Priority Report

Cancer Research

Discrepancies in Cancer Genomic Sequencing Highlight Opportunities for Driver Mutation Discovery

Andrew M. Hudson[1], Tim Yates[2], Yaoyong Li[3], Eleanor W. Trotter[1], Shameem Fawdar[1], Phil Chapman[3], Paul Lorigan[4], Andrew Biankin[5], Crispin J. Miller[2,3], and John Brognard[1]

Abstract

Cancer genome sequencing is being used at an increasing rate to identify actionable driver mutations that can inform therapeutic intervention strategies. A comparison of two of the most prominent cancer genome sequencing databases from different institutes (Cancer Cell Line Encyclopedia and Catalogue of Somatic Mutations in Cancer) revealed marked discrepancies in the detection of missense mutations in identical cell lines (57.38% conformity). The main reason for this discrepancy is inadequate sequencing of GC-rich areas of the exome. We have therefore mapped over 400 regions of consistent inadequate sequencing (cold-spots) in known cancer-causing genes and kinases, in 368 of which neither institute finds mutations. We demonstrate, using a newly identified PAK4 mutation as proof of principle, that specific targeting and sequencing of these GC-rich cold-spot regions can lead to the identification of novel driver mutations in known tumor suppressors and oncogenes. We highlight that cross-referencing between genomic databases is required to comprehensively assess genomic alterations in commonly used cell lines and that there are still significant opportunities to identify novel drivers of tumorigenesis in poorly sequenced areas of the exome. Finally, we assess other reasons for the observed discrepancy, such as variations in dbSNP filtering and the acquisition/loss of mutations, to give explanations as to why there is a discrepancy in pharmacogenomic studies, given recent concerns with poor reproducibility of data. *Cancer Res; 74(22); 6390–6. ©2014 AACR.*

Introduction

Personalized therapeutic approaches that target genetically activated drivers have significantly improved patient outcome in a number of common and rare cancers. The development of personalized therapeutics relies on affordable, efficient, and accurate cancer genomic sequencing to identify genetic aberrations present in a given tumor, from which actionable mutations can then be obtained (1). To aid novel driver and targeted therapy discovery, the Sanger Institute (Cambridge, United Kingdom) and Broad Institute (Boston, MA) have developed extensive catalogues of mutations found in a large cohort of cell lines. These resources, which are readily accessible to most biomedical researchers via database portals, have greatly facilitated the process of driver gene discovery. Through an initial evaluation of genetic dependencies in non–small cell lung cancer cell lines, we observed inconsistencies in the mutational profiles as reported by the Sanger Institute's Catalogue of Somatic Mutations in Cancer (COSMIC) database and the Broad Institute's Cancer Cell Line Encyclopedia (CCLE; refs. 2–4). We therefore investigated the extent and causes of these discrepancies to identify opportunities to improve the discovery of driver mutations in oncogenes and tumor suppressors.

Materials and Methods

18 cell line comparison between COSMIC and CCLE data

Commercially available cell lines previously sequenced by COSMIC were identified from the Greenman and colleagues paper (5). Eighteen of these cell lines were also sequenced by CCLE using the Hybrid Capture method using the SureSelect Target Enrichment System (Agilent Technologies) and sequencing on Illumina instruments (76bp paired read ends). Mutational data were downloaded from CCLE website on May 14, 2013 (*CCLE_hybrid_capture1650_hg19_NoCommonSNPs_NoNeutralVariants_CDS_2012.05.07.maf*). COSMIC data were downloaded for each cell line from their respective webpages on May 14, 2013. Common genes reported as

[1]Signalling Networks in Cancer Group, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, United Kingdom. [2]RNA Biology Group, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, United Kingdom. [3]Computational Biology Support Team, Cancer Research UK Manchester Institute, The University of Manchester, Manchester, United Kingdom. [4]University of Manchester and The Christie NHS Foundation Trust, Manchester, United Kingdom. [5]Wolfson Wohl Translational Cancer Research Centre, University of Glasgow, United Kingdom.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Y. Li and E.W. Trotter contributed equally to this article.

Corresponding Authors: John Brognard, Signalling Networks in Cancer Group, Cancer Research UK Manchester Institute, Manchester M20 4BX, United Kingdom. Phone: 44-1613065301; Fax: 44-1614463109; E-mail: John.Brognard@cruk.manchester.ac.uk; and Crispin J. Miller, RNA Biology Group and Computational Biology Support Team, Cancer Research UK Manchester Institute, Manchester, M20 4BX, United Kingdom. Phone: 44-1614463176; Fax: 44-1614463109; E-mail: Crispin.Miller@cruk.manchester.ac.uk

doi: 10.1158/0008-5472.CAN-14-1020

©2014 American Association for Cancer Research.

Downloaded from http://aacrjournals.org/cancerres/article-pdf/74/22/6390/2711181/6390.pdf by guest on 09 December 2023

sequenced by both institutes were used to compare both datasets. Script A (Supplementary Data) was written in Groovy programming language to compare the genetic location of missense nontruncating mutations recorded by each institute and compare the lists to find conformity. Sequencing bam files for the CCLE hybrid capture sequencing (COSMIC data unavailable) were viewed using the Integrative Genomics Viewer (IGV: Broad Institute; ref. 6) to categorize the mutations only reported in COSMIC. GC content of the missed mutations was calculated with Ensembl Rest API (version 70) reference genome and capturing the sequence 100 bp either side of the mutation.

568 cell line comparison

COSMIC cell line names were compared with the list of cell lines sequenced by CCLE to find 568 mutually sequenced cell lines. CCLE data were downloaded in the filtered MAF file as described above. COSMIC data were downloaded as a complete file from the COSMIC FTP site on November 12, 2013 (*CosmicCellLineProject_v67_241013.tsv.gv*). The comparison of the sequencing of 1,630 mutually sequenced genes by the two datasets was performed using Script B (Supplementary Data). Mutations were matched by genomic location. Given the variability of gene transcripts from which amino acid changes are calculated, the amino acid change reported was derived from the most common resultant amino acid change and where there was no majority change, the CCLE change was reported followed by COSMIC when comparing COSMIC and CRUK MI data only. CCLE data that were unfiltered (data for common polymorphisms, putative neutral variants, and mutations located outside of the CDS not filtered out) and contained all variants with an allelic fraction >10% were obtained from the CCLE website on November 22, 2013 (*CCLE_hybrid_capture1650_hg19_allVariants_2012.05.07.maf.gz*). The COSMIC-only mutations were cross-referenced against the unfiltered CCLE list to identify further mutation matches. Cancer Census genes were identified from the COSMIC Cancer Census webpage (http://cancer.sanger.ac.uk/cancergenome/projects/census/; ref. 7).

Whole-exome sequencing of four cell lines

Cell lines were obtained from ATCC and DNA extracted within three passages of delivery from ATCC corresponding to less than one month from time of receipt. ATCC authenticates cell lines through short tandem repeat profiling, morphology analysis, cytochrome C oxidase I testing, and karyotyping. On arrival from ATCC, the total passage number for each cell line was H2009 = 23, H2087 = 21, H2122 = 21, H1437 = 46. Cells are maintained in RPMI medium-1640 (Invitrogen) with additional 10% FCS (Lonza Group) and 4 mmol/L GlutaMAX (Invitrogen). Cells are split 1:10 at 80% confluency. DNA extraction is performed using DNeasy Blood and Tissue Kit (Qiagen). Whole-exome sequencing was performed using Agilent Sure Select XT Target Enrichment System for Illumina Pair-end Multiplex Sequencing, enriching with the SureSelect XT Human All Exon V4 library and performing 2 × 100 bp paired-end sequencing on the Illumina HiSeq 2500 with TruSeq SBS v3 chemistry (read density: Supplementary Table S5). Average read density for each sample was calculated using the Lander/Waterman equation as detailed in the Illumina Estimating Coverage Technical Note (http://res.illumina.com/documents/products/technotes/technote_coverage_calculation.pdf). Variant calling was made using the Genome Analysis Tool Kit (GATK: Broad Institute; ref. 8). Comparison of conformity with the COSMIC and CCLE mutation calls was made using Script B with data filtered and unfiltered for mutations with dbSNP ids.

Cold-spot analysis

Bam files from hybrid capture used to create the CCLE database are not available for download so ten independent CCLE whole-exome bam files (performed on Illumina HiSeq 2000) were downloaded on January 9, 2014, via the Cancer Genomics Hub (bam files and metadata with experimental info available from https://browser.cghub.ucsc.edu). These files were analyzed for 986 kinase and Cancer Census genes (Supplementary Table S6), among which 969 genes are protein-coding genes as annotated in ENSEMBL human gene database version 70. The lung cancer sequencing files used were: CCLE-NCI-H2286-DNA-08, CCLE-NCI-H1944-DNA-08, CCLE-COR-L95-DNA-08, CCLE-NCI-H1373-DNA-08, CCLE-NCI-H1184-DNA-08, CCLE-HLF-a-DNA-08, CCLE-JL-1-DNA-08, CCLE-HCC-78-DNA-08, CCLE-DV-90-DNA-08, CCLE-DMS153-DNA-08. The reads in the bam files were mapped onto the reference genome hg19. From each bam file, the read coverage at each base of the protein-coding exonic regions of the 969 selected genes was obtained using Samtools Mpileup (9). Sequencing read cold-spots were defined as protein-coding exonic regions spanning 100 nucleotide bps or more and with the averaged read coverage ≤ 4 at each base. Read cold-spots were identified in the sequencing data and the GC content calculated using the bases corresponding to the read cold-spot. Multiple transcripts of the same gene were removed if the genetic location of the identified cold-spot was identical or the start or end genomic location was the same between same gene transcripts (retaining the transcript with the longest read cold-spot). Top 20 cold-spots are defined as gene transcripts (that were sequenced by CCLE and COSMIC) with the largest cold-spot regions. The average GC content for all coding exons was calculated using the longest transcript (Ensemble Version 70) for each of the 969 genes screened for cold-spots. Circos plots were constructed using the Circos software (from http://www.circos.ca; ref. 10).

Verification of PAK4 mutation

Amplification PCR of region of interest was performed using Phusion High Fidelity PCR Master Mix with H.F. Buffer (New England Biolabs; 12.5 μL) with Betaine 5M (Sigma; 5 μL), 250 ng DNA, forward and reverse primers (Eurofin MWG Operon; 1.25 μL each), and water to make reaction volume of 25 μL. PCR was carried out on S1000 Thermal Cycler (Bio-Rad) with the following PCR steps for a total of 40 cycles; (i) 98.0° 30 seconds; (ii) 98.0° 10 seconds; (iii) 62.0° 30 seconds; and (iv) 72.0° for 150 seconds. PCR product purification was carried out with Illustra ExoProstar Enzymatic PCR and Sequencing Clean-up (GE Healthcare). Sequencing was carried out using an ABI13130 16 capillary system (Life Technologies) and sequencing data were analyzed using 4Peaks software (MekenTosj).

Downloaded from http://aacrjournals.org/cancerres/article-pdf/74/22/6390/2711181/6390.pdf by guest on 09 December 2023

PAK4 transient overexpression

Wild-type PAK4 plasmid (Addgene 23713) was obtained from Addgene (deposited by Hahn and Root; ref. 11). The plasmid was cloned into a Flag-tagged destination vector. STOP codon and the E119Q mutation were introduced by site-directed mutagenesis (Quick Change II Kit, Agilent Technologies). Plasmid was transfected into HEK293T cells in a 12-well format using Attractene according to the manufacturer's protocol. Cells were lysed on ice after 48 hours using Triton X-100 Cell Lysis Buffer supplemented with protease inhibitor tablet (Roche). Lysates were resolved on SDS-PAGE gels followed by Western blotting. Primary antibodies used were: Flag M2 and α- tubulin (Sigma); pERK1/2 (T202/Y204) and pJNK (T183/Y185; Cell Signaling Technology). Mouse or rabbit horseradish peroxidase-conjugated antibodies were used as secondary (Cell Signaling Technology). All Western blot analyses are representative of three independent experiments.

Results and Discussion

We compared missense mutations found in 568 cancer cell lines sequenced by CCLE and COSMIC (v67) across 1,630 mutually sequenced genes (3). A total of 45,377 mutations were reported, of which 26,038 were consistent between institutes (57.38%). A total of 4,496 (9.91%) and 14,843 (32.71%) mutations were found solely by CCLE or COSMIC, respectively (Fig. 1). The ISHIKAWAHERAKLIO02ER cell line, sequenced by both institutes using their standard protocols, showed a total of 263 mutations (52 in COSMIC and 213 in CCLE) but no matches, suggesting different cell lines may have been sequenced. Cross-referencing to Cancer Census genes (7) found that 4,058 mutations reported in one, but not both, of the databases were in known cancer-causing genes (Supplementary Table S1). These included mutations in *EGFR*, *TP53*, *BRAF*, *MAP2K1*, and *PIK3CA* (Table 1), highlighting the difficulties faced when using NGS to identify driver mutations even in well-known cancer-causing genes. Our data reveal a marked discrepancy in mutation reporting between the two most prominent resources and that cross-referencing between the databases is imperative.

We had previously performed a pilot comparison of mutational profiles in 18 cancer cell lines sequenced by the Broad Institute's CCLE using Hybrid Capture sequencing (3), and an earlier release of Sanger Institute's COSMIC database (5, 12). Similar to our larger-scale comparison, we observed low consensus between missense mutation detection in mutually sequenced genes (mean 41.33%; Supplementary Fig. S1). Analyzing the raw read data (6) from CCLE suggested that the most common source of discrepancy was poor sequencing read coverage (41%; Fig. 2). We therefore analyzed 10 randomly selected CCLE whole-exome sequencing files to identify regions of poor coverage (cold-spots). We discovered over 400 cold-spots (100 bp or larger) in Cancer Census and kinase genes that we have mapped as a resource for the research community (Fig. 3 and Supplementary Table S2; ref. 10). These cold-spots are rich in GC nucleotides (63.49% compared with 51.74% average GC-content of all exons in target genes) indicating that high GC-content is a major cause of inadequate sequencing coverage. Importantly, we found for CCLE and COSMIC data combined, an 18-fold reduction in mutation

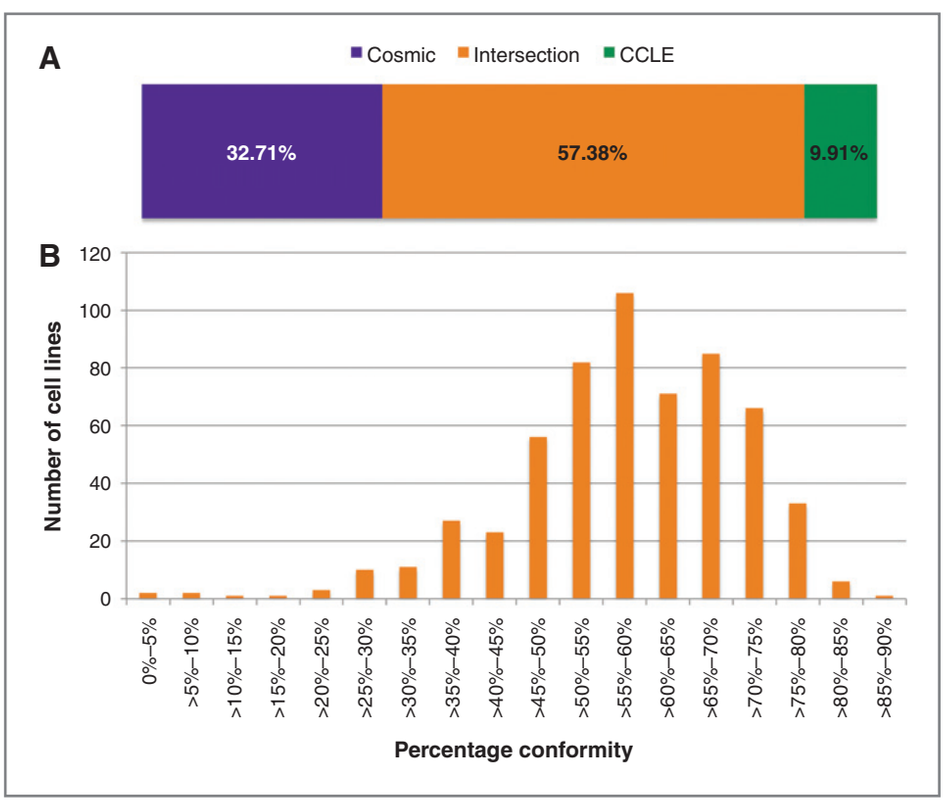

Figure 1. Marked discrepancy is seen in mutation calling between CCLE and COSMIC. A, overall percentage conformity of 46,409 mutations detected by COSMIC and/or CCLE. The intersection between datasets (mutations found by both institutes) accounted for 57.38%. COSMIC-only mutations comprised 32.71% of the dataset and CCLE-only mutations 9.91%. B, the percentage agreement between mutations reported in the 568 cell lines sequenced by both institutes.

Downloaded from http://aacrjournals.org/cancerres/article-pdf/74/22/6390/2711181/6390.pdf by guest on 09 December 2023

Table 1. Mutations in well-known oncogenes and tumor suppressor genes that were detected by only one institute (COSMIC or CCLE)

BRAF	P74A, S76P, V120I, E296K, I326T, **I326V(5)**, R506G, S727G
EGFR	Q71L, R98Q, E282K, S306, V323, K327E, Q408R, L469W, G614S, V654M, G659R, P672R, R677C, T678M, G682V, Q701R, A702D, V738D, A750E, A755D, L815F, L861Q, R973Q, A1076T, T1085N, A1118T, D1127N
FGFR2	R6C, C9S, G89V, E163K, P187S, A315T, Y328S, T341M, A355S, G364E, K401R, L451I, P559H, I643T, C809Y, P814T
HRAS	A11T, G12D, L171P
IDH2	Q95R, H358R, **S408R(2)**
JAK2	V80M, Y96H, T108A, V563I, V617F, L905P, N1129S
KRAS	**G12D(2)**, Q61H, I171M, M188V
MAP2K1	**Q56P(3)**, V85I, A158T, R160K, K185T, V211A
NRAS	**Q61K(6), Q61R(2)**
PIK3CA	K111E, C420R, E542K, E545K, F666L, R770Q
STK11	I46T, Y49D, G56V, K62N, K78N, L105S, D196R, **S216F(2)**, G242W, M392I
TP53	V31I, P47R, D48N, D49H, W53L, A74P, A74S, Y103F, **R110L(2)**, R110P, F113C, F113V, K120N, V122L, C124R, Y126D, M133K, C135R, C176W, E180G, **R181C(2)**, I195T, R213L, **V216L(2)**, V218L, Y220C, N239S, S241F, C242F, M246V, R249S, **R273H(14)**, R283C, R290C, P309S, D324A, R337L, F341C, A347P, G360V, G389W

NOTE: Mutations in bold occurred multiple times (number of occurrences is in parentheses). Supplementary Tables S1a and S1b list the mutations, stratified according to the reporting institute.

density at these loci relative to the remaining exonic regions in the dataset. Extrapolating these data suggests that an additional 1,871 mutations would have been detected in Cancer Census and kinase genes across the 568 cell lines (corresponding to a mean of over three new mutations in Cancer Census or kinase genes per cell line) had the read coverage in the cold-spots been adequate. The *TET2* cold-spot (Fig. 3) is one of the largest of such loci identified, and is not associated with high GC-content. Mutations were reported for this locus in COSMIC, suggesting a sequencing issue specific to the CCLE protocol. This demonstrates that factors, other than inadequate sequencing of GC-content, such as library preparation, reagents, and amplification efficiency can also affect mutation detection at certain loci.

We performed whole-exome sequencing on four of the sequenced lung cancer cell lines (H2009, H1437, H2122, H2087) using an Illumina HiSeq 2500 (achieving over 98% uniquely mapped reads) and a GATK pipeline for mutation detection (8). Our own sequencing identified 27 novel mutations in these four cell lines that were undocumented by COSMIC or CCLE (Supplementary Table S3). Two thirds of these were located in areas of poor read coverage as defined by the CCLE hybrid capture sequencing (less than four reads) but reasonable coverage in our data (mean read depth = 63). The average GC-content 100 bp either side of these newly identified mutations was significantly higher than those where all three institutes were in agreement (60.85% vs. 47.13%, $P = < 10^{-4}$). These findings suggest that the new mutations were previously missed because of being located in GC-rich cold-spots. Although the contribution of factors such as different library preparation and reagents may play a role, our data indicate that NGS efficiency of high GC-rich regions is improving, but earlier datasets are more likely to have missed mutations in GC-rich regions. The majority of The Cancer Genome Atlas and International Cancer Genome Consortium data are of a similar age to CCLE and COSMIC, and therefore subject to similar limitations. Our own more recent sequencing faired better in these regions but still had many GC-rich cold-spots in cancer-associated genes. This is a significant problem, particularly in cancers, including lung cancers, which have a mutational signature predominantly favoring GC-rich trinucleotides (13).

One of the novel mutations identified by our group was in PAK4 (E119Q) in H2009. This mutation lies in a GC-rich (> 76%) area of poor read coverage in CCLE (2 reads; neither reporting the mutation). In contrast, the locus was covered by 39 reads in our data, of which 51% identified the mutation (Supplementary Fig. S2). Given the importance of the PAK kinases in the cancer proliferation and survival pathways (2, 14), we further characterized this mutation. Overexpression of the PAK4 E119Q mutant in 293T cells showed enhanced activation of the ERK pathway compared with the wild-type kinase, suggesting that this is a gain-of-function mutation (Supplementary Fig. S3). These data indicate that additional cancer driver mutations in GC-rich regions will be consistently missed by next-generation cancer genomic sequencing studies, and highlight the potential of developing sequencing platforms to target cold-spot regions for novel cancer gene discovery.

Difference in computational protocols represent another important cause of discrepancy, and includes differences in dbSNP filtering as well as the threshold allelic fraction required to call a mutation. We investigated the effects of dbSNP filtering by comparing the COSMIC-only mutations with unfiltered data from CCLE (the equivalent COSMIC data were

Downloaded from http://aacrjournals.org/cancerres/article-pdf/74/22/6390/2711181/6390.pdf by guest on 09 December 2023

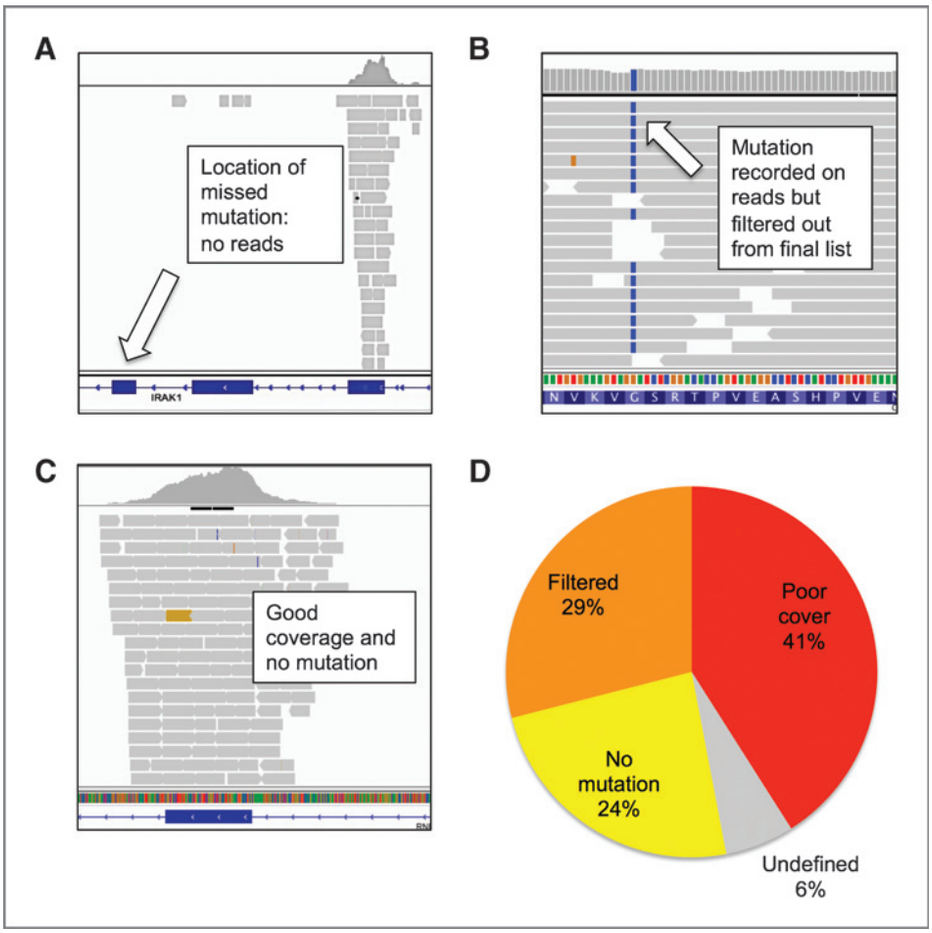

Figure 2. In the original 18 cell line comparison, mutations detected by COSMIC but not CCLE were categorized into: poor coverage with 5 or less reads (A); good read coverage (over 20 reads) and mutation detected on reads but annotated as a dbSNP, neutral variant, outside coding region in all transcripts, or detected on less than 10% of reads, and removed (B); and good coverage, no mutation (C). D, reveals that the most common cause for mutations being missed by CCLE was poor read coverage (41%). Images of read coverage were taken using the Integrative Genomics Viewer.

unavailable). Conformity increased to 67.85% although 10,091 COSMIC-only mutations remained unmatched to CCLE (Supplementary Fig. S4). Therefore, one third of mutations detected only by COSMIC were present on CCLE sequencing reads but were discarded because they were thought to be germline variants. This observation recapitulated the original 18-cell line comparison and our own sequencing also confirmed this with a similar percentage of mutations unreported as a consequence of dbSNP filtering (Supplementary Fig. S5).

By comparing the COSMIC and CCLE data with the four cell lines that we sequenced, we found that 86.34% of the mutations reported by only one database were actually present in our data, suggesting that a minority (approximately 15%–20% based on our two comparisons) of the discrepancy between cell lines is due to acquisition/loss of mutations (Supplementary Table S4). Although a relatively minor factor in our comparisons, the effect of gaining a mutation in a cell line has the potential to greatly affect pharmacogenomic studies. This is highlighted by eight cell lines in the larger comparison that contained activating codon 61 *NRAS* mutations that were reported in only one of the databases (seven reported by COSMIC alone; one by CCLE alone). Analysis of the sequencing data covering the seven *NRAS* mutations not detected by CCLE confirmed good read coverage (mean 220 reads) without evidence of mutation in all seven cases, suggesting loss or gain of the mutation by cell passaging. Passage number is not generally reported in online databases but would greatly assist researchers characterizing the role of specific mutations by indicating whether a mutation has been lost or acquired during passaging.

Although the retrospective nature of our study is unable to control for many sequencing variables such as reagents, polymerases, and platform parameters, we have identified important factors for the discrepancies between the two main cancer genomics databases. These are important findings in the context of a recent study that identified inconsistencies in large pharmacogenomics studies (15). Comparing only 64 genes, this study found some acceptable discrepancies in mutational profiles of cells reported by CCLE and COSMIC but concluded that they were due to differences in the sequencing platforms and variant filtering. Our analysis of a larger panel of genes shows that there is marked discrepancy in sequencing results caused by inadequate sequencing and acquisition of new mutations in addition to variances in dbSNP calling. The authors also concluded that mutational profile was not a major cause of discrepancy in pharmacogenomics data based on the finding that mutational status was not significantly associated with drug response. However, our data show that mutations of

Downloaded from http://aacrjournals.org/cancerres/article-pdf/74/22/6390/2711181/6390.pdf by guest on 09 December 2023

Figure 3. The 20 largest cold-spots detected in cancer census or kinase genes transcripts (of those that were sequenced by both COSMIC and CCLE hybrid capture) using CCLE whole-exome sequencing data. All but one of these cold-spots was located in a high GC-content area and resulted in no mutations being detected by either institute. The *TET2* cold-spot was not located in high-GC content areas and contained mutations detected by COSMIC, indicating that this cold-spot was not present in the COSMIC data. The outer shaded gray plot shows the GC content at each base (calculated as 50 bp either side) with GC content over 70% shaded in red. The middle light green plot shows sequencing read coverage with white troughs representing poor read coverage. The inner three rings record the position of mutations found by both institutes (orange), COSMIC-only (violet), and CCLE (green). Light blue shards show cold-spots over 100 bp in length with the top 20 shaded darker. Data were plotted using a combination of Circos and custom scripts.

Downloaded from http://aacrjournals.org/cancerres/article-pdf/74/22/6390/2711181/6390.pdf by guest on 09 December 2023

cancer-causing genes in sequencing read cold-spots will be frequently undetected, and therefore greatly weaken any analysis attempting to correlate mutation status with drug response. These unsequenced regions of the exome will undoubtedly contain driver mutations, thus mapping cold-spot regions will facilitate novel therapeutic target discovery.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: A.M. Hudson, S. Fawdar, P. Lorigan, A. Biankin, C.J. Miller, J. Brognard

Development of methodology: A.M. Hudson, S. Fawdar, A. Biankin, C.J. Miller, J. Brognard

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): A.M. Hudson, E.W. Trotter, J. Brognard

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.M. Hudson, T. Yates, Y. Li, P. Chapman, C.J. Miller, J. Brognard

Writing, review, and/or revision of the manuscript: A.M. Hudson, Y. Li, P. Lorigan, A. Biankin, C.J. Miller, J. Brognard

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A.M. Hudson, T. Yates, P. Chapman

Study supervision: P. Lorigan, C.J. Miller, J. Brognard

Acknowledgments

The authors thank members of the Signalling Networks in Cancer Group and RNA Biology Groups for helpful discussions, the Core Facility for their advice and support, and Drs. William Newman and Ged Brady for helpful comments and suggestions

Grant Support

This work was fully supported by Cancer Research UK.

Received April 3, 2014; revised August 14, 2014; accepted September 15, 2014; published OnlineFirst September 25, 2014.

References

1. Kim ES, Herbst RS, Wistuba II, Lee JJ, Blumenschein GR, Tsao A, et al. The battle trial: personalizing therapy for lung cancer. Cancer Discov 2011;1:44–53.
2. Fawdar S, Trotter EW, Li Y, Stephenson NL, Hanke F, Marusiak AA, et al. Targeted genetic dependency screen facilitates identification of actionable mutations in FGFR4, MAP3K9, and PAK5 in lung cancer. Proc Natl Acad Sci U S A 2013;110:12426–31.
3. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603–7.
4. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. Nucleic Acids Res 2011;39:945–50.
5. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. Nature 2007;446:153–8.
6. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14:178–92.
7. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, et al. A census of human cancer genes. Nat Rev Cancer 2004;4:177–83.
8. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491–8.
9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9.
10. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res 2009;19:1639–45.
11. Johannessen CM, Boehm JS, Kim SY, Thomas SR, Wardwell L, Johnson LA, et al. COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. Nature 2010;468:968–72.
12. van Haaften G, Dalgliesh GL, Davies H, Chen L, Bignell G, Greenman C, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet 2009;41:521–3.
13. Feng Z, Hu W, Hu Y, Tang MS. Acrolein is a major cigarette-related lung cancer agent: Preferential binding at p53 mutational hotspots and inhibition of DNA repair. Proc Natl Acad Sci U S A 2006;103:15404–9.
14. Radu M, Semenova G, Kosoff R, Chernoff J. PAK signalling during the development and progression of cancer. Nat Rev Cancer 2013;14:13–25.
15. Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJ, et al. Inconsistency in large pharmacogenomic studies. Nature 2013;504:389–93.

Downloaded from http://aacrjournals.org/cancerres/article-pdf/74/22/6390/2711181/6390.pdf by guest on 09 December 2023