and trametinib. *N Engl J Med* **372**, 30–39.

[40] Long GV, Stroyakovskiy D, Gogas H, Levchenko E, de Braud F, Larkin J, Garbe C, Jouary T, Hauschild A, and Grob JJ, et al (2014). Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. *N Engl J Med* **371**, 1877–1888.

[41] Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, and Ghandi M, et al (2013). Punctuated evolution of prostate cancer genomes. *Cell* **153**, 666–677.

[42] Nakata T, Kitamura Y, Shimizu K, Tanaka S, Fujimori M, Yokoyama S, Ito K, and Emi M (1999). Fusion of a novel gene, ELKS, to RET due to translocation t(10;12)(q11;p13) in a papillary thyroid carcinoma. *Genes Chromosomes Cancer* **25**, 97–103.

[43] Lu C and Luo J (2013). Decoding the androgen receptor splice variants. *Transl Androl Urol* **2**, 178–186.

[44] Antonarakis ES, Lu C, Wang H, Luber B, Nakazawa M, Roeser JC, Chen Y, Mohammad TA, Fedor HL, and Lotan TL, et al (2014). AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med* **371**, 1028–1038.

[45] Kong-Beltran M, Seshagiri S, Zha J, Zhu W, Bhawe K, Mendoza N, Holcomb T, Pujara K, Stinson J, and Fu L, et al (2006). Somatic mutations lead to an oncogenic deletion of met in lung cancer. *Cancer Res* **66**, 283–289.

[46] Druillennec S, Dorard C, and Eychene A (2012). Alternative splicing in oncogenic kinases: from physiological functions to cancer. *J Nucleic Acids* **2012**, 639062.

[47] Barbieri CE and Tomlins SA (2014). The prostate cancer genome: perspectives and potential. *Urol Oncol* **32**(53), e15–e22.

[48] Beltran H and Rubin MA (2013). New strategies in prostate cancer: translating genomics into the clinic. *Clin Cancer Res* **19**, 517–523.

[49] Brenner JC, Chinnaiyan AM, and Tomlins SA (2013). ETS fusion genes in prostate cancer. In: Tindall DJ, editor. Vol. Protein Reviews. New York: Springer New York; 2013. p. 139–183.

[50] Rubin MA, Maher CA, and Chinnaiyan AM (2011). Common gene rearrangements in prostate cancer. *J Clin Oncol* **29**, 3659–3668.

[51] Beltran H, Tomlins S, Aparicio A, Arora V, Rickman D, Ayala G, Huang J, True L, Gleave ME, and Soule H, et al (2014). Aggressive variants of castration-resistant prostate cancer. *Clin Cancer Res* **20**, 2846–2850.

[52] Chen H, Sun Y, Wu C, Magyar CE, Li X, Cheng L, Yao JL, Shen S, Osunkoya AO, and Liang C, et al (2012). Pathogenesis of prostatic small cell carcinoma involves the inactivation of the P53 pathway. *Endocr Relat Cancer* **19**, 321–331.

[53] Bluemn EG and Nelson PS (2012). The androgen/androgen receptor axis in prostate cancer. *Curr Opin Oncol* **24**, 251–257.

[54] Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, Dumpit RF, Coleman I, Ng SB, Salipante SJ, and Rieder MJ, et al (2011). Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc Natl Acad Sci U S A* **108**, 17087–17092.

[55] Rajan P, Sudbery IM, Villasevil ME, Mui E, Fleming J, Davis M, Ahmad I, Edwards J, Sansom OJ, and Sims D, et al (2014). Next-generation sequencing of advanced prostate cancer treated with androgen-deprivation therapy. *Eur Urol* **66**, 32–39.

[56] Beltran H, Yelensky R, Frampton GM, Park K, Downing SR, Macdonald TY, Jarosz M, Lipson D, Tagawa ST, and Nanus DM, et al (2013). Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. *Eur Urol* **63**, 920–926.