Published in final edited form as:

Sci Transl Med. 2015 April 15; 7(283): 283ra53. doi:10.1126/scitranslmed.aaa7161.

Personalized genomic analyses for cancer mutation discovery and interpretation

Siân Jones¹, Valsamo Anagnostou², Karli Lytle¹, Sonya Parpart-Li¹, Monica Nesselbush¹, David R. Riley¹, Manish Shukla¹, Bryan Chesnick¹, Maura Kadan¹, Eniko Papp², Kevin G. Galens¹, Derek Murphy¹, Theresa Zhang¹, Lisa Kann¹, Mark Sausen¹, Samuel V. Angiuoli¹, Luis A. Diaz Jr.², and Victor E. Velculescu^{2,*}

¹Personal Genome Diagnostics, Baltimore, MD 21224, USA

²The Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

Abstract

Massively parallel sequencing approaches are beginning to be used clinically to characterize individual patient tumors and to select therapies based on the identified mutations. A major question in these analyses is the extent to which these methods identify clinically actionable alterations and whether the examination of the tumor tissue alone is sufficient or whether matched normal DNA should also be analyzed to accurately identify tumor-specific (somatic) alterations. To address these issues, we comprehensively evaluated 815 tumor-normal paired samples from patients of 15 tumor types. We identified genomic alterations using next-generation sequencing of whole exomes or 111 targeted genes that were validated with sensitivities >95% and >99%, respectively, and specificities >99.99%. These analyses revealed an average of 140 and 4.3 somatic mutations per exome and targeted analysis, respectively. More than 75% of cases had somatic alterations in genes associated with known therapies or current clinical trials. Analyses of matched normal DNA identified germline alterations in cancer-predisposing genes in 3% of

*Corresponding author. velculescu@jhmi.edu.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/7/283/283ra53/DC1

Fig. S1. Bioinformatic approach to classify somatic and germline mutations on the basis of the consequence of the alteration.

Fig. S2. Bioinformatic approach to classify somatic and germline mutations based on the affected protein domain.

Table S1. Genes analyzed in the targeted approach (provided in a separate Excel file).

Table S2. Summary of sequencing statistics (provided in a separate Excel file).

Table S3. Summary of performance characteristics of whole-exome and targeted analyses (provided in a separate Excel file).

Table S4. Characteristics of the tumor and normal samples (provided in a separate Excel file).

Table S5. Fraction of cases with somatic mutations in actionable genes (provided in a separate Excel file).

Table S6. Fraction of cases with evidence for clinical actionability in different tumor types (provided in a separate Excel file).

Table S7. Hereditary cancer predisposition genes (provided in a separate Excel file).

Table S8. Putative germline predisposing mutations (provided in a separate Excel file).

Table S9. Germline false-positive mutations in actionable genes (provided in a separate Excel file).

Competing interests: Some of the work described in this publication is included in a pending patent application. L.A.D. and V.E.V. are cofounders of Personal Genome Diagnostics and are members of its Scientific Advisory Board and Board of Directors. L.A.D. and V.E.V. own Personal Genome Diagnostics stock, which is subject to certain restrictions under university policy. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict of interest policies.

Author contributions: S.J. designed the study, performed analyses, and wrote the paper. V.A., K.L., S.P.-L., M.N., D.R.R., M.K., E.P., K.G.G., D.M., and T.Z. performed analyses; M.S. and S.V.A. performed analyses and wrote the paper. M.S., B.C., and L.K. performed experiments. L.A.D. and V.E.V. designed the study, performed analyses, provided funding, and wrote the paper.

patients with apparently sporadic cancers. In contrast, a tumor-only sequencing approach could not definitively identify germline changes in cancer-predisposing genes and led to additional false-positive findings comprising 31% and 65% of alterations identified in targeted and exome analyses, respectively, including in potentially actionable genes. These data suggest that matched tumor-normal sequencing analyses are essential for precise identification and interpretation of somatic and germline alterations and have important implications for the diagnostic and therapeutic management of cancer patients.

INTRODUCTION

High-complexity genomic analyses are changing the diagnostic landscape of oncology (1–7). Therapies targeting specific genetic alterations can be safer and more effective than traditional chemotherapies when used in an appropriate patient population (8). This has been successfully demonstrated for a number of therapeutics targeting the protein products of specific genes that are altered in human cancer, including the use of imatinib in chronic myeloid leukemias carrying the *BCR-ABL* fusion, trastuzumab in *ERBB2* (*HER-2/neu*) amplified breast cancer, and vemurafenib in *BRAF*-mutated melanoma. Molecular alterations have also been shown to have a predictive or prognostic effect. For example, mutations at codons 12 and 13 of *KRAS* predict a poor response to anti-EGFR (epidermal growth factor receptor) monoclonal antibodies such as cetuximab and panitumumab, so the use of these drugs is contraindicated in colorectal cancer patients with such mutations (9). Glioblastoma patients with *IDH1*-mutated tumors have an increased overall survival compared to those without such changes (10). In addition to established therapies, off-label indications and drugs in clinical trials can be used with knowledge of alterations in specific genes. Because the mutations driving each tumor are unique, identifying the specific mutations in each patient's cancer is critical for the development of a personalized treatment plan that takes advantage of the growing number of targeted therapies.

Each tumor contains inherited (germline) and tumor-specific (somatic) variants. Somatic alterations in oncogenes and tumor suppressors drive the development and growth of the tumor and are typically the targets of personalized therapies. Sequencing and comparison of matched normal DNA to tumor DNA from an affected individual would theoretically allow for accurate identification and subtraction of germline alterations from somatic changes. However, this method is not routinely used in cancer diagnostic assays, including next-generation sequencing approaches, where only tumor DNA is assessed, likely as a result of logistical difficulties in obtaining a blood or saliva sample, increased cost, and an underappreciation of the potential value of the matched normal (2, 11–13). Additionally, a major question in the development and progression of human cancer has been the contribution of germline alterations to cancer predisposition. Although estimates have been proposed in specific tumor types (14, 15), the comprehensive examination of cancer-predisposing alterations in apparently sporadic cancer patients has not been investigated.

To evaluate the clinical use of large-scale cancer genome analyses that incorporate these aspects, we performed whole-exome and targeted next-generation sequencing analyses in tumor and normal samples from cancer patients. We analyzed matched tumor and normal

data together as well as separately for somatic mutation detection, potential clinical actionability, and identification of predisposing alterations.

RESULTS

Overview of the approach

To systematically assess somatic alterations in tumor samples, we designed capture probes for the targeted analysis of a set of 111 clinically relevant genes (table S1) and sequenced these regions or the complete set of coding genes (20,766 genes) using next-generation sequencing approaches (Fig. 1). These data were aligned to the human reference sequence and annotated using the Consensus Coding DNA Sequences (CCDS), RefSeq, and Ensembl databases. Tumor and normal data were compared to identify somatic and germline alterations using the VariantDx software pipeline, focusing on single-base substitutions as well as small insertions and deletions. Stringent criteria were used to ensure sufficient coverage at analyzed bases and to exclude mapping and sequencing errors (table S2). All candidate somatic alterations were visually inspected to remove remaining artifactual changes. Analysis of samples using both whole-exome Sanger and next-generation sequencing was used to demonstrate that the next-generation sequencing and bioinformatic approaches were able to detect somatic mutations in frozen and formalin-fixed paraffin-embedded (FFPE) tumor tissues with high sensitivity and specificity and to accurately distinguish between somatic and germline alterations (table S3).

Clinical actionability of targeted and exome analyses

Using the above approach, we analyzed matched tumor and normal specimens from 815 patients, with the tumor types indicated in table S4. A total of 105,672 somatic alterations were identified, with an average of 4.34 somatic mutations (range, 0 to 29) in the targeted analyses and an average of 140 somatic alterations (range, 1 to 6219) in the exome analyses. The number of somatic alterations in various tumor types was largely consistent with previous analyses of cancer exomes (10, 16–30). To explore whether genetic alterations may be useful clinically, we investigated whether mutant genes observed in individual cases may be clinically actionable using existing or investigational therapies. We examined altered genes that were associated with (i) U.S. Food and Drug Administration (FDA)–approved therapies for oncologic indications, (ii) therapies in published prospective clinical studies, and (iii) ongoing clinical trials for patients with tumor types analyzed. Through these analyses, we identified somatic alterations in genes with potentially actionable consequences in 580 of the 753 patients analyzed (77%) (Fig. 2 and tables S5 and S6). Some tumor types, such as colorectal cancer and melanoma, had a much higher fraction of actionable changes than others. More than 90% of genes with potentially actionable alterations were mutated in <5% of cases, suggesting that actionable changes are predominantly different among cancer patients (table S5). Although the fraction of patients who had at least one actionable alteration was high, most of the actionable changes were associated with current clinical trials (67%) rather than established or investigational therapies (33%).

Identification of patients with putative germline cancer predisposition mutations

In addition to the detection of somatic alterations, we assessed whether our analyses identified cancer-predisposing changes in the genomes of apparently sporadic cancer patients. To perform this analysis, we examined a set of 85 genes associated with known cancer predisposition syndromes (table S7) in DNA from blood, saliva, or unaffected tissue samples from the 815 cancer patients. To conservatively identify protein-altering changes in these genes, we focused on truncating alterations, including insertions or deletions resulting in a frameshift, splice site changes, and nonsense alterations. Through these analyses, we identified 27 of the 815 patients (~3%) with truncating alterations in these genes (table S8). All but one of these cases were not previously known to have a cancer-predisposing alteration in their germ line. Fifteen mutations were predicted to be pathogenic or likely pathogenic based on previous publications. Examples of germline alterations included changes in genes in expected tumor types, such as *BRCA1* alterations in breast and ovarian cancer patients and a nonsense mutation (50Q>X) in *CDKN2A* in a melanoma case. However, less well-described examples were also detected, including *BRCA2* alterations in patients with other solid tumor types such as colorectal cancer and cholangiocarcinoma, *ATM* changes in an esophageal cancer case, *FANC* alterations in patients with a variety of tumor types, and alterations in the *BRIP1* (BRCA1 interacting protein C-terminal helicase 1) gene in a cholangiocarcinoma (800Y>X) and in an anal cancer case (624S>X).

Bioinformatic approaches for distinguishing germline and somatic mutations

Because many newly developed tests for alterations in cancer genes only examine the tumor tissue (2, 11, 13, 31), we evaluated how effective bioinformatic approaches could be in separation of somatic from germline mutations without the use of a matched normal (Fig. 1). First, we reanalyzed only the tumor data from all 58 targeted cases and 100 whole-exome cases composed of about half frozen and half FFPE samples from a representative range of tumor types. We compared these to an unmatched normal sample that had been sequenced using the same methods as for the matched normal samples. We used these data to remove common germline variants, as well as sequencing and alignment errors. All candidate alterations were visually inspected to remove any remaining artifacts. An average of 11.53 mutations (range, 3 to 34) and 1401 mutations (range, 919 to 2651) were observed in the targeted and exome cases, respectively (Fig. 3).

To identify additional germline variants in the tumors that were not present in the unmatched normals, we compared the observed tumor alterations to those in single-nucleotide polymorphism (SNP) databases (dbSNP, version 38) and filtered variants identified through the 1000 Genomes Project or other sources (32) (including 42,886,118 total candidate variants). This approach removed between 0 and 9 alterations (average, 5.25) in the targeted analyses, including all germline alterations in 10 of 58 cases. However, an average of 1.95 germline variants remained per case through the tumor-only approach, resulting in a total of 113 remaining germline changes in the 58 cases analyzed (Fig. 3). A total of 1019 mutations were removed using dbSNP filters in each of the exome cases (range, 623 to 1911), but an average of 382 mutations remained per case. A considerable proportion of the remaining germline variants affecting 48% of patients analyzed included alterations that could have been classified as potentially actionable changes (Fig. 3 and table

S9). For example, a *JAK2* (Janus kinase 2) mutation in the catalytic domain (1021Y>F), multiple missense alterations in *ERBB2*, an in-frame deletion (1508PF>P) in *TSC2* (tuberous sclerosis complex 2), and an *ALK* (anaplastic lymphoma kinase) change in the catalytic domain (1200A>V) would have been incorrectly identified through a tumor-only approach. Approved or investigational therapies targeting the altered protein products are available for these genes, including ruxolitinib for *JAK2*, neratinib for *ERBB2*, everolimus for *TSC2*, and crizotinib for *ALK* among others, that could have been inappropriately administered to patients on the basis of a tumor-only analysis. Overall, most of the cases filtered using germline databases had remaining germline alterations, with about half in potentially actionable genes.

The filtering of tumor-only data with variants present in germline databases has the potential to inadvertently remove somatic variants that may be identical to germline variants. In our targeted analyses, two somatic mutations in *PDGFRA* (478S>P) and *ATRX* (929Q>E) matched identical mutations at the nucleotide level in dbSNP and were erroneously removed by this method. The analysis of all coding genes revealed 155 somatic mutations that were removed using this approach, including the 114R>C change in the catalytic domain of the mitogen-activated protein kinase *MAPK4* and 320P>R in the transcription factor *ESX1*, which have been previously reported to be somatically mutated in skin, and thyroid and liver cancers, respectively.

To further examine detection of somatic alterations using a tumor-only approach, we attempted to separate out the somatic mutations from the remaining germline alterations after dbSNP filtering using data from the COSMIC (Catalogue of Somatic Mutations in Cancer) database (Fig. 4). Mutations in our data set were considered more likely to be somatic if tumor-specific alterations had previously been reported within the same codon of the gene. In total, 108 mutations in 47 of the cases analyzed for the targeted set of genes and 1806 mutations in the exome cases were classified into this category. This approach was useful in identifying well-characterized mutations at hotspots in oncogenes such as *KRAS*, *TP53*, and *PIK3CA* but did not identify less frequent nonsynonymous somatic mutations. Nine of the potential somatic mutations in the targeted genes that overlapped with COSMIC were present in the matched normal samples and were, in fact, germline. In the exome data, 778 germline mutations occurred at codons in which somatic mutations had been previously described. Because somatic mutations can be clustered within certain regions of a gene, we expanded our COSMIC criteria to include any mutations within five codons of the observed alteration. This increased the number of potential somatic mutations in the targeted genes by 152 to give a total of 270 (4.48 per patient) and increased the number by almost 15,000 in the exome cases to give a total of 16,731 (168 per patient). However, the specificity of the approach was substantially reduced, with 48 and 8929 of these mutations actually occurring in the matched normal in the targeted and exome genes, respectively. To determine the overall number of identical changes in the genome that had been reported as both germline variants and somatic changes through other studies, we examined the overall overlap between common dbSNP variants and the COSMIC databases. After excluding variants of known medical impact or annotated as somatic in dbSNP, we found 8606 nonsynonymous mutations that were present in both databases, of which 63 mutations were observed more

than five times in COSMIC. These analyses suggest that a considerable number of variants in the germ line may be identical to those in somatic disease such as cancer, and the number of identical variants will increase as additional somatic and germline genomes are analyzed.

Because somatic mutations in tumor suppressor genes are often truncating, we also examined this mutation type as a means to positively select for alterations in the tumor-only data after filtering of common germline variants (fig. S1). Seventy-five mutations affecting genes such as *CDH1* (splice site), *PIK3R1* (frameshift), and *ARID1B* (nonsense) in 43 cases of the targeted analyses fell into this category. However, similar to the COSMIC approach, 13 of the alterations identified as candidate somatic changes using this method were germline. In the exome cases, there were 7424 truncating mutations, but 5108 of these were germline, not somatic. Finally, we looked to see whether any of the mutations were present in the kinase domains of the proteins, because activating somatic mutations often occur in these regions. Forty-two mutations, including the *EGFR* exon 19 deletion 745KELREA>T, 542E>K in *PIK3CA*, 1021Y>F in *JAK2*, and 867E>K in *RET*, were identified in the targeted analyses, and 786 mutations, including 309P>L in *MAPK12* and 201P>S in *CDK10*, were present in the exome cases. However, four mutations in the targeted set (including the alteration in *JAK2*) and 295 alterations in the exome set were, in fact, germline (fig. S2).

Using a combination of the COSMIC, truncating alteration, and kinase domain approaches, we correctly identified 216 of 252 somatic mutations in the targeted analyses. Of the 36 somatic mutations that were missed, several occurred in genes such as *ERBB2*, *ERBB3*, and *TSC2* that are under active clinical investigation and may have been clinically actionable. These approaches also identified 71 mutations (1.22 per case) that were germline from the analyses of the matched normal samples. These included changes in actionable genes such as *ERBB2* (1128V>I), *MSH6* (726F>L), and *RET* (977S>R). Furthermore, there were 78 mutations that were not removed by the SNP filters or positively selected by the additional criteria and could not be classified by these methods. When the entire coding region was analyzed, only 8941 of the 13,314 true somatic mutations were identified, 14,734 germline variants were incorrectly categorized as likely to be tumor-specific, and the remaining 14,508 mutations including 10,135 germline alterations could not be classified.

Use of tumor cellularity in distinguishing germline from somatic mutations

As an independent measure of the somatic or germline status of a variant, we examined the fraction of mutant alleles in an analyzed tumor sample. Germline mutations would be expected to have variant allele frequency close to 50% for heterozygous and 100% for homozygous changes, whereas the proportion of variant tags for somatic mutations would depend on the extent of normal tissue contamination in the tumor sample and would presumably be lower. Of the 43 targeted cases where tumor cellularity was available, only 5 had a pathological purity of less than 50%. In these cases, all of the alterations were correctly called as somatic or germline using this method. However, in most cases, the tumor cellularity exceeded 50%. In these cases, this approach could not reliably distinguish between somatic and germline alterations, correctly identifying on average only 48% of somatic mutations. Likewise, of the assessable 16 cancer-predisposing germline variants in

these cases, only 2 could be distinguished from somatic alterations through an analysis of allele fractions.

DISCUSSION

Overall, these data provide a comprehensive analysis of the detection and interpretation of somatic and germline alterations in human cancer. These observations suggest that a high fraction of human tumors have alterations that may be clinically actionable and that a small but notable fraction of apparently sporadic cancer patients have pathogenic germline changes in cancer-predisposing genes. Additionally, these data support the notion that accurate identification and clinical interpretation of alterations benefit from analysis of both tumor and normal DNA from cancer patients.

As with all large-scale studies, our analyses have limitations. Although a variety of bioinformatic approaches were used in our analyses, additional computational methods could improve tumor-only analyses in the future. These include the use of additional germline databases, including the Exome Sequencing Project as well as other ongoing large-scale germline analyses such as the Genomics England 100,000 Genomes Project (33) and the Human Longevity sequencing initiative (34), that may not be well represented in the current dbSNP or the 1000 Genomes data sets. However, even if such approaches improve the filtering of germline changes, they are likely to increase the fraction of somatic variants that are inadvertently removed. Tools such as CHASM (cancer-specific high-throughput annotation of somatic mutations) (35), SIFT (36), PolyPhen (37), and others could potentially be used to predict whether a somatic mutation is likely a driver or passenger even in the absence of normal DNA. However, using CHASM as a final filter in our tumor-only data set approach did not identify any additional somatic mutations. Increasing the number of protein domains examined may also be helpful, although the mutations identified using this approach may be expected to be identified by the COSMIC clustering filter. Additionally, it is conceivable that some of the somatic changes identified represent genetic mosaicism affecting precursor cells from which the tumor originated. Although such changes would likely not affect actionable genes, careful genomic analyses of normal cells adjacent to neoplastic cells could be performed to resolve this issue.

From a clinical perspective, the use of matched tumor and normal DNA for genomic analyses as we have described is the most direct approach for accurate identification of actionable somatic and germline changes in cancer specimens. Although hotspot mutations in a few oncogenes can be readily detected with high sensitivity and specificity by analyses of tumor tissue alone, we expect that as many as a third of actionable changes in tumor-only analyses may be incorrectly classified as somatic changes when these actually represent constitutional alterations. Use of additional bioinformatic filtering approaches can improve the specificity but will miss a sizable fraction of somatic changes in actionable genes. Additionally, as we have shown, without analysis of germline DNA, cancer patients cannot be accurately screened for hereditary mutations in cancer predisposition genes that could inform the clinical management of the patient and indicate additional family members that could benefit from regular cancer screening. Larger-scale studies in patients with a family history of the tumor types identified in this study could be used to determine the contribution

of mutations such as those indicated in table S8 to different cancers. Conversely, the identification of alterations in cancer-predisposing genes such as *BRCA1* and *BRCA2* in tumor-only analyses may lead to unnecessary referrals for genetic counseling and additional germline-specific testing in cases where these alterations are truly somatic. For example, we identified somatic alterations in *BRCA1* or *BRCA2* in 53 patients without any evidence of additional germline changes in these genes that may have led to unnecessary additional clinical follow-up had these been identified through a tumor-only analysis.

The current use of tumor-only sequencing analyses in many diagnostic laboratories may be a result of previous implementation of mutation hotspot assays but also reflects practical challenges in performing matched tumor and normal analyses, including obtaining normal DNA and the potential need to consent patients for such studies. However, germline DNA can now be routinely obtained from saliva samples and unaffected resected tissue in addition to blood, potentially simplifying logistical challenges. Additionally, patient consents may not be needed when constitutional changes are analyzed only for the purposes of filtering somatic alterations and germline changes are not directly reported. Some institutions are moving to blanket consents that permit comprehensive genomic analyses, including those identifying somatic as well as predisposing alterations in cancer and other diseases.

Given the anticipated widespread adoption of genomic analyses for cancer patients, these studies suggest that such genetic tests need to be carefully designed and implemented in the clinical setting. These results highlight that the sensitivity and specificity of clinical genetic tests can be compromised when analytical methods are used in an attempt to identify somatic mutations in the place of sequencing a matched normal sample. Our studies suggest that the use of tumor-only analyses may lead to inappropriate administration of cancer therapies with substantial effects on patient safety and health care costs. The consequences of such analyses will become even more important through discovery of additional actionable genes and as new targeted therapies continue to be developed. The design of diagnostic assays must be carefully considered to ensure that patients receive the full benefit of these advances.

MATERIALS AND METHODS

Study design

The study was a retrospective analysis of targeted and whole-exome sequencing data from cancer patients with a range of different tumor types. We evaluated the clinical actionability of the mutations identified, determined the fraction of cases with a hereditary mutation in a known cancer predisposition gene and the effectiveness of different bioinformatic approaches in distinguishing between germline and somatic variants in the absence of matched normal samples.

Samples

Eight hundred fifteen tumor samples and matched normal tissues were obtained and analyzed with Western Institutional Review Board approval. A large range of cancer types including brain, breast, colorectal, cholangiocarcinoma, head and neck, neuroendocrine,

renal, gastric, gynecological, esophageal, lung, melanoma, and pancreatic cancers, hematopoietic malignancies, and sarcomas were studied. Sample types included FFPE and frozen tissue, cell lines, DNA, and early-passage patient-derived xenografts. Patient-derived xenografts were included because we have previously shown a high concordance between such samples and matching primary tumors (38). Samples provided as FFPE blocks or frozen tissue underwent pathological review to determine tumor cellularity. Tumors were macrodissected to remove contaminating normal tissue. Matched normal samples were provided as blood, saliva, unaffected tissue, or normal cell lines (table S4).

Sample preparation and next-generation sequencing

Sample preparation, library construction, exome and targeted capture, next-generation sequencing, and bioinformatic analyses of tumor and normal samples were performed as previously described (19). In brief, DNA was extracted from frozen or FFPE tissue, along with matched blood or saliva samples using the Qiagen DNA FFPE Tissue Kit or Qiagen DNA Blood Mini Kit (Qiagen). Genomic DNA from tumor and normal samples was fragmented and used for Illumina TruSeq library construction (Illumina) according to the manufacturer's instructions or as previously described (19). Briefly, 50 ng to 3 µg of genomic DNA in 100 µl of TE (tris-EDTA) was fragmented in a Covaris sonicator to a size of 150 to 450 base pairs(bp). To remove fragments smaller than 150 bp, DNA was purified using Agencourt AMPure XP beads (Beckman Coulter) in a ratio of 1.0:0.9 of polymerase chain reaction (PCR) product to beads twice and washed using 70% ethanol per the manufacturer's instructions. Purified, fragmented DNA was mixed with 36 µl of H₂O, 10 µl of End Repair Reaction Buffer, 5 µl of End Repair Enzyme Mix [cat# E6050, New England BioLabs (NEB)]. The 100-µl end-repair mixture was incubated at 20°C for 30 min and purified using Agencourt AMPure XP beads (Beckman Coulter) in a ratio of 1.0:1.25 of PCR product to beads and washed using 70% ethanol per the manufacturer's instructions. To A-tail, 42 µl of end-repaired DNA was mixed with 5 µl of 10× dA-Tailing Reaction Buffer and 3 µl of Klenow (exo-) (cat# E6053, NEB). The 50-µl mixture was incubated at 37°C for 30 min and purified using Agencourt AMPure XP beads (Beckman Coulter) in a ratio of 1.0:1.0 of PCR product to beads and washed using 70% ethanol per the manufacturer's instructions. For adapter ligation, 25 µl of A-tailed DNA was mixed with 6.7 µl of H₂O, 3.3 µl of paired-end (PE) adapter (Illumina), 10 µl of 5× ligation buffer and 5 µl of Quick T4 DNA ligase (cat# E6056, NEB). The ligation mixture was incubated at 20°C for 15 min and purified using Agencourt AMPure XP beads (Beckman Coulter) in a ratio of 1.0:0.95 and 1.0 of PCR product to beads twice and washed using 70% ethanol per the manufacturer's instructions. To obtain an amplified library, 12 PCRs of 25 µl each were set up, each including 15.5 µl of H₂O, 5 µl of 5× Phusion HF buffer, 0.5 µl of a deoxynucleotide triphosphate (dNTP) mix containing 10 mM of each dNTP, 1.25 µl of dimethyl sulfoxide (DMSO), 0.25 µl of Illumina PE primer #1, 0.25 µl of Illumina PE primer #2, 0.25 µl of Hot Start Phusion polymerase, and 2 µl of the DNA. The PCR program used was 98°C for 2 min; 12 cycles of 98°C for 15 s, 65°C for 30 s, and 72°C for 30 s; and 72°C for 5 min. DNA was purified using Agencourt AMPure XP beads (Beckman Coulter) in a ratio of 1.0:1.0 of PCR product to beads and washed using 70% ethanol per the manufacturer's instructions. Exonic or targeted regions were captured in solution using the Agilent SureSelect version 4 kit or a custom-targeted panel for the 111 genes of interest according to the manufacturer's

instructions (Agilent). The captured library was then purified with a Qiagen MinElute column purification kit and eluted in 17 μ l of 70°C elution buffer to obtain 15 μ l of captured DNA library. The captured DNA library was amplified in the following way: eight 30- μ l PCR reactions each containing 19 μ l of H₂O, 6 μ l of 5× Phusion HF buffer, 0.6 μ l of 10 mM dNTP, 1.5 μ l of DMSO, 0.30 μ l of Illumina PE primer #1, 0.30 μ l of Illumina PE primer #2, 0.30 μ l of Hot Start Phusion polymerase, and 2 μ l of captured exome library were set up. The PCR program used was 98°C for 30 s; 14 cycles (exome) or 16 cycles (targeted) of 98°C for 10 s, 65°C for 30 s, and 72°C for 30 s; and 72°C for 5 min. To purify PCR products, a NucleoSpin Extract II purification kit (Macherey-Nagel) was used following the manufacturer's instructions. PE sequencing, resulting in 100 bases from each end of the fragments for exome libraries and 150 bases from each end of the fragment for targeted libraries, was performed using Illumina HiSeq 2000/2500 and Illumina MiSeq instrumentation (Illumina).

Primary processing of next-generation sequencing data and identification of putative somatic mutations

Somatic mutations were identified using VariantDx custom software for identifying mutations in matched tumor and normal samples. Before mutation calling, primary processing of sequence data for both tumor and normal samples was performed using Illumina CASA VA (Consensus Assessment of Sequence and Variation) software (version 1.8), including masking of adapter sequences. Sequence reads were aligned against the human reference genome (version hg18) using ELAND (Efficient Large-Scale Alignment of Nucleotide Databases) with additional realignment of select regions using the Needleman-Wunsch method (39). Candidate somatic mutations, consisting of point mutations, insertions, and deletions, were then identified using VariantDx across either the whole exome or regions of interest. VariantDx examines sequence alignments of tumor samples against a matched normal while applying filters to exclude alignment and sequencing artifacts. In brief, an alignment filter was applied to exclude quality-failed reads, unpaired reads, and poorly mapped reads in the tumor. A base quality filter was applied to limit inclusion of bases with reported Phred quality scores >30 for the tumor and >20 for the normal (www.phrap.com/phred/). A mutation in the tumor was identified as a candidate somatic mutation only when (i) distinct paired reads contained the mutation in the tumor; (ii) the number of distinct paired reads containing a particular mutation in the tumor was at least 2% of the total distinct read pairs for targeted analyses and 10% of read pairs for exome; (iii) the mismatched base was not present in >1% of the reads in the matched normal sample as well as not present in a custom database of common germline variants derived from dbSNP; and (iv) the position was covered in both the tumor and normal. Mutations arising from misplaced genome alignments, including paralogous sequences, were identified and excluded by searching the reference genome.

Candidate somatic mutations were further filtered on the basis of gene annotation to identify those occurring in protein-coding regions. Functional consequences were predicted using snpEff and a custom database of CCDS, RefSeq, and Ensembl annotations using the latest transcript versions available on hg18 from University of California, Santa Cruz (<https://genome.ucsc.edu/>). Predictions were ordered to prefer transcripts with canonical start and

stop codons and CCDS or RefSeq transcripts over Ensembl when available. Finally, mutations were filtered to exclude intronic and silent changes and retain mutations resulting in missense mutations, nonsense mutations, frameshifts, or splice site alterations. A manual visual inspection step was used to further remove artifactual changes.

Identification of putative somatic mutations without matched normal sample

One hundred cases with exome sequencing data and 58 targeted cases were selected for analyses both with and without their matched normal sample. For the identification of putative somatic mutations without a matched normal, additional filters were applied. First, mutations present in an unmatched normal sample, sequenced to a similar coverage and on the same platform as the matched normal, were removed. Second, alterations reported in the 1000 Genomes Project, present in >1% of the population, or listed as Common in dbSNP138 were filtered. In an attempt to positively select for somatic changes in the resulting data set, mutations occurring within the same amino acid or within five codons of previously reported somatic alterations were identified by comparison to the COSMIC database (version 68, <http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/>). In addition, frameshift, nonsense, and splice site changes predicted to truncate the protein as well as nonsynonymous mutations within the catalytic domain of protein kinases (40) were selected.

Comparison between dbSNP and COSMIC

Common germline mutations were obtained from the dbSNP human variation sets in VCF (variant call format) version 138 labeled “common” with a germline minor allele frequency of 0.01 and indicated with “no known medical impact.” Mutations were filtered on dbSNP fields to exclude synonymous changes and those annotated with somatic origin. Mutations were compared to COSMIC version 68 to identify mutations that matched dbSNP both on genomic position and genomic change.

Clinical actionability analyses

We identified 196 well-characterized genes with potential clinical relevance and assessed the level of evidence for clinical actionability in three ways. First, we determined which of the genes were associated with FDA-approved therapies (www.fda.gov/Drugs/). Second, we carried out a literature search to identify published prospective clinical studies pertaining to genomic alterations of each gene and their association with outcome for cancer patients. Genes that served as targets for specific agents or were predictors of response or resistance to cancer therapies when mutated were considered actionable. Third, we identified clinical trials (<http://clinicaltrials.gov/>) that specified altered genes within the inclusion criteria and were actively recruiting patients in August 2014. In all cases, the tumor type relevant to the FDA approval or studied in the clinical trials was determined to allow the clinical information to be matched to the mutational data by both gene and cancer type.

Identification of germline mutations in cancer predisposition genes

We evaluated the coding regions of 85 previously well-characterized cancer predisposition genes for alterations in normal DNA from blood, saliva, or normal tissue of 815 cases using the VariantDx pipeline adapted to run on germline samples. Mutations resulting in

frameshifts, nonsense, or splice site alterations were considered most likely to be causative and selected for further analysis. Each mutation was compared to published alterations using the ClinVar database (41) and locus-specific databases including the Breast Cancer Information Core (BIC) database for *BRCA1* and *BRCA2* (42), the Leiden Open Variation Databases (LOVD) for *ATM*, *BRIP1*, *FA* genes, and *PALB2* (43), the International Agency for Research on Cancer (IARC) *TP53* database (44), and the International Society for Gastrointestinal Hereditary Tumours Incorporated (InSiGHT) database for the mismatch repair genes *MLH1*, *MSH2*, *MSH6*, and *PMS2* (45). Any alteration designated in these databases as benign or likely benign was excluded.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank V. Adleff and L. D. Wood for technical assistance and J. R. Eshelman, R. H. Hruban, and members of our laboratories for helpful discussions.

Funding: This work was supported in part by the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation, the Commonwealth Foundation, American Association for Cancer Research Stand Up To Cancer–Dream Team Translational Cancer Research grant, John G. Ballenger Trust, FasterCures Research Acceleration Award, Swim Across America, and U.S. NIH grant CA121113.

REFERENCES AND NOTES

1. Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, Marlow S, Jane-Valbuena J, Friedrich DC, Kryukov G, Carter SL, McKenna A, Sivachenko A, Rosenberg M, Kiezun A, Voet D, Lawrence M, Lichtenstein LT, Gentry JG, Huang FW, Fostel J, Farlow D, Barbie D, Gandhi L, Lander ES, Gray SW, Joffe S, Janne P, Garber J, MacConaill L, Lindeman N, Rollins B, Kantoff P, Fisher SA, Gabriel S, Getz G, Garraway LA. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med*. 2014; 20:682–688. [PubMed: 24836576]
2. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, Schnall-Levin M, White J, Sanford EM, An P, Sun J, Juhn F, Brennan K, Iwanik K, Maillet A, Buell J, White E, Zhao M, Balasubramanian S, Terzic S, Richards T, Banning V, Garcia L, Mahoney K, Zwirko Z, Donahue A, Beltran H, Mosquera JM, Rubin MA, Dogan S, Hedvat CV, Berger MF, Pusztai L, Lechner M, Boshoff C, Jarosz M, Vietz C, Parker A, Miller VA, Ross JS, Curran J, Cronin MT, Stephens PJ, Lipson D, Yelensky R. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol*. 2013; 31:1023–1031. [PubMed: 24142049]
3. Garraway LA, Jänne PA. Circumventing cancer drug resistance in the era of personalized medicine. *Cancer Discov*. 2012; 2:214–226. [PubMed: 22585993]
4. Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, Pochanard P, Ducar M, Van Hummelen P, Macconail LE, Hahn WC, Meyerson M, Gabriel SB, Garraway LA. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. *Cancer Discov*. 2012; 2:82–93. [PubMed: 22585170]
5. Dias-Santagata D, Akhavanfard S, David SS, Vernovsky K, Kuhlmann G, Boisvert SL, Stubbs H, McDermott U, Settleman J, Kwak EL, Clark JW, Isakoff SJ, Sequist LV, Engelman JA, Lynch TJ, Haber DA, Louis DN, Ellisen LW, Borger DR, Iafrate AJ. Rapid targeted mutational analysis of human tumours: A clinical platform to guide personalized cancer medicine. *EMBO Mol Med*. 2010; 2:146–158. [PubMed: 20432502]
6. Kerick M, Isau M, Timmermann B, Sultmann H, Herwig R, Krobitch S, Schaefer G, Verdorfer I, Bartsch G, Klocker H, Lehrach H, Schweiger MR. Targeted high throughput sequencing in clinical

cancer settings: Formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. *BMC Med Genomics*. 2011; 4:68. [PubMed: 21958464]

7. Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, Cao X, Kalyana-Sundaram S, Sam L, Balbin OA, Quist MJ, Barrette T, Everett J, Siddiqui J, Kunju LP, Navone N, Araujo JC, Troncoso P, Logothetis CJ, Innis JW, Smith DC, Lao CD, Kim SY, Roberts JS, Gruber SB, Pienta KJ, Talpaz M, Chinnaiyan AM. Personalized oncology through integrative high-throughput sequencing: A pilot study. *Sci Transl Med*. 2011; 3:111ra121.
8. Stegmeier F, Warmuth M, Sellers WR, Dorsch M. Targeted cancer therapies in the twenty-first century: Lessons from imatinib. *Clin Pharmacol Ther*. 2010; 87:543–552. [PubMed: 20237469]
9. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD, Chang DD. Wild-type *KRAS* is required for panitumumab efficacy in patients with metastatic colorectal cancer. *J Clin Oncol*. 2008; 26:1626–1634. [PubMed: 18316791]
10. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, Olivi A, McLendon R, Rasheed BA, Keir S, Nikolskaya T, Nikolsky Y, Busam DA, Tekleab H, Diaz LA Jr, Hartigan J, Smith DR, Strausberg RL, Marie SK, Shinjo SM, Yan H, Riggins GJ, Bigner DD, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008; 321:1807–1812. [PubMed: 18772396]
11. Fumagalli D, Gavin PG, Taniyama Y, Kim SI, Choi HJ, Paik S, Pogue-Geile KL. A rapid, sensitive, reproducible and cost-effective method for mutation profiling of colon cancer and metastatic lymph nodes. *BMC Cancer*. 2010; 10:101. [PubMed: 20233444]
12. Lin MT, Mosier SL, Thiess M, Beierl KF, Debeljak M, Tseng LH, Chen G, Yegnasubramanian S, Ho H, Cope L, Wheelan SJ, Gocke CD, Eshleman JR. Clinical validation of *KRAS*, *BRAF*, and *EGFR* mutation detection using next-generation sequencing. *Am J Clin Pathol*. 2014; 141:856–866. [PubMed: 24838331]
13. Wong NA, Gonzalez D, Salto-Tellez M, Butler R, Diaz-Cano SJ, Ilyas M, Newman W, Shaw E, Taniere P, Walsh SV. Association of Clinical Pathologists Molecular Pathology and Diagnostics Group, RAS testing of colorectal carcinoma—A guidance document from the Association of Clinical Pathologists Molecular Pathology and Diagnostics Group. *J Clin Pathol*. 2014; 67:751–757. [PubMed: 24996433]
14. Kraus C, Rau TT, Lux P, Erlenbach-Wünsch K, Löhr S, Krumbiegel M, Thiel CT, Stöhr R, Agaimy A, Croner RS, Stürzl M, Hohenberger W, Hartmann A, Reis A. Comprehensive screening for mutations associated with colorectal cancer in unselected cases reveals penetrant and nonpenetrant mutations. *Int J Cancer*. 2015; 136:E559–E568. [PubMed: 25142776]
15. De Leeneer K, Coene I, Crombez B, Simkens J, Van den Broecke R, Bols A, Stragier B, Vanhoutte I, De Paepe A, Poppe B, Claes K. Prevalence of *BRCA1/2* mutations in sporadic breast/ovarian cancer patients and identification of a novel de novo *BRCA1* mutation in a patient diagnosed with late onset breast and ovarian cancer: Implications for genetic testing. *Breast Cancer Res Treat*. 2012; 132:87–95. [PubMed: 21553119]
16. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. The consensus coding sequences of human breast and colorectal cancers. *Science*. 2006; 314:268–274. [PubMed: 16959974]
17. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*. 2008; 321:1801–1806. [PubMed: 18772397]
18. Parsons DW, Li M, Zhang X, Jones S, Leary RJ, Lin JC, Boca SM, Carter H, Samayoa J, Bettegowda C, Gallia GL, Jallo GI, Binder ZA, Nikolsky Y, Hartigan J, Smith DR, Gerhard DS, Fults DW, VandenBerg S, Berger MS, Marie SK, Shinjo SM, Clara C, Phillips PC, Minturn JE, Biegel JA, Judkins AR, Resnick AC, Storm PB, Curran T, He Y, Rasheed BA, Friedman HS, Keir

- ST, McLendon R, Northcott PA, Taylor MD, Burger PC, Riggins GJ, Karchin R, Parmigiani G, Bigner DD, Yan H, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE. The genetic landscape of the childhood cancer medulloblastoma. *Science*. 2011; 331:435–439. [PubMed: 21163964]
19. Sausen M, Leary RJ, Jones S, Wu J, Reynolds CP, Liu X, Blackford A, Parmigiani G, Diaz LA Jr, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE, Hogarty MD. Integrated genomic analyses identify *ARID1A* and *ARID1B* alterations in the childhood cancer neuroblastoma. *Nat Genet*. 2013; 45:12–17. [PubMed: 23202128]
 20. Jones S, Wang TL, Shih le-M, Mao TL, Nakayama K, Roden R, Glas R, Slamon D, Diaz LA Jr, Vogelstein B, Kinzler KW, Velculescu VE, Papadopoulos N. Frequent mutations of chromatin remodeling gene *ARID1A* in ovarian clear cell carcinoma. *Science*. 2010; 330:228–231. [PubMed: 20826764]
 21. Wei X, Walia V, Lin JC, Teer JK, Prickett TD, Gartner J, Davis S, Stemke-Hale K, Davies MA, Gershenwald JE, Robinson W, Robinson S, Rosenberg SA, Samuels Y. NISC Comparative Sequencing Program. Exome sequencing identifies *GRIN2A* as frequently mutated in melanoma. *Nat Genet*. 2011; 43:442–446. [PubMed: 21499247]
 22. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011; 474:609–615. [PubMed: 21720365]
 23. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. 2013; 499:43–49. [PubMed: 23792563]
 24. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013; 368:2059–2074. [PubMed: 23634996]
 25. Gao YB, Chen ZL, Li JG, Hu XD, Shi XJ, Sun ZM, Zhang F, Zhao ZR, Li ZT, Liu ZY, Zhao YD, Sun J, Zhou CC, Yao R, Wang SY, Wang P, Sun N, Zhang BH, Dong JS, Yu Y, Luo M, Feng XL, Shi SS, Zhou F, Tan FW, Qiu B, Li N, Shao K, Zhang LJ, Zhang LJ, Xue Q, Gao SG, He J. Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet*. 2014; 46:1097–1102. [PubMed: 25151357]
 26. Shern JF, Chen L, Chmielecki J, Wei JS, Patidar R, Rosenberg M, Ambrogio L, Auclair D, Wang J, Song YK, Tolman C, Hurd L, Liao H, Zhang S, Bogen D, Brohl AS, Sindiri S, Catchpoole D, Badgett T, Getz G, Mora J, Anderson JR, Skapek SX, Barr FG, Meyerson M, Hawkins DS, Khan J. Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. *Cancer Discov*. 2014; 4:216–231. [PubMed: 24436047]
 27. Jiao Y, Pawlik TM, Anders RA, Selaru FM, Streppel MM, Lucas DJ, Niknafs N, Guthrie VB, Maitra A, Argani P, Offerhaus GJ, Roa JC, Roberts LR, Gores GJ, Popescu I, Alexandrescu ST, Dima S, Fassan M, Simbolo M, Mafficini A, Capelli P, Lawlor RT, Ruzzenente A, Guglielmi A, Tortora G, de Braud F, Scarpa A, Jarnagin W, Klimstra D, Karchin R, Velculescu VE, Hruban RH, Vogelstein B, Kinzler KW, Papadopoulos N, Wood LD. Exome sequencing identifies frequent inactivating mutations in *BAP1*, *ARID1A* and *PBRM1* in intrahepatic cholangiocarcinomas. *Nat Genet*. 2013; 45:1470–1473. [PubMed: 24185509]
 28. Zang ZJ, Cutcutache I, Poon SL, Zhang SL, McPherson JR, Tao J, Rajasegaran V, Heng HL, Deng N, Gan A, Lim KH, Ong CK, Huang D, Chin SY, Tan IB, Ng CC, Yu W, Wu Y, Lee M, Wu J, Poh D, Wan WK, Rha SY, So J, Salto-Tellez M, Yeoh KG, Wong WK, Zhu YJ, Futreal PA, Pang B, Ruan Y, Hillmer AM, Bertrand D, Nagarajan N, Rozen S, Teh BT, Tan P. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat Genet*. 2012; 44:570–574. [PubMed: 22484628]
 29. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014; 511:543–550. [PubMed: 25079552]
 30. Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A, Schlick RD, Tang LH, Wolfgang CL, Choti MA, Velculescu VE, Diaz LA Jr, Vogelstein B, Kinzler KW, Hruban RH, Papadopoulos N. *DAXX/ATRX*, *MEN1*, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science*. 2011; 331:1199–1203. [PubMed: 21252315]
 31. Li A, Swift M. Mutations at the ataxia-telangiectasia locus and clinical phenotypes of A-T patients. *Am J Med Genet*. 2000; 92:170–177. [PubMed: 10817650]

32. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467:1061–1073. [PubMed: 20981092]
33. UK launches 1000,000-genome initiative. *Cancer Discov*. 2013; 3:OF6.
34. Human Longevity Inc. launches massive sequencing effort. *Cancer Discov*. 2014; 4:OF5.
35. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, Vogelstein B, Karchin R. Cancer-specific high-throughput annotation of somatic mutations: Computational prediction of driver missense mutations. *Cancer Res*. 2009; 69:6660–6667. [PubMed: 19654296]
36. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*. 2001; 11:863–874. [PubMed: 11337480]
37. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7:248–249. [PubMed: 20354512]
38. Jones S, Chen WD, Parmigiani G, Diehl F, Beerenwinkel N, Antal T, Traulsen A, Nowak MA, Siegel C, Velculescu VE, Kinzler KW, Vogelstein B, Willis J, Markowitz SD. Comparative lesion sequencing provides insights into tumor evolution. *Proc Natl Acad Sci USA*. 2008; 105:4283–4288. [PubMed: 18337506]
39. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 1970; 48:443–453. [PubMed: 5420325]
40. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH. CDD: A Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res*. 2011; 39:D225–D229. [PubMed: 21109532]
41. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014; 42:D980–D985. [PubMed: 24234437]
42. Couch FJ, Weber BL. Mutations and polymorphisms in the familial early-onset breast cancer (*BRCA1*) gene. *Hum Mutat*. 1996; 8:8–18. [PubMed: 8807330]
43. Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT. LOVD v.2.0: The next generation in gene variant databases. *Hum Mutat*. 2011; 32:557–563. [PubMed: 21520333]
44. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, Olivier M. Impact of mutant p53 functional properties on *TP53* mutation patterns and tumor phenotype: Lessons from recent developments in the IARC TP53 database. *Hum Mutat*. 2007; 28:622–629. [PubMed: 17311302]
45. Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, Bapat B, Bernstein I, Capellá G, den Dunnen JT, du Sart D, Fabre A, Farrell MP, Farrington SM, Frayling IM, Frebourg T, Goldgar DE, Heinen CD, Holinski-Feder E, Kohonen-Corish M, Robinson KL, Leung SY, Martins A, Moller P, Morak M, Nystrom M, Peltomaki P, Pineda M, Qi M, Ramesar R, Rasmussen LJ, Royer-Pokora B, Scott RJ, Sijmons R, Tavtigian SV, Tops CM, Weber T, Wijnen J, Woods MO, Macrae F, Genuardi M. InSiGHT, Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet*. 2014; 46:107–115. [PubMed: 24362816]

Fig. 1. Schematic description of whole-exome or targeted next-generation sequencing analyses

The approaches used tumor-only (blue arrow) or matched tumor and normal DNA (red arrow) to identify sequence alterations. Bioinformatic methods to separate germline and somatic changes included comparison to dbSNP, COSMIC, and kinase domain databases. Identified gene alterations were compared to databases of established and experimental therapies to identify potential clinical actionability and predisposing alterations.

Fig. 2. Clinically actionable somatic genomic alterations in various tumor types

Each bar represents the fraction of cases with mutations in clinically actionable genes as determined by the comparison of alterations to genes that were associated with established FDA-approved therapies (brown), previously published clinical trials (green), or current clinical trials in the same tumor type (blue). For approved therapies and previously published clinical trials, potential actionability was also considered in tumor types that were different from those where the clinical use has been described (light brown and light green, respectively). Some of the colorectal tumors analyzed were from patients with tumors known to be *KRAS* wild type, resulting in a lower fraction of cases with actionable changes related to FDA-approved therapies.

Fig. 3. Detection of tumor-specific and germline alterations using tumor-only and matched tumor and normal analyses

(A and B) Bar graphs show the number of true somatic alterations (blue) and germline false-positive changes (red) in each case for tumor-only targeted (A) and exome (B) analyses. The fraction of changes in actionable genes is indicated for both somatic (dark blue) and germline changes (dark red). For exome analyses, actionable alterations for somatic and germline changes are also indicated in the inset graph. (C) Summary of overall characteristics and the number of somatic and germline variants detected for each type of analysis. Total sequence coverage, the number of samples analyzed, and the number of

somatic mutations per tumor in the matched tumor/normal analyses are included for reference.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 4. Bioinformatic filtering approaches for detection of somatic and germline changes (A and B) Somatic candidate mutations identified through targeted (A) and whole-exome (B) analyses. A total of 669 and 140,107 candidate mutations were found before any filtering in targeted and exome analyses, respectively. After filtering using dbSNP, 304 germline variants could be distinguished from 365 candidate somatic mutations in the targeted analyses; 101,924 germline changes were similarly filtered from 38,183 candidate somatic mutations in the exome analyses. Comparison to matched normal samples in each case allowed for distinction between true somatic mutations and germline variants. Filtered variants were compared to COSMIC data to determine the number of somatic mutations that could be distinguished from germline changes using this approach. In parallel, candidate somatic mutations were compared to genes described in FDA approval trials, published clinical trials, and active clinical trials to identify alterations present in clinically actionable genes. The overlaps between the COSMIC data and the categories indicated above are indicated with the designated areas in both targeted and exome analyses.