ARTICLES

The landscape of somatic copy-number alteration across human cancers

Rameen Beroukhim[1,3,4,5]*, Craig H. Mermel[1,3]*, Dale Porter[8], Guo Wei[1], Soumya Raychaudhuri[1,4], Jerry Donovan[8], Jordi Barretina[1,3], Jesse S. Boehm[1], Jennifer Dobson[1,3], Mitsuyoshi Urashima[9], Kevin T. Mc Henry[8], Reid M. Pinchback[1], Azra H. Ligon[4], Yoon-Jae Cho[6], Leila Haery[1,3], Heidi Greulich[1,3,4,5], Michael Reich[1], Wendy Winckler[1], Michael S. Lawrence[1], Barbara A. Weir[1,3], Kumiko E. Tanaka[1,3], Derek Y. Chiang[1,3,13], Adam J. Bass[1,3,4], Alice Loo[8], Carter Hoffman[1,3], John Prensner[1,3], Ted Liefeld[1], Qing Gao[1], Derek Yecies[3], Sabina Signoretti[3,4], Elizabeth Maher[10], Frederic J. Kaye[11], Hidefumi Sasaki[12], Joel E. Tepper[13], Jonathan A. Fletcher[4], Josep Tabernero[14], José Baselga[14], Ming-Sound Tsao[15], Francesca Demichelis[16], Mark A. Rubin[16], Pasi A. Janne[3,4], Mark J. Daly[1,17], Carmelo Nucera[7], Ross L. Levine[18], Benjamin L. Ebert[1,4,5], Stacey Gabriel[1], Anil K. Rustgi[19], Cristina R. Antonescu[18], Marc Ladanyi[18], Anthony Letai[3], Levi A. Garraway[1,3], Massimo Loda[3,4], David G. Beer[20], Lawrence D. True[21], Aikou Okamoto[22], Scott L. Pomeroy[6], Samuel Singer[18], Todd R. Golub[1,3,23], Eric S. Lander[1,2,5], Gad Getz[1], William R. Sellers[8] & Matthew Meyerson[1,3,5]

A powerful way to discover key genes with causal roles in oncogenesis is to identify genomic regions that undergo frequent alteration in human cancers. Here we present high-resolution analyses of somatic copy-number alterations (SCNAs) from 3,131 cancer specimens, belonging largely to 26 histological types. We identify 158 regions of focal SCNA that are altered at significant frequency across several cancer types, of which 122 cannot be explained by the presence of a known cancer target gene located within these regions. Several gene families are enriched among these regions of focal SCNA, including the *BCL2* family of apoptosis regulators and the NF-κB pathway. We show that cancer cells containing amplifications surrounding the *MCL1* and *BCL2L1* anti-apoptotic genes depend on the expression of these genes for survival. Finally, we demonstrate that a large majority of SCNAs identified in individual cancer types are present in several cancer types.

The development of cancer is driven by the acquisition of somatic genetic alterations, including single base substitutions, translocations, infections, and copy-number alterations[1,2]. Recent advances in genome characterization technologies have enabled increasingly systematic efforts to characterize these alterations in human cancer samples[3]. Identification of these genome alterations can provide important insights into the cellular defects that cause cancer and suggest potential therapeutic strategies[2].

Somatic copy-number alterations (SCNAs, distinguished from germline copy-number variations, CNVs; see Supplementary Note 1a) are extremely common in cancer[4–6]. Genomic analyses of cancer samples, by cytogenetic studies and more recently by array-based profiling, have identified recurrent alterations associated with particular cancer types[4–6]. In some cases, focal SCNAs have led to the identification of cancer-causing genes and suggested specific therapeutic approaches[7–14].

A critical challenge in the genome-wide analysis of SCNAs is distinguishing the alterations that drive cancer growth from the numerous, apparently random alterations that accumulate during tumorigenesis (see Supplementary Note 1b). By studying a sufficiently large collection of cancer samples, it should ultimately be possible to create a comprehensive, high-resolution catalogue of all SCNAs consistently associated with the development of all major types of cancer. Key open questions include: the extent to which significant SCNAs are associated with known cancer-related genes or indicate the presence of new cancer-related genes in particular cancer types; the extent to which large sample collections can be used to pinpoint the precise 'targets' of recurrent amplifications or deletions and thereby to identify cancer-related genes

[1]Cancer Program and Medical and Population Genetics Group, The Broad Institute of M.I.T. and Harvard, 7 Cambridge Center, [2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [3]Departments of Medical Oncology, Pediatric Oncology, and Cancer Biology, and Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, 44 Binney Street, [4]Departments of Medicine and Pathology, Brigham and Women's Hospital, 75 Francis Street, [5]Departments of Medicine, Pathology, Pediatrics, and Systems Biology, Harvard Medical School, 25 Shattuck Street, [6]Department of Neurology, Children's Hospital Boston, 300 Longwood Avenue, [7]Department of Pathology, Beth Israel Deaconess Medical Center, 3 Blackfan Circle, Boston, Massachusetts 02115, USA. [8]Novartis Institutes for BioMedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [9]Division of Molecular Epidemiology, Jikei University School of Medicine, 3-25-8 Nishi-shimbashi, Minato-ku, Tokyo 105-8461, Japan. [10]Department of Internal Medicine, UT Southwestern Medical Center, Dallas, Texas 75390-9186, USA. [11]Genetics Branch, Center for Cancer Research, National Cancer Institute and National Naval Medical Center, Bethesda, Maryland 20889, USA. [12]Department of Surgery II, Nagoya City University Medical School, Nagoya 467-8601, Japan. [13]Department of Genetics and Radiation Oncology, UNC/Lineberger Comprehensive Cancer Center, University of North Carolina, School of Medicine, Chapel Hill, North Carolina 27599, USA. [14]Medical Oncology Program, Vall d'Hebron University Hospital Research Institute, Vall d'Hebron Institute of Oncology, and Autonomous University of Barcelona, 08035 Barcelona, Spain. [15]Department of Pathology and Division of Applied Molecular Oncology, University Health Network, Princess Margaret Hospital and Ontario Cancer Institute, Toronto, Ontario M5G 2M9, Canada. [16]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York 10065, USA. [17]Center for Human Genetic Research, Massachusetts General Hospital, Richard B. Simches Research Center, Boston, Massachusetts 02114, USA. [18]Departments of Medicine and Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [19]Departments of Medicine (GI Division) and Genetics, and Abramson Cancer Center, University of Pennsylvania, 415 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA. [20]Section of Thoracic Surgery, Department of Surgery, University of Michigan, Ann Arbor, Ann Arbor, Michigan 48109, USA. [21]Department of Pathology, University of Washington Medical Center, 1959 North East Pacific Street, Seattle, Washington 98195-6100, USA. [22]Department of Obstetrics and Gynecology, Jikei University School of Medicine, 3-25-8 Nishi-shimbashi, Minato-ku, Tokyo 105-8461, Japan. [23]Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.
*These authors contributed equally to this work.

(see Supplementary Note 2); and the extent to which SCNAs are restricted to particular types or shared across many cancer types, suggesting common biological pathways.

In this paper, we explore these issues by studying copy-number profiles from 3,131 cancers across more than two dozen cancer types, with the data all derived from a single experimental platform and analysed with a common, rigorous statistical methodology.

A collection of 3,131 copy-number profiles across cancer

The 3,131 cancer copy-number profiles consisted of 2,509 profiles determined by our laboratory (see references in Supplementary Note 3), including more than 800 previously unpublished profiles, and 622 profiles determined by other groups[11,15,16]. Most (2,965) come from 26 cancer types, each represented by more than 20 specimens. Seventeen cancer types are represented by at least 40 specimens each (Supplementary Table 1). Most profiles (2,520) were obtained from tissue specimens, with the remainder from cancer cell lines (541) and melanoma short-term cultures (70).

Copy-number measurements were obtained on the Affymetrix 250 K Sty array, containing probes for 238,270 single nucleotide polymorphisms (SNPs). We compared the signal intensities from each cancer specimen to array data from 1,480 normal tissue specimens (of which 1,140 were paired with cancer specimens from the same individual) to identify regions of somatically generated SCNA. We recorded the genomic position, length and amplitude of change in normalized copy-number for every SCNA (Supplementary Fig. 1a and Supplementary Methods).

We observed a total of 75,700 gains and 55,101 losses across the 3,131 cancers, for a mean of 24 gains (median = 12) and 18 losses (median = 12) per sample. For most (17 out of 26) cancer types, the mean number of SCNAs per sample was within twofold of these overall means (Supplementary Fig. 1b). Across all samples, 8.3% of amplification and 8.7% of deletion breakpoints (excluding those occurring within centromeres or telomeres) occurred in regions of segmental duplication, which is enriched relative to the proportion of the genome in such regions (5.1% of SNPs; $P < 10^{-20}$ in each case) and probably reflects a predisposition to SCNA formation[17]. An average of 17% of the genome was amplified and 16% deleted in a typical cancer sample, compared to averages of 0.35% and less than 0.1% in normal samples (representing germline CNVs and occasional analytic artefacts).

Background rates of focal and arm-level SCNAs

Across the entire genome, the most prevalent SCNAs are either very short (focal) or almost exactly the length of a chromosome arm or whole chromosome (arm-level) (Fig. 1a). The focal SCNAs occur at a

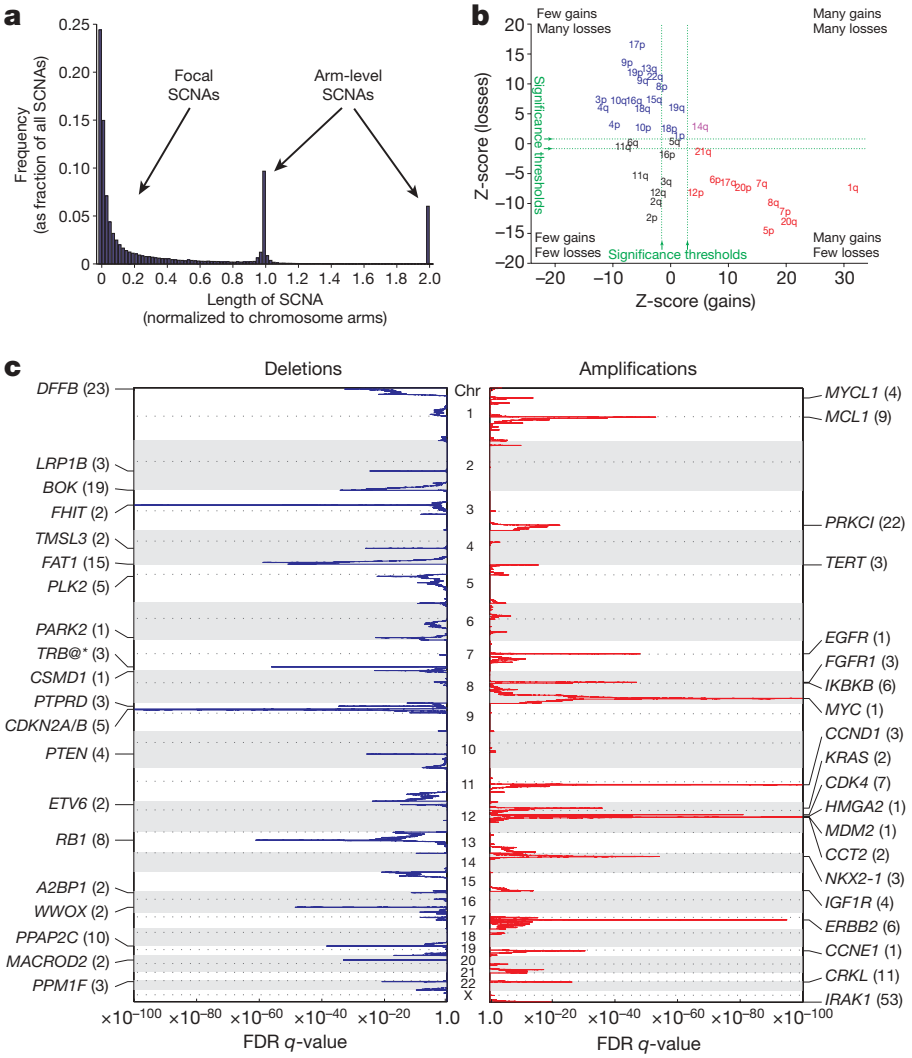

Figure 1 | Identification of significant arm-level and focal SCNAs across cancer. **a**, Length distribution of SCNAs. **b**, The significance of arm-level SCNAs. The length-adjusted Z-scores for gains (*x*-axis) and losses (*y*-axis) of indicated chromosome arms are shown. Arms in red, blue, purple and black show significant gain, loss, both, or neither, respectively. **c**, The significance of focal SCNAs. GISTIC *q*-values (*x*-axis) for deletions (left, blue) and amplifications (right, red) are plotted across the genome (*y*-axis). Known or putative gene targets within the peak regions (*TRB@*, indicated by an asterisk, is immediately adjacent) are indicated for the 20 most significant peaks; values in parentheses represent the number of genes in the peak region.

frequency inversely related to their lengths, with a median length of 1.8 megabases (Mb) (range 0.5 kilobases (kb)–85 Mb).

Arm-level SCNAs occur approximately 30 times more frequently than would be expected by the inverse-length distribution associated with focal SCNAs (Fig. 1a). This observation is seen across all cancer types (Supplementary Fig. 2), and applies to both copy gains and losses (data not shown). As a result, in a typical cancer sample, 25% of the genome is affected by arm-level SCNAs and 10% by focal SCNAs, with 2% overlap. All arm-level (and most focal) SCNAs are of low amplitude (usually single-copy changes), but some focal SCNAs can range to very high amplitude. When analysing SCNAs for evidence of significant alteration in cancer, we accounted for the difference in background rates between arm-level and focal SCNAs by considering them separately.

Several studies have analysed patterns of arm-level SCNAs across large numbers of cancer specimens[4–6], and our results are mainly in agreement with theirs. We also observed that the frequency of arm-level SCNAs decreases with the length of chromosome arms. Adjusted for this trend, most chromosome arms show strong evidence of preferential gain or loss, but rarely both, across many cancer lineages (see Fig. 1b and Supplementary Note 4).

The large size of arm-level SCNAs makes it difficult to determine the specific target gene or genes. By contrast, mapping of focal SCNAs has great power to pinpoint the important genes targeted by these events[7–14].

Pooled analysis of focal SCNAs

We determined the regions in which SCNAs occur at a significantly high frequency. For this purpose, we calculated the genome-average 'background' rates for SCNAs in our data set as a function of length and amplitude, and used the GISTIC (genomic identification of significant targets in cancer) algorithm[18] with improvements as described in Supplementary Methods.

We identified 158 independent regions of significant focal SCNAs, including 76 amplifications and 82 deletions, in the pooled analysis of all our data (Fig. 1c and Supplementary Table 2). This number was relatively robust to changes in the number of samples (Supplementary Fig. 3a) and removal of individual cancer types from the pooled analysis (Supplementary Fig. 3b). Indeed, a stratified analysis of 680 samples distributed evenly across the 17 most highly represented cancer types identified 76% of these significant SCNAs, similar to the number expected based on the reduced power of this smaller sample set (Supplementary Fig. 3a).

The most frequent of these significant focal SCNAs (*MYC* amplifications and *CDKN2A/B* deletions) involve 14% of samples, whereas the least frequent are observed in 2.3% of samples for amplifications and 1.5% for deletions. The frequency of significant arm-level SCNAs is higher (15–29% of samples; Supplementary Fig. 3c). These frequencies are likely to be underestimates, as some SCNAs are not detected owing to contamination of cancer samples with DNA from adjacent normal cells, technical error, and the incomplete spatial resolution afforded by the SNP array platform.

For each of the 158 significant focal SCNAs, we determined a confidence interval ('peak region') that has a 95% likelihood of containing the targeted gene (Supplementary Fig. 3d). Our large data set enables more sensitive and high-resolution detection of peak regions than previous copy-number analyses (see Supplementary Note 5 and Supplementary Table 3). An even larger data set would be desirable, on the basis of analyses showing that the increase in resolution with sample size has not reached a plateau (Supplementary Fig. 3e).

The 76 focal amplification peak regions contain a median of 6.5 genes each (range 0–143, including microRNAs). Sixteen regions contain more than 25 genes each; the remaining 60 regions contain in aggregate 364 potential target genes. We found that 25 of the 76 regions (33%) contain functionally validated oncogenes documented to be activated by amplification (Supplementary Table 2), including nine of the top ten regions (*MYC*, *CCND1*, *ERBB2*, *CDK4*, *NKX2-1*, *MDM2*, *EGFR*, *FGFR1* and *KRAS*; Fig. 1c and Supplementary Table 2). The tenth region, on 1q, contains nine genes; we present evidence later that the target gene in this region is the anti-apoptotic *BCL2* family member, *MCL1*.

The 82 focal deletion peaks contain a median of seven genes each (range 1–173). Nineteen regions contain at least 25 genes each; the remaining 63 regions contain in aggregate 474 potential target genes. Nine of the 82 regions (11%) contain functionally validated tumour suppressor genes documented to be inactivated by deletion (Supplementary Table 2). Two other deletions (involving *ETV6* and the span from *TMPRSS2* to *ERG*) are associated with translocation events that create oncogenes. Another deletion adjacent to the T-cell-receptor-β locus occurs in acute lymphoblastic leukaemia and likely is not associated with cancer, as it occurs during normal T-cell development.

The remaining 70 deletion peaks do not contain known tumour suppressor genes, translocation sites, or somatic rearrangements. More than one-third (26) contain large genes, the genomic loci of which span more than 750 kilobases (kb); none of these genes has been convincingly demonstrated to be a tumour suppressor gene. Conversely, 19 of the 40 largest genes in the genome occur in deletion peaks (Fig. 2a; $P = 3 \times 10^{-9}$). This association between deletions and large genes could be due to a propensity for both to occur in regions of low gene density. Indeed, large genes tend to occur in gene-poor regions (Fig. 2a, bottom), and an analysis of all SCNAs in the data set shows that deletions (but not amplifications) show a bias towards regions of low gene density (up to 30% below the genome average; Fig. 2b). Even after removing the 26 SCNAs containing large genes, the gene density among the remaining deletions is still 25% below the genome average. These observations suggest that some of the deletions may not be related to cancer aetiology, but rather may reflect a high frequency of deletion or low levels of selection against deletion in these regions.

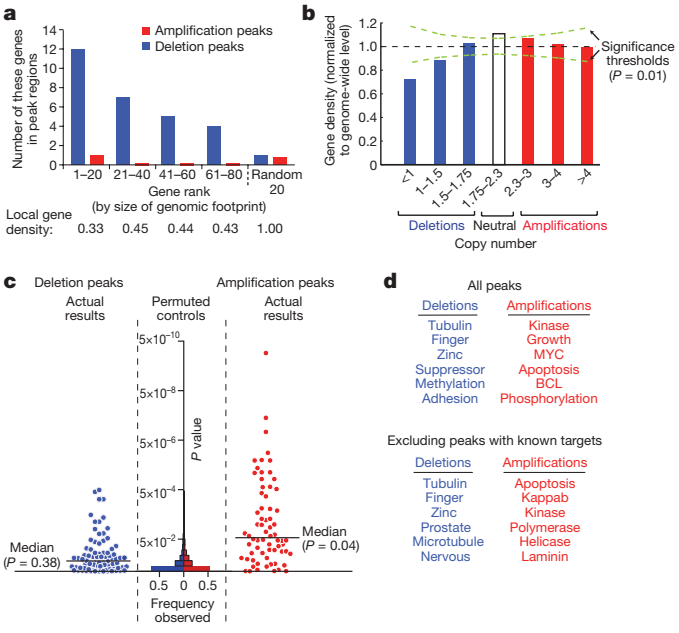

Figure 2 | Characteristics of significant focal SCNAs. **a**, Enrichment of large genes in deletion peaks. Genes are ranked by the amount of genome occupied. Local gene density is normalized against the genome-wide average. **b**, Average gene density among genomic regions as a function of their copy number. **c**, GRAIL analysis[20] *P*-values, plotted for each peak region, reflect the similarity between genes in that region compared to genes in all other regions. Increasing significance is plotted towards the top and reflects greater similarity. Histograms of *P*-values are shown for randomly placed regions ('permuted controls'). **d**, The literature terms most associated with genes in either deletion or amplification peaks, but not both.

Most known amplified oncogenes reside within the 76 amplified regions, although there are exceptions. For example, *MITF*[19] is probably undetected because it is a lineage-specific oncogene restricted to melanoma. At least ten known deleted tumour suppressor genes do not reside in the deleted regions in the pooled analysis (*BRCA2*, *FBXW7*, *NF2*, *PTCH1*, *SMARCB1*, *STK11*, *SUFU*, *VHL*, *WT1* and *WTX* (also known as FAM123B)). Some of these are specific to cancer types not represented in our data set (for example, *NF2*, *WT1* and *WTX*), whereas others (for example, *BRCA2*, *FBXW7*, *STK11* and *VHL*) primarily suffer arm-level deletions (with possible further deletions beyond the resolution of the array platform). Other tumour suppressor genes may be missed if they lie within regions in which the background deletion rates are lower than the genome-wide average, or if they are adjacent to genes in which deletion is poorly tolerated (which would be expected to occur more readily in regions of high gene density; see Supplementary Note 1b). Such tumour suppressors might be inactivated by point mutations more often than SCNAs.

Over-represented gene families and pathways

We assessed potential cancer-causing genes in the SCNAs using GRAIL (gene relationships among implicated loci)[20], an algorithm that searches for functional relationships among genomic regions. GRAIL scores each gene in a collection of genomic regions for its 'relatedness' to genes in other regions based on textual similarity between published abstracts for all papers citing the genes, on the notion that some target genes will function in common pathways.

We found that 47 of the 158 peak regions (34 of the 76 amplification peaks and 13 of the 82 deletion peaks) contain genes significantly related to genes in other peak regions (Fig. 2c). In 21 of these regions, the highest-scoring gene was a previously validated target of SCNA in human cancer (Supplementary Table 2). Across all peak regions, the literature terms most significantly enriched refer to gene families important in cancer pathogenesis, such as kinases, cell cycle regulators, and *MYC* family members (Fig. 2d, top and Supplementary Table 4).

To discover new genes, we next examined the 122 regions without previously documented SCNA targets. The most significantly enriched literature term associated with the amplification peaks was 'apoptosis' (Fig. 2d, bottom and Supplementary Table 4). Two of the five known anti-apoptotic members of the *BCL2* family[21] (*MCL1* and *BCL2L1*) are in amplification peaks. Two of eleven pro-apoptotic members (*BOK* and *BBC3*) were also found among deletion peaks, for a total of four of the 16 known *BCL2* family members, with anti-apoptotic genes amplified but not deleted and vice versa for pro-apoptotic genes (Fig. 3a; $P = 3 \times 10^{-10}$). Although some *BCL2* family members are known to be translocation and point mutation targets[22–26], pathway dysregulation by copy-number change has not been well-described. Later, we describe functional validation that *MCL1* and *BCL2L1* are targets of amplifications that encompass them.

The second-ranking term among amplification peaks without known targets was 'NF-κB', reflecting a preponderance of members

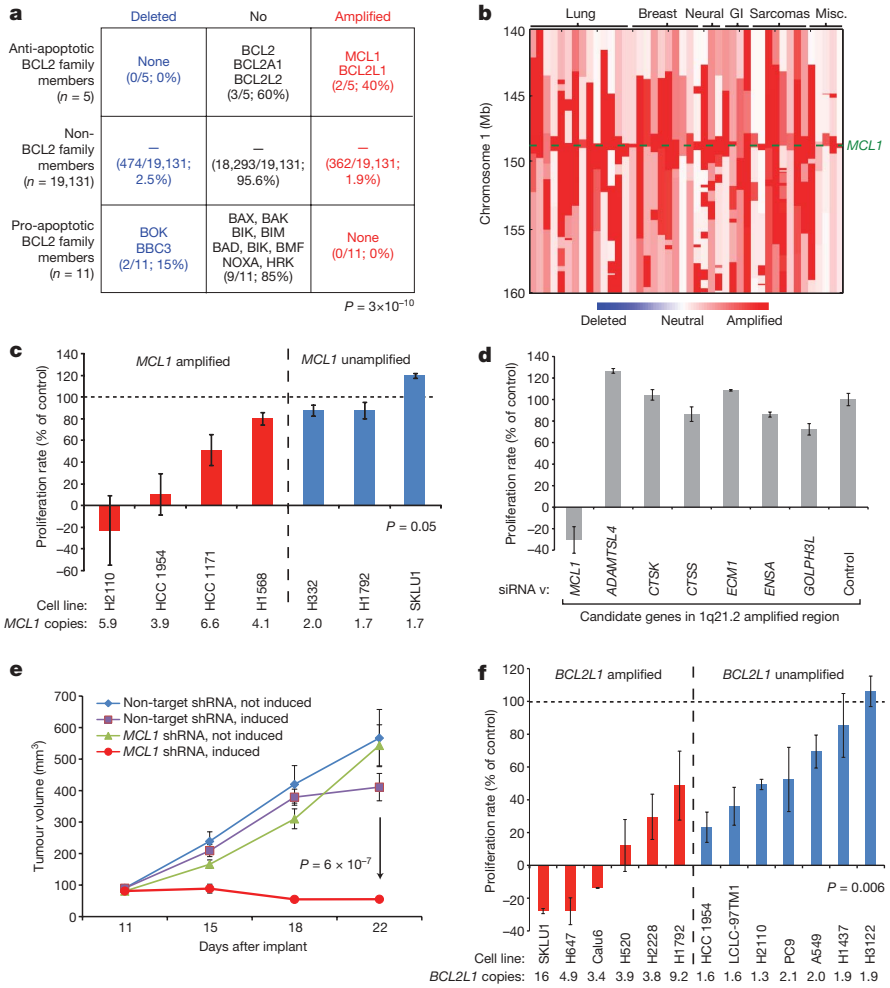

Figure 3 | Dependency of cancer cell lines on the amplified *BCL2* family members, *MCL1* and *BCL2L1*. **a**, Enrichment of pro- and anti-apoptotic *BCL2* family members deletion and amplification peaks. **b**, Copy-number profiles among 50 cancer samples around *MCL1* (lineages are across the top; genomic locations are on the left). **c**, Changes in cell number after *MCL1* knockdown relative to controls. **d**, Proliferation rates in NCI-H2110 cells after siRNA transfection against the listed genes. **e**, Effect of *MCL1* knockdown on growth of NCI-H2110 xenografts. **f**, Changes in cell number after *BCL2L1* knockdown relative to controls. Error bars represent s.e.m. across three experiments.

of this pathway (*TRAF6*, *IKBKB*, *IKBKG*, *IRAK1* and *RIPK1*; $P = 0.001$ for pathway enrichment[27]) and consistent with an emerging recognition of its importance in several cancer types[28–30].

Because some gene families may have been missed by GRAIL, we separately analysed gene ontology (GO) terms for association with amplification peaks (data not shown). We identified significant enrichment of genes associated with 'molecular adaptor activity' (GO: 0060090, $P = 4 \times 10^{-10}$), including *IRS2*, *GRB2*, *GRB7*, *GAB2*, *GRAP*, *TRAF2*, *TRAF6* and *CRKL*. *IRS2* and *GAB2* are known to be transforming when overexpressed[31,32], and *CRKL* has been reported as an essential gene among cells in which it is amplified[33].

Amplifications of *MCL1* and *BCL2L1*

MCL1 is one of nine genes in an amplification peak in cytoband 1q21.2 (Fig. 3b and Supplementary Table 2) with focal amplifications observed in 10.9% of cancers across multiple tissue types. Fluorescence *in situ* hybridization (FISH) of the *MCL1* region in lung and breast cancers showed much higher rates of focal amplification (Supplementary Fig. 4a–b). Amplifications of 1q21.2 were previously reported in two studies of lung adenocarcinoma[7,34] and one of melanoma[35], but the peak regions in these studies contained 86, 36 and 53 genes respectively.

We examined whether cell growth depends on *MCL1* in the presence of gene amplification by measuring the rate of change in cell number after activating an inducible short hairpin RNA (shRNA) against *MCL1* in cells with and without 1q21.2 amplification. We observed a more pronounced reduction in proliferation rates among four *MCL1*-amplified cell lines, compared to three *MCL1*-unamplified control cell lines ($P = 0.05$; Fig. 3c) (all achieved >70% knockdown; Supplementary Fig. 4c). Reducing the expression of six of the other genes (all by >70%; Supplementary Fig. 4d) within the 1q21.2 amplicon in NCI-H2110 cells produced no significant effects (Fig. 3d). Similar effects were observed after *MCL1* depletion with many shRNAs and short interfering RNAs (siRNAs) (Supplementary Fig. 4e). Growth of NCI-H2110 xenografts were also inhibited by induction of anti-*MCL1* shRNA (Fig. 3e).

BCL2L1 is one of five genes in a peak region of amplification on 20q11.21 (Supplementary Fig. 5a). Amplifications of this region have been previously noted in lung cancer[36], giant-cell tumour of bone[37], and embryonic stem cell lines (the latter also amplifying a region including *BCL2*)[38,39], but functional validation of *BCL2L1* as a gene targeted by these amplifications has not been reported. We examined *BCL2L1* dependency using shRNA against *BCL2L1* in cells with and without 20q11.21 amplification. We observed a more pronounced reduction in proliferation rates among six *BCL2L1*-amplified lines (including SKLU1, which was *MCL1*-independent), compared to seven *BCL2L1*-unamplified lines ($P = 0.006$; Fig. 3f). These decreased proliferative rates were associated with increased apoptosis (Supplementary Fig. 5b).

We then sought to explore how amplification of these *BCL2* family members might act in cancer by examining other SCNAs found in cancers carrying *MCL1* or *BCL2L1* amplifications. The most frequent other focal SCNA in these cancers was amplification of the region carrying *MYC* (observed in approximately two-thirds of these cases). *BCL2* has previously been shown to reduce *MYC*-induced apoptosis in lymphoid cells[40]. We found that overexpression of *MCL1* and *BCL2L1* in immortalized bronchial epithelial cells also reduces *MYC*-induced apoptosis (Supplementary Fig. 5c, d). Oncogenic roles for *MCL1*and *BCL2L1* have been previously suggested by reports of increased rates of lymphoma and leukaemia in transgenic mice[41,42]. Somatic amplification of *MCL1* and *BCL2L1* may therefore be a common mechanism for cancers, including carcinomas, to increase cell survival.

Sharing of focal SCNAs across cancer types

Our analysis of a large number of cancer types with a high-resolution platform afforded an opportunity to quantify the degree to which significant focal SCNAs are shared across cancer types. We performed separate analyses of each of the 17 cancer types represented by at least 40 samples and compared the significant SCNAs to those from a pooled analysis of the remaining samples, excluding the cancer type in question.

Most focal SCNAs identified in any one of these 17 cancer types are also found in the pooled analysis excluding that cancer type (median 79% overlap, versus 10% for randomly permuted regions, $P < 0.001$; Fig. 4) and, indeed, in the 158 regions from the overall pool. Nonetheless, cancer-type-restricted analyses identified a further 199 significant SCNAs (145 regions of amplification, 54 regions of deletion; Supplementary Table 5). (These exclude 79 regions of amplification on chromosome 12 found only in dedifferentiated liposarcomas that are probably related to the ring chromosomes in that disease[43]). However, many of these regions were even found to occur in more than one cancer type (median two). As would be expected, the 158 regions in the pooled analysis were found in more cancer types (median five) and were better localized (median size 1.5 Mb versus 11 Mb in the lineage-restricted analyses).

Arm-level alterations, like focal SCNAs, tend to be shared among several cancer types (Supplementary Note 4). Previous studies have demonstrated a tendency for cancers of similar developmental lineages to exhibit similar recurrent arm-level SCNAs[44]. We found that this tendency was much more apparent for arm-level than focal SCNAs (see Supplementary Note 6), suggesting that arm-level SCNAs are shaped to a greater extent by developmental context.

Portal for cancer genomics

The raw data and analyses from this study are available at http://www.broadinstitute.org/tumorscape, including segmented copy-number data (viewable using the Integrative Genomics Viewer; J. Robinson *et al.*, manuscript in preparation) and profiles describing the significance of copy-number changes. The portal also supports gene copy-number queries across and within individual cancer types (instructions are in Supplementary Note 7).

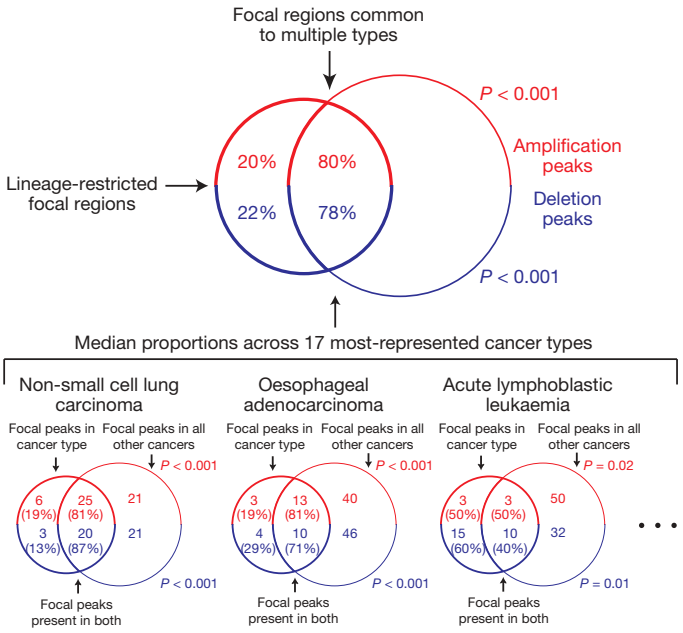

Figure 4 | Most significant focal SCNA peaks identified in any one cancer type are also identified in the rest of the data set (its complement). The top Venn diagram represents median results across the 17 cancer types represented by >40 samples. Venn diagrams representing the specific examples of non-small cell lung cancer, oesophageal adenocarcinoma, and acute lymphoblastic leukaemia are shown along the bottom. The three dots indicate similar analyses were performed on the remaining 14 cancer types. Diagrams are not drawn to scale.

Discussion

This study represents the largest analysis so far of high-resolution copy-number profiles of cancer specimens. Several features of the copy-number landscape apply to the vast majority of cancer types. There is a notably high prevalence of arm-level SCNAs[4–6], which probably reflects the ease with which such mutational events occur compared to focal events[45,46]. The analysis also shows a strong tendency for significant focal SCNAs in one cancer type to be also found in several others.

We identified a total of 357 significant regions of focal SCNA, including 158 regions in the pooled analysis and 199 regions in analyses of individual cancer types. These are surely underestimates of the number of regions that are significantly altered in cancer. Many cancer types were represented by relatively few samples; others were not represented at all. Some SCNAs were missed owing to the resolution limit of the array platform. Further efforts will be needed to characterize larger numbers of cancer genomes at higher resolution to create a comprehensive catalogue of the significant SCNAs and define their occurrence in difference cancer types.

An important challenge is to identify the cancer gene targets of each of these SCNAs. Less than one-quarter of the 158 common peak regions are associated with previously validated targets of SCNAs in human cancer. Although a subset of the SCNAs may represent deletion events that are tolerated but not causally involved in cancer (as suggested by the correlation with gene-poor regions) or frequent owing to mechanistic bias (for example, associated with fragile sites)[47], many more cancer-causing genes are likely to be found through analysis of SCNAs. The GRAIL analysis of our peak regions points to more than a dozen probable candidates, and the functional analysis of *MCL1* and *BCL2L1* strongly implicates these genes as amplification targets. Moreover, some SCNAs may contain several functional targets[10].

Identification of the target genes will require both genomic and functional studies. For focal events, the copy-number profiles of further samples at higher resolution can help narrow the lists of candidates. Nucleotide sequencing may identify point mutations, especially in the case of heterozygous deletions. Because overlapping SCNAs in different cancer types may target different genes, all candidates should be functionally tested separately in each cancer type in appropriate model systems.

Although many canonical oncogenes and tumour suppressor genes are known to be altered across several cancer types and functionally relevant in model systems of diverse tissue origins[1], it has not been clear whether these genes are typical or represent a discovery bias towards genes relevant to many cancer types. By studying a large number of cancers of multiple types, we have found that most of the significant SCNAs within any single cancer type tend to be found in other cancer types as well. Similar findings for point mutations and translocations would suggest that the appearance of tremendous diversity across cancer genomes may reflect the combinations of a limited number of functionally relevant events.

METHODS SUMMARY

DNA extracted from cancer specimens and normal tissue was labelled and hybridized to the Affymetrix 250K Sty I array to obtain signal intensities and genotype calls. Signal intensities were normalized against data from 1,480 normal samples. Copy-number profiles were inferred using GLAD[48] and changes of >0.1 copies in either direction were called SCNAs. The significance of focal SCNAs (covering <0.5 chromosome arms) was determined using GISTIC[18], with modifications to score SCNAs directly proportional to amplitude and to allow summation of non-overlapping deletions affecting the same gene. Peak region boundaries were determined so that the change in the GISTIC score from peak to boundary had $<5\%$ likelihood of occurring by random fluctuation. *P*-values for Figs 2b and 4 were determined by comparing the gene densities of SCNAs and fraction overlap of peak regions, respectively, to the same quantities calculated from random permutations of the locations of these SCNAs and peak regions. RNA interference was performed by inducible and stable expression of shRNA lentiviral vectors and by siRNA transfection. Proliferation in inducible shRNA experiments was measured in triplicate every half-hour on 96-well plates by a real-time electric sensing system (ACEA Bioscience), and in stable shRNA expression and siRNA transfection experiments by CellTiterGlo (Promega). Apoptosis was measured by immunoblot against cleaved PARP, and FACS analysis of cells stained with antibody to annexin V and propidium iodide. Tumour growth in nude mice was measured by caliper twice weekly. Expression of *MYC*, *MCL1* and *BCL2L1* was performed with retroviral vectors in lung epithelial cells immortalized by introduction of SV40 and hTERT[49].

Received 2 June; accepted 23 December 2009.

1. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4,** 177–183 (2004).
2. Stuart, D. & Sellers, W. R. Linking somatic genetic alterations in cancer to therapeutics. *Curr. Opin. Cell Biol.* **21,** 304–310 (2009).
3. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–724 (2009).
4. Baudis, M. Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data. *BMC Cancer* **7,** 226 (2007).
5. *Mitelman Database of Chromosome Aberrations in Cancer* (eds Mitelman, F. Johansson, B. and Mertens, F.) ⟨http://cgap.nci.nih.gov/Chromosomes/Mitelman⟩ (2009).
6. *NCI and NCBI's SKY/M-FISH and CGH Database* ⟨http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi⟩ (2001).
7. Weir, B. A. *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450,** 893–898 (2007).
8. Eder, A. M. *et al.* Atypical PKCiota contributes to poor prognosis through loss of apical-basal polarity and cyclin E overexpression in ovarian cancer. *Proc. Natl Acad. Sci. USA* **102,** 12519–12524 (2005).
9. Lahortiga, I. *et al.* Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia. *Nature Genet.* **39,** 593–595 (2007).
10. Zender, L. *et al.* Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* **125,** 1253–1267 (2006).
11. Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446,** 758–764 (2007).
12. Wiedemeyer, R. *et al.* Feedback circuit among INK4 tumor suppressors constrains human glioblastoma development. *Cancer Cell* **13,** 355–364 (2008).
13. Cancer_Genome_Atlas_Research_Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455,** 1061–1068 (2008).
14. Chitale, D. *et al.* An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. *Oncogene* **28,** 2773–2783 (2009).
15. GlaxoSmithKline. *GSK Cancer Cell Line Genomic Profiling Data* ⟨https://cabig.nci.nih.gov/tools/caArray_GSKdata⟩ (2008).
16. Mullighan, C. G. *et al.* BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature* **453,** 110–114 (2008).
17. Mazzarella, R. & Schlessinger, D. Pathological consequences of sequence duplications in the human genome. *Genome Res.* **8,** 1007 (1998).
18. Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl Acad. Sci. USA* **104,** 20007–20012 (2007).
19. Garraway, L. A. *et al.* Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436,** 117–122 (2005).
20. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5,** e1000534 (2009).
21. Letai, A. G. Diagnosing and exploiting cancer's addiction to blocks in apoptosis. *Nature Rev. Cancer* **8,** 121–132 (2008).
22. Tsujimoto, Y., Gorham, J., Cossman, J., Jaffe, E. & Croce, C. M. The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. *Science* **229,** 1390–1393 (1985).
23. Cleary, M. L. & Sklar, J. Nucleotide sequence of a t(14;18) chromosomal breakpoint in follicular lymphoma and demonstration of a breakpoint-cluster region near a transcriptionally active locus on chromosome 18. *Proc. Natl Acad. Sci. USA* **82,** 7439–7443 (1985).
24. Bakhshi, A. *et al.* Cloning the chromosomal breakpoint of t(14;18) human lymphomas: clustering around JH on chromosome 14 and near a transcriptional unit on 18. *Cell* **41,** 899–906 (1985).
25. Rampino, N. *et al.* Somatic frameshift mutations in the *BAX* gene in colon cancers of the microsatellite mutator phenotype. *Science* **275,** 967–969 (1997).
26. Arena, V., Martini, M., Luongo, M., Capelli, A. & Larocca, L. M. Mutations of the BIK gene in human peripheral B-cell lymphomas. *Genes Chromosom. Cancer* **38,** 91–96 (2003).
27. Wu, C. & Ponnappan, U. *NF-κB Signaling Pathway* ⟨http://www.biocarta.com/pathfiles/h_nfkbPathway.asp#contributors⟩ (2009).
28. Boehm, J. S. *et al.* Integrative genomic approaches identify *IKBKE* as a breast cancer oncogene. *Cell* **129,** 1065–1079 (2007).
29. Barbie, D. A. *et al.* Systematic RNA interference reveals that oncogenic *KRAS*-driven cancers require TBK1. *Nature* **462,** 108–112 (2009).
30. Scholl, C. *et al.* Synthetic lethal interaction between oncogenic *KRAS* dependency and suppression of *STK33* in human cancer cells. *Cell* **137,** 821–834 (2009).

31. Nishida, K. & Hirano, T. The role of Gab family scaffolding adapter proteins in the signal transduction of cytokine and growth factor receptors. *Cancer Sci.* **94,** 1029–1033 (2003).
32. Dearth, R. K., Cui, X., Kim, H. J., Hadsell, D. L. & Lee, A. V. Oncogenic transformation by the signaling adaptor proteins insulin receptor substrate (IRS)-1 and IRS-2. *Cell Cycle* **6,** 705–713 (2007).
33. Luo, B. *et al.* Highly parallel identification of essential genes in cancer cells. *Proc. Natl Acad. Sci. USA* **105,** 20380–20385 (2008).
34. Kendall, J. *et al.* Oncogenic cooperation and coamplification of developmental transcription factor genes in lung cancer. *Proc. Natl Acad. Sci. USA* **104,** 16663–16668 (2007).
35. Lin, W. M. *et al.* Modeling genomic diversity and tumor dependency in malignant melanoma. *Cancer Res.* **68,** 664–673 (2008).
36. Tonon, G. *et al.* High-resolution genomic profiles of human lung cancer. *Proc. Natl Acad. Sci. USA* **102,** 9625–9630 (2005).
37. Smith, L. T. *et al.* 20q11.1 amplification in giant-cell tumor of bone: Array CGH, FISH, and association with outcome. *Genes Chromosom. Cancer* **45,** 957–966 (2006).
38. Spits, C. *et al.* Recurrent chromosomal abnormalities in human embryonic stem cells. *Nature Biotechnol.* **26,** 1361–1363 (2008).
39. Lefort, N. *et al.* Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. *Nature Biotechnol.* **26,** 1364–1366 (2008).
40. Fanidi, A., Harrington, E. A. & Evan, G. I. Cooperative interaction between c-*myc* and *bcl*-2 proto-oncogenes. *Nature* **359,** 554–556 (1992).
41. Zhou, P. *et al. MCL1* transgenic mice exhibit a high incidence of B-cell lymphoma manifested as a spectrum of histologic subtypes. *Blood* **97,** 3902–3909 (2001).
42. Beverly, L. J. & Varmus, H. E. MYC-induced myeloid leukemogenesis is accelerated by all six members of the antiapoptotic BCL family. *Oncogene* **28,** 1274–1279 (2009).
43. Dal Cin, P. *et al.* Cytogenetic and fluorescence in situ hybridization investigation of ring chromosomes characterizing a specific pathologic subgroup of adipose tissue tumors. *Cancer Genet. Cytogenet.* **68,** 85–90 (1993).
44. Myllykangas, S., Bohling, T. & Knuutila, S. Specificity, selection and significance of gene amplifications in cancer. *Semin. Cancer Biol.* **17,** 42–55 (2007).
45. Albertson, D. G. Gene amplification in cancer. *Trends Genet.* **22,** 447–455 (2006).
46. Fukasawa, K. Centrosome amplification, chromosome instability and cancer development. *Cancer Lett.* **230,** 6–19 (2005).
47. Smith, D. I., Zhu, Y., McAvoy, S. & Kuhn, R. Common fragile sites, extremely large genes, neural development and cancer. *Cancer Lett.* **232,** 48–57 (2006).
48. Hupé, P., Stransky, N., Thiery, J.-P., Radvanyi, F. & Barillot, E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* **20,** 3413–3422 (2004).
49. Lundberg, A. S. *et al.* Immortalization and transformation of primary human airway epithelial cells by gene transfer. *Oncogene* **21,** 4577–4586 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the National Institutes of Health (NIH) (Dana-Farber/Harvard Cancer Center and Pacific Northwest Prostate Cancer SPOREs, P50CA90578, R01CA109038, R01CA109467, P01CA085859, P01CA 098101 and K08CA122833), the Doris Duke Charitable Foundation, the Sarah Thomas Monopoli Lung Cancer Research Fund, the Seaman Corporation Fund for Lung Cancer Research, and the Lucas Foundation. Medulloblastoma samples were obtained in collaboration with the Children's Oncology Group. N. Vena provided technical assistance with FISH, and I. Mellinghoff, P. S. Mischel, L. Liau and T. F. Cloughesy provided DNA samples. We thank T. Ried, R. Weinberg and B. Vogelstein for critical review of the manuscript and for comments about its context in the field of cancer genetics.

Author Contributions R.B., C.H.M., E.S.L., G.G., W.R.S. and M.M. conceived and designed the study; R.B., J.B., M.U., A.H.L., Y.-J.C., W.W., B.A.W., D.Y.C., A.J.B., J.P., S.S., E.M., F.J.K., H.S., J.E.T., J.A.F., J.T., J.B., M.-S.T., F.D., M.A.R., P.A.J., C.N., R.L.L., B.L.E., S.G., A.K.R., C.R.A., M.L., L.A.G., M.L., D.G.B., L.D.T., A.O., S.L.P., S.S. and M.M. contributed primary samples and/or assisted in the generation of the data; R.B., C.H.M., S.R., J.Dob., M.S.L., B.A.W., M.J.D. and G.G. performed the data analysis; R.B., D.P., G.W., J.Don., J.S.B., K.T.M., L.H., H.G., K.E.T., A.L., C.H., D.Y., A.L., L.A.G., T.R.G. and M.M. designed and performed the functional experiments on *BCL2* family member genes; R.B., C.H.M., R.M.P., M.R., T.L. and Q.G. designed and built the cancer copy-number portal; R.B., C.H.M., E.S.L. and M.M. wrote, and all other authors have critically read and commented on, the manuscript.

Author Information The SNP array data have been deposited to the Gene Expression Omnibus (GEO) under accession number GSE19399. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.S.L. (lander@broadinstitute.org), G.G. (gadgetz@broadinstitute.org), W.R.S. (william.sellers@novartis.com) or M.M. (matthew_meyerson@dfci.harvard.edu).