The landscape of somatic copy-number alteration across human cancers

A full list of authors and affiliations appears at the end of the article.

Abstract

A powerful way to discover key genes playing causal roles in oncogenesis is to identify genomic regions that undergo frequent alteration in human cancers. Here, we report high-resolution analyses of somatic copy-number alterations (SCNAs) from 3131 cancer specimens, belonging largely to 26 histological types. We identify 158 regions of focal SCNA that are altered at significant frequency across multiple 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 harboring amplifications surrounding the *MCL1* and *BCL2L1* anti-apoptotic genes depend upon expression of these genes for survival. Finally, we demonstrate that a large majority of SCNAs identified in individual cancer types are present in multiple cancer types.

Background

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].

Correspondence to: Eric S. Lander; Gad Getz; William R. Sellers; Matthew Meyerson.

*These authors contributed equally to this work.

Competing Interests Statement

The authors declare that they have no competing financial interests.

Author Contributions

RB, CHM, ESL, GG, WRS, and MM conceived and designed the study; RB, JB, MU, AHL, YC, WW, BAW, DYC, AJB, JP, SS, EM, FJK, HS, JET, JAF, JT, JB, MST, FD, MAR, PAJ, CN, RLL, BLE, SG, AKR, CRA, ML, LAG, ML, DGB, LDT, AO, SLP, SS, and MM contributed primary samples and/or assisted in the generation of the data; RB, CHM, SR, JD, MSL, BAW, MJD, and GG performed the data analysis; RB, DP, GW, JD, JSB, KTM, LH, HG, KET, AL, CH, DY, AL, LAG, TRG, and MM designed and performed the functional experiments on *BCL2* family member genes; RB, CHM, RMP, MR, TL, and QG designed and built the cancer copy number portal; RB, CHM, ESL, and MM wrote, and all other authors have critically read and commented on, the manuscript.

Somatic copy-number alterations (SCNAs, distinguished from germline copy-number variations, CNVs; see Supplementary Note 1a) are extremely common in cancer[4],[5],[6]. Genomic analysis of tumor samples, by cytogenetic studies and more recently array-based profiling, have identified recurrent alterations associated with particular cancer types[4],[5],[6]. In some cases, focal SCNAs have led to the identification of cancer-causing genes and suggested specific therapeutic approaches[7],[8],[9],[10],[11],[12],[13],[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 tumors, it should ultimately be possible to create a comprehensive, high-resolution catalog 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 tumor 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 (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 analyzed with a common, rigorous statistical methodology.

Results

A collection of 3,131 copy-number profiles from multiple cancer types

The 3,131 cancer copy-number profiles consisted of 2,509 profiles determined by our laboratory (see references in Supplementary Note 3), including over eight hundred previously unpublished profiles, and 622 profiles determined by other groups[11],[15],[16]. The majority (2965) 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 a single array platform, the Affymetrix 250K Sty array, containing probes for 238,270 single nucleotide polymorphisms (SNPs). We compared the signal intensities from each cancer specimen to array data from 1480 normal tissue specimens (of which 1140 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 Figure 1a and Supplementary Methods).

We observed a total of 75,700 gains and 55,101 losses across the 3131 cancers, for a mean of 24 gains (median = 12) and 18 losses (median = 12) per sample. For most (17/26) cancer types, the mean number of SCNAs per sample was within two-fold of these overall means

(Supplementary Figure 1b). Across all samples, 8.3% of amplification and 8.7% of deletion breakpoints (excluding those occurring within centromeres or telomeres) occured in regions of segmental duplication, which is enrichment relative to the proportion of genome in such regions (5.1% of SNPs; $p<10^{-20}$ in each case) and likely 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 artifacts).

Focal and chromosome arm-level alterations have different background rates

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) (Figure 1a). The focal SCNAs occur at a frequency inversely related to their lengths, with a median length of 1.8 Mb (range 0.5 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 (Figure 1a). This observation is seen across all cancer types (Supplementary Figure 2) and applies to both copy gains and losses (data not shown). As a result, in a typical tumor, 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 analyzing 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.

Multiple studies have analyzed patterns of arm-level SCNAs across large numbers of cancer specimens[4,5,6], and our results are largely in agreement with theirs. We additionally observed that the frequency of arm-level SCNAs decreases with the length of chromosome arms. Adjusted for this trend, the majority of chromosome arms exhibit strong evidence of preferential gain or loss, but rarely both, across multiple cancer lineages (see Figure 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[7,8,9,10,11,12,13,14].

Pooled analysis of focal SCNAs

We determined those regions in which SCNAs occur at a significantly high frequency. For this purpose, we calculated the genome-average "background" rates for SCNAs in our dataset as a function of length and amplitude, and used the Genomic Identification of Significant Targets In Cancer (GISTIC) algorithm[18] with improvements as described in Supplementary Methods.

We identified 158 independently regions of significant focal SCNAs, including 76 amplifications and 82 deletions, in the pooled analysis of all our data (Figure 1c and Supplementary Table 2). This number was relatively robust to changes in the number of samples (Supplementary Figure 3a) and removal of individual cancer types from the pooled

analysis (Supplementary Figure 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 upon the reduced power of this smaller sample set (Supplementary Figure 3a).

The most frequent of these significant focal SCNAs (*MYC* amplifications and *CDKN2A/B* deletions) involve 14% of samples, while 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 Figure 3c). These frequencies are likely to be underestimates, as some SCNAs are not detected due to contamination of tumor 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 Figure 3d). Our large dataset enables more sensitive and high-resolution detection of peak regions than prior copy-number analyses (see Supplementary Note 5 and Supplementary Table 3). An even larger dataset would be desirable, based on analyses showing that the increase in resolution with sample size has not reached a plateau (Supplementary Figure 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 9 of the top 10 regions *(MYC, CCND1, ERBB2, CDK4, NKX2-1, MDM2, EGFR, FGFR1,* and *KRAS*; Figure 1c, Supplementary Table 2). The tenth region, on 1q, contains nine genes; we present evidence below 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 tumor suppressor genes documented to be inactivated by deletion (Supplementary Table 2). Two additional 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 beta locus occurs in acute lymphoblastic leukemia and likely is not associated with cancer, as it occurs during normal T-cell development.

The remaining 70 deletion peaks do not contain known tumor suppressor genes, translocation sites, or somatic rearrangements. Over one-third (26) contain large genes, whose genomic loci span more than 750 kb; none of these genes has been convincingly demonstrated to be a tumor suppressor gene. Conversely, 19 of the 40 largest genes in the genome occur in deletion peaks (Figure 2a; p = 3×10^{-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 (Figure 2a, bottom), and an analysis of all SCNAs in the dataset reveals that deletions (but not amplifications) show a

bias toward regions of low gene density (up to 30% below the genome average; Figure 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 etiology, but rather may reflect a high frequency of deletion or low levels of selection against deletion in these regions.

The majority of known amplified oncogenes reside within the 76 amplified regions, although there are exceptions. For example, *MITF*[19] is likely undetected because it is a lineage-specific oncogene restricted to melanoma. At least 10 known deleted tumor suppressor genes do not reside in the deleted regions in the pooled analysis (*BRCA2*, *FBXW7*, *NF2*, *PTCH1*, *SMARCB1*, *STK11*, *SUFU*, *VHL*, *WT1*, and *WTX*). Some of these are specific to cancer types not represented in our dataset (e.g. *NF2*, *WT1*, and *WTX*), while others primarily suffer arm-level deletions (with possible additional deletions beyond the resolution of the array platform) (e.g. *BRCA2*, *FBXW7*, *STK11* and *VHL*). Other tumor suppressor genes may be missed if they lie within regions whose background deletion rates are lower than the genome-wide average, or if they are adjacent to genes whose deletion is poorly tolerated (which would be expected to occur more readily in regions of high gene density) (see Supplementary Note 1). Such tumor suppressors might be inactivated by point mutations more often than SCNAs.

Specific gene families and pathways are over-represented among focal SCNAs

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 (Figure 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 (Figure 2d, top; 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" (Figure 2d, bottom; 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 11 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 (Figure 3a; p = 3e-10). Although some *BCL2* family members are known to be translocation and point mutation targets[22],[23],[24],[25],[26], pathway dysregulation by copy-number change has not been

well-described. Below, 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 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 multiple cancer types[28],[29],[30].

Because some gene families may have been missed by GRAIL, we separately analyzed 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=4e-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* are associated with *MCL1* and *BCL2L1* dependence

MCL1 is one of nine genes in an amplification peak in cytoband 1q21.2 (Figure 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 revealed much higher rates of focal amplification (Supplementary Figure 4a–b). Amplifications of 1q21.2 were previously reported in lung adenocarcinoma and melanoma[7],[34],[35], but the peak regions in those studies contained 86, 36 and 53 genes, respectively.

We examined whether cell growth depends upon *MCL1* in the presence of gene amplification by measuring the rate of change in cell number after activating an inducible 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* non-amplified control cell lines (p = 0.05; Figure 3c) (all achieved >70% knockdown; Supplementary Figure 4c). Reducing the expression of 6 of the other genes (all by >70%; Supplementary Figure 4d) within the 1q21.2 amplicon in NCI-H2110 cells produced no significant effects (Figure 3d). Similar effects were observed following *MCL1* depletion with multiple shRNAs and siRNAs (Supplementary Figure 4e). Growth of NCI-H2110 xenografts were also inhibited by induction of anti-*MCL1* shRNA (Figure 3e).

BCL2L1 is one of five genes in a peak region of amplification on 20q11.21 (Supplementary Figure 5a). Amplifications of this region have been previously noted in lung cancer[36], giant-cell tumor 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* non-amplified lines (p = 0.006; Figure 3f). These decreased proliferative rates were associated with increased apoptosis (Supplementary Figure 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 additional focal SCNA in these cancers was amplification of the region carrying *MYC* (62% and 69%, respectively). *BCL2* has previously been shown to reduce *MYC*-induced apoptosis in lymphoid cells[40]. We found that over-expression of *MCL1* and *BCL2L1* in immortalized bronchial epithelial cells also reduces *MYC*-induced apoptosis (Supplementary Figure 5c–d). Oncogenic roles for *MCL1* and *BCL2L1* have been previously suggested by reports of increased rates of lymphoma and leukemia 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.

The vast majority of significant focal SCNAs in each cancer type are prevalent across other 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.

The majority of 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; Figure 4) and, indeed, in the 158 regions from the overall pool. Nonetheless, cancer type-restricted analyses identified an additional 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 and likely to be related to the ring chromosomes in that disease[43]). However, even many of these regions were 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 vs. 11 Mb in the lineage-restricted analyses).

Arm-level alterations, like focal SCNAs, tend to be shared among multiple cancer types (Supplementary Note 4). Prior 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 www.broadinstitute.org/tumorscape, including segmented copy-number data (viewable using the Integrative Genomics Viewer [Robinson et al, 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).

Discussion

This study represents the largest analysis to date 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 strikingly high prevalence of arm-level SCNAs[4,5,6], which likely reflects the ease with which such mutational events occur compared to focal events[45,46]. The analysis also reveals a strong tendency for significant focal SCNAs in one cancer type to be also found in multiple 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 due 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 catalog of the significant SCNAs and define their occurrence in difference cancer types.

A key challenge is to identify the cancer gene targets of each of these SCNAs. Fewer than one-quarter of the 158 common peak regions are associated with previously validated targets of SCNAs in human cancer. While 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 due to mechanistic bias (e.g. 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 likely candidates, while the functional analysis of *MCL1* and *BCL2L1* strongly implicates these genes as amplification targets. Moreover, some SCNAs may contain multiple functional targets[10].

Identification of the target genes will require both genomic and functional studies. For focal events, the copy-number profiles of additional 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.

While many canonical oncogenes and tumor suppressor genes are known to be altered across multiple 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 toward genes relevant to multiple 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 combinations of a limited number of functionally relevant events.

Methods summary

DNA extracted from cancer specimens and normal tissue was labeled and hybridized to the Affymetrix 250K Sty I array to obtain signal intensities and genotype calls. Signal intensities were normalized against data from 1480 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 Figures 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. RNAi 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. Tumor 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].

Full methods are described in Supplementary Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

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

Affiliations

[1] Cancer Program, Medical and Population Genetics Group, The Broad Institute of M.I.T. and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA

[2] Departments of Medical Oncology, Pediatric Oncology, and Cancer Biology, and Center for Cancer Genome Discovery, Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA

[3] Departments of Medicine and Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA

[4] Departments of Medicine, Pathology, and Pediatrics, and Systems Biology, Harvard Medical School, Boston, MA 02115, USA

[5] Novartis Institutes for BioMedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

[6] Division of Molecular Epidemiology, Jikei University School of Medicine, 3-25-8 Nishi-shimbashi, Minato-ku, Tokyo 105-8461 Japan

[7] Department of Neurology, Children's Hospital Boston, Boston, MA 02115, USA

[8] Department of Internal Medicine, UT Southwestern Medical Center, Dallas, Texas 75390-9186

[9] Genetics Branch, Center for Cancer Research, National Cancer Institute and National Naval Medical Center, Bethesda, Maryland 20889, USA

[10] Department of Surgery II, Nagoya City University Medical School, Nagoya 467-8601, Japan

[11] Department of Genetics and Radiation Oncology, UNC/Lineberger Comprehensive Cancer Center, University of North Carolina, School of Medicine, Chapel Hill, NC 27599, USA

[12] Medical Oncology Program, Vall d'Hebron University Hospital Research Institute, Vall d'Hebron Institute of Oncology, and Autonomous University of Barcelona, 08035 Barcelona, Spain

[13] Department of Pathology and Division of Applied Molecular Oncology, University Health Network, Princess Margaret Hospital and Ontario Cancer Institute, Toronto, ON, Canada

[14] Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York

[15] Center for Human Genetic Research, Massachusetts General Hospital, Richard B. Simches Research Center, Boston, Massachusetts 02114, USA

[16] Department of Pathology, Beth Israel Deaconess Medical Center, 3 Blackfan Circle, Boston MA 02115, USAA

[17] Departments of Medicine and Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065

[18] Departments of Medicine (GI Division) and Genetics, and Abramson Cancer Center, University of Pennsylvania, 415 Curie Blvd. Philadelphia, PA 19104

[19] Section of Thoracic Surgery, Department of Surgery, University of Michigan, Ann Arbor, Ann Arbor, MI, 48109

[20] Department of Pathology, University of Washington Medical Center, 1959 NE Pacific St, Seattle, WA, 98195-6100

[21] Department of Obstetrics and Gynecology. Jikei University School of Medicine, 3-25-8 Nishi-shimbashi, Minato-ku, Tokyo 105-8461 Japan

[22] Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

[23] Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

Acknowledgments

References

1. Futreal PA, et al. A census of human cancer genes. Nat Rev Cancer. 2004; 4:177–183. [PubMed: 14993899]

2. Stuart D, Sellers WR. Linking somatic genetic alterations in cancer to therapeutics. Curr Opin Cell Biol. 2009; 21:304–310. [PubMed: 19328671]

3. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009; 458:719–724. [PubMed: 19360079]

4. Baudis M. Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data. BMC Cancer. 2007; 7:226. [PubMed: 18088415]

5. Mitelman F, Johansson B, Mertens F. Mitelman Database of Chromosome Aberrations in Cancer. 2009http://cgap.nci.nih.gov/Chromosomes/Mitelman

6. NCI and NCBI's SKY/M-FISH and CGH Database. 2001http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi

7. Weir BA, et al. Characterizing the cancer genome in lung adenocarcinoma. Nature. 2007; 450:893–898. [PubMed: 17982442]

8. Eder AM, 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 U S A. 2005; 102:12519–12524. [PubMed: 16116079]

9. Lahortiga I, et al. Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia. Nat Genet. 2007; 39:593–595. [PubMed: 17435759]

10. Zender L, et al. Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. Cell. 2006; 125:1253–1267. [PubMed: 16814713]

11. Mullighan CG, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature. 2007; 446:758–764. [PubMed: 17344859]

12. Wiedemeyer R, et al. Feedback circuit among INK4 tumor suppressors constrains human glioblastoma development. Cancer Cell. 2008; 13:355–364. [PubMed: 18394558]

13. Cancer_Genome_Atlas_Research_Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–1068. [PubMed: 18772890]

14. Chitale D, et al. An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. Oncogene. 2009; 28:2773–2783. [PubMed: 19525976]

15. GlaxoSmithKline. GSK Cancer Cell Line Genomic Profiling Data. 2008<https://cabig.nci.nih.gov/tools/caArray_GSKdata>

16. Mullighan CG, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. Nature. 2008; 453:110–114. [PubMed: 18408710]

17. Mazzarella R, Schlessinger D. Pathological Consequences of Sequence Duplications in the Human Genome. Genome Research. 1998; 8:1007. [PubMed: 9799789]

18. Beroukhim R, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci U S A. 2007; 104:20007–20012. [PubMed: 18077431]

19. Garraway LA, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature. 2005; 436:117–122. [PubMed: 16001072]

20. Raychaudhuri S, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. PLoS Genet. 2009; 5:e1000534. [PubMed: 19557189]

21. Letai AG. Diagnosing and exploiting cancer's addiction to blocks in apoptosis. Nat Rev Cancer. 2008; 8:121–132. [PubMed: 18202696]

22. Tsujimoto Y, Gorham J, Cossman J, Jaffe E, Croce CM. The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. Science. 1985; 229:1390–1393. [PubMed: 3929382]

23. Cleary ML, 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 U S A. 1985; 82:7439–7443. [PubMed: 2865728]

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. 1985; 41:899–906. [PubMed: 3924412]

25. Rampino N, et al. Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science. 1997; 275:967–969. [PubMed: 9020077]

26. Arena V, Martini M, Luongo M, Capelli A, Larocca LM. Mutations of the BIK gene in human peripheral B-cell lymphomas. Genes Chromosomes Cancer. 2003; 38:91–96. [PubMed: 12874789]

27. Wu C, Ponnappan U. NF-κB Signaling Pathway. 2009<http://www.biocarta.com/pathfiles/h_nfkbPathway.asp#contributors>

28. Boehm JS, et al. Integrative Genomic Approaches Identify IKBKE as a Breast Cancer Oncogene. Cell. 2007; 129:1065–1079. [PubMed: 17574021]

29. Barbie DA, et al. Systematic RNA interference reveals that oncogenic *KRAS*-driven cancers require TBK1. Nature. 2009; 462:108–112. [PubMed: 19847166]

30. Scholl C, et al. Synthetic Lethal Interaction Between Oncogenic *KRAS* Dependency and Suppression of *STK33* in Human Cancer Cells. Cell. 2009; 137:821–834. [PubMed: 19490892]

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. 2003; 94:1029–1033. [PubMed: 14662016]

32. Dearth RK, Cui X, Kim HJ, Hadsell DL, Lee AV. Oncogenic transformation by the signaling adaptor proteins insulin receptor substrate (IRS)-1 and IRS-2. Cell Cycle. 2007; 6:705–713. [PubMed: 17374994]

33. Luo B, et al. Highly parallel identification of essential genes in cancer cells. Proceedings of the National Academy of Sciences. 2008; 105:20380–20385.

34. Kendall J, et al. Oncogenic cooperation and coamplification of developmental transcription factor genes in lung cancer. Proc Natl Acad Sci U S A. 2007; 104:16663–16668. [PubMed: 17925434]

35. Lin WM, et al. Modeling genomic diversity and tumor dependency in malignant melanoma. Cancer Res. 2008; 68:664–673. [PubMed: 18245465]

36. Tonon G, et al. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci U S A. 2005; 102:9625–9630. [PubMed: 15983384]

37. Smith LT, et al. 20q11.1 amplification in giant-cell tumor of bone: Array CGH, FISH, and association with outcome. Genes Chromosomes Cancer. 2006; 45:957–966. [PubMed: 16847944]

38. Spits C, et al. Recurrent chromosomal abnormalities in human embryonic stem cells. Nat Biotechnol. 2008; 26:1361–1363. [PubMed: 19029912]

39. Lefort N, et al. Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. Nat Biotechnol. 2008; 26:1364–1366. [PubMed: 19029913]

40. Fanidi A, Harrington EA, Evan GI. Cooperative interaction between c-myc and bcl-2 proto-oncogenes. Nature. 1992; 359:554–556. [PubMed: 1406976]

41. Zhou P, et al. MCL1 transgenic mice exhibit a high incidence of B-cell lymphoma manifested as a spectrum of histologic subtypes. Blood. 2001; 97:3902–3909. [PubMed: 11389033]

42. Beverly LJ, Varmus HE. MYC-induced myeloid leukemogenesis is accelerated by all six members of the antiapoptotic BCL family. Oncogene. 2009; 28:1274–1279. [PubMed: 19137012]

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 Genetics and Cytogenetics. 1993; 68:85–90. [PubMed: 8353809]

44. Myllykangas S, Bohling T, Knuutila S. Specificity, selection and significance of gene amplifications in cancer. Semin Cancer Biol. 2007; 17:42–55. [PubMed: 17161620]

45. Albertson DG. Gene amplification in cancer. Trends in Genetics. 2006; 22:447–455. [PubMed: 16787682]

46. Fukasawa K. Centrosome amplification, chromosome instability and cancer development. Cancer Lett. 2005; 230:6–19. [PubMed: 16253756]

47. Smith DI, Zhu Y, McAvoy S, Kuhn R. Common fragile sites, extremely large genes, neural development and cancer. Cancer Letters. 2006; 232:48–57. [PubMed: 16221525]

48. Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics. 2004; 20:3413–3422. [PubMed: 15381628]

49. Lundberg AS, et al. Immortalization and transformation of primary human airway epithelial cells by gene transfer. Oncogene. 2002; 21:4577–4586. [PubMed: 12085236]
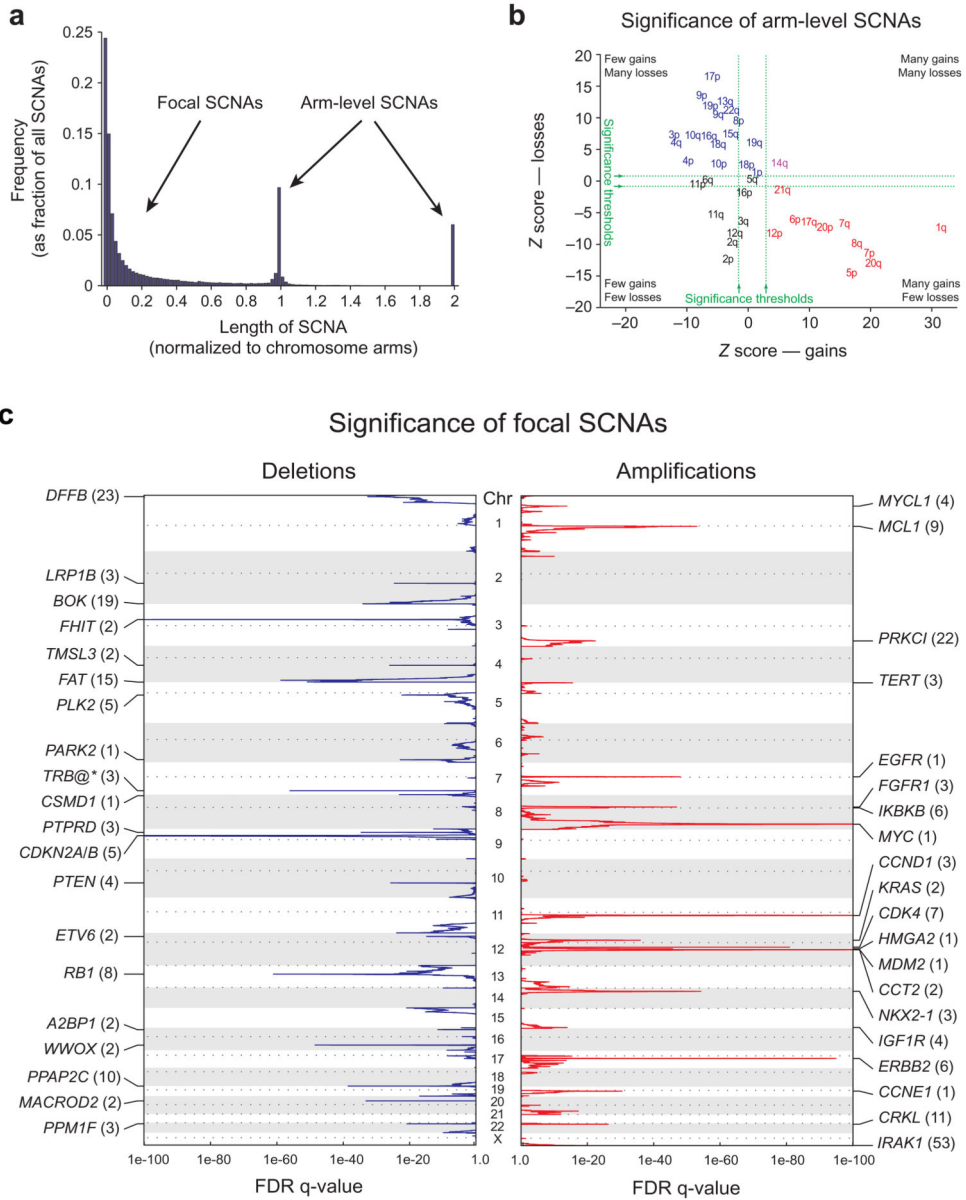
Figure 1.
Identification of significant arm-level and focal SCNAs across cancer. a) Length distribution of SCNAs. b) Length-adjusted Z-scores for gains (x-axis) and losses (y-axis) of indicated chromosome arms. Arms in red, blue, purple, and black exhibit significant gain, loss, both, or neither, respectively. c) 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.
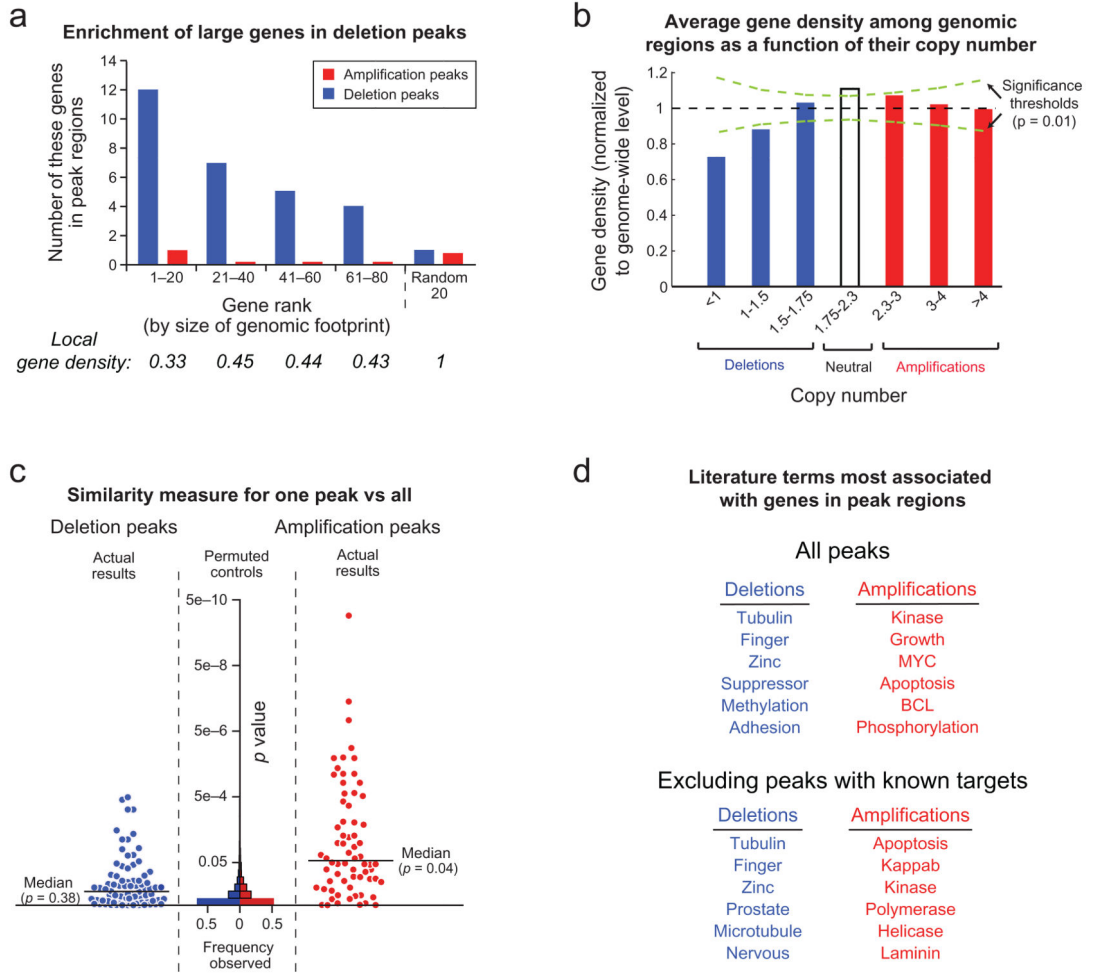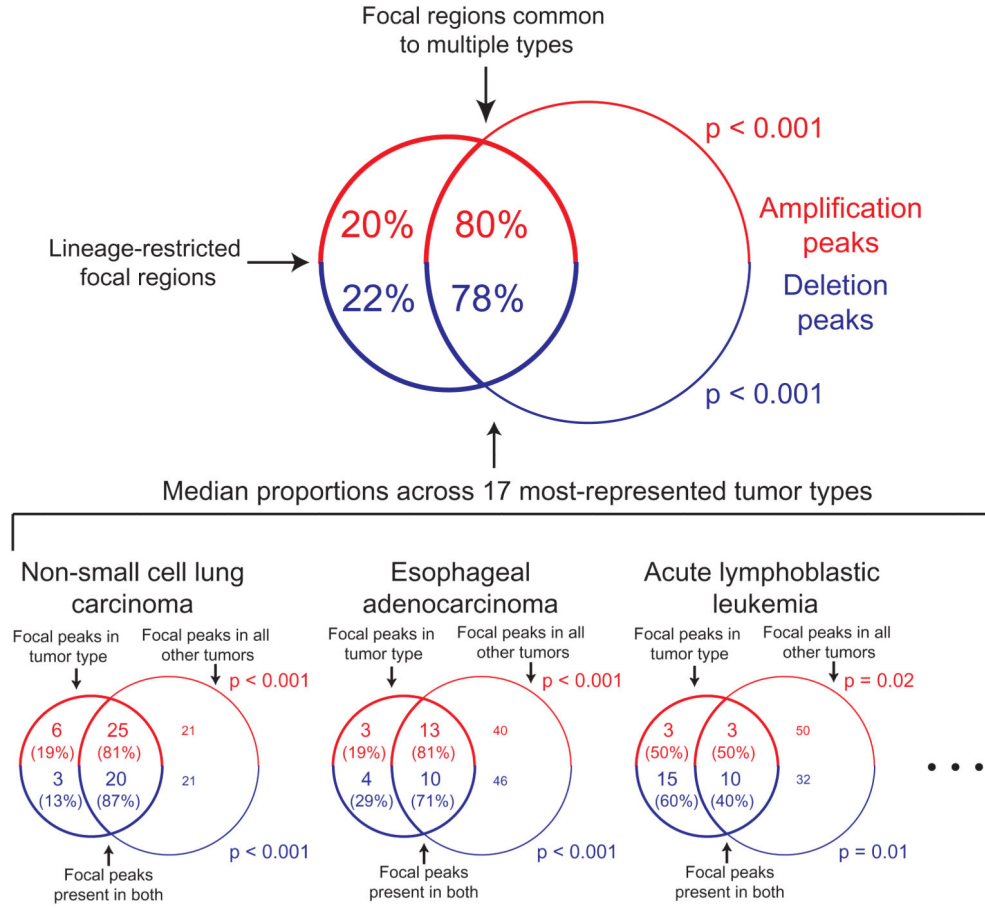
Figure 2.
Characteristics of significant focal SCNAs. a) Genes are ranked by the amount of genome occupied. Local gene density is normalized against the genome-wide average. b) Average gene density as a function of copy number. c) GRAIL analysis[20] p-values, plotted for each peak region, reflect the 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 displayed for randomly placed regions ("Permuted controls"). d) The literature terms most associated with genes in either deletion or amplification peaks, but not both.
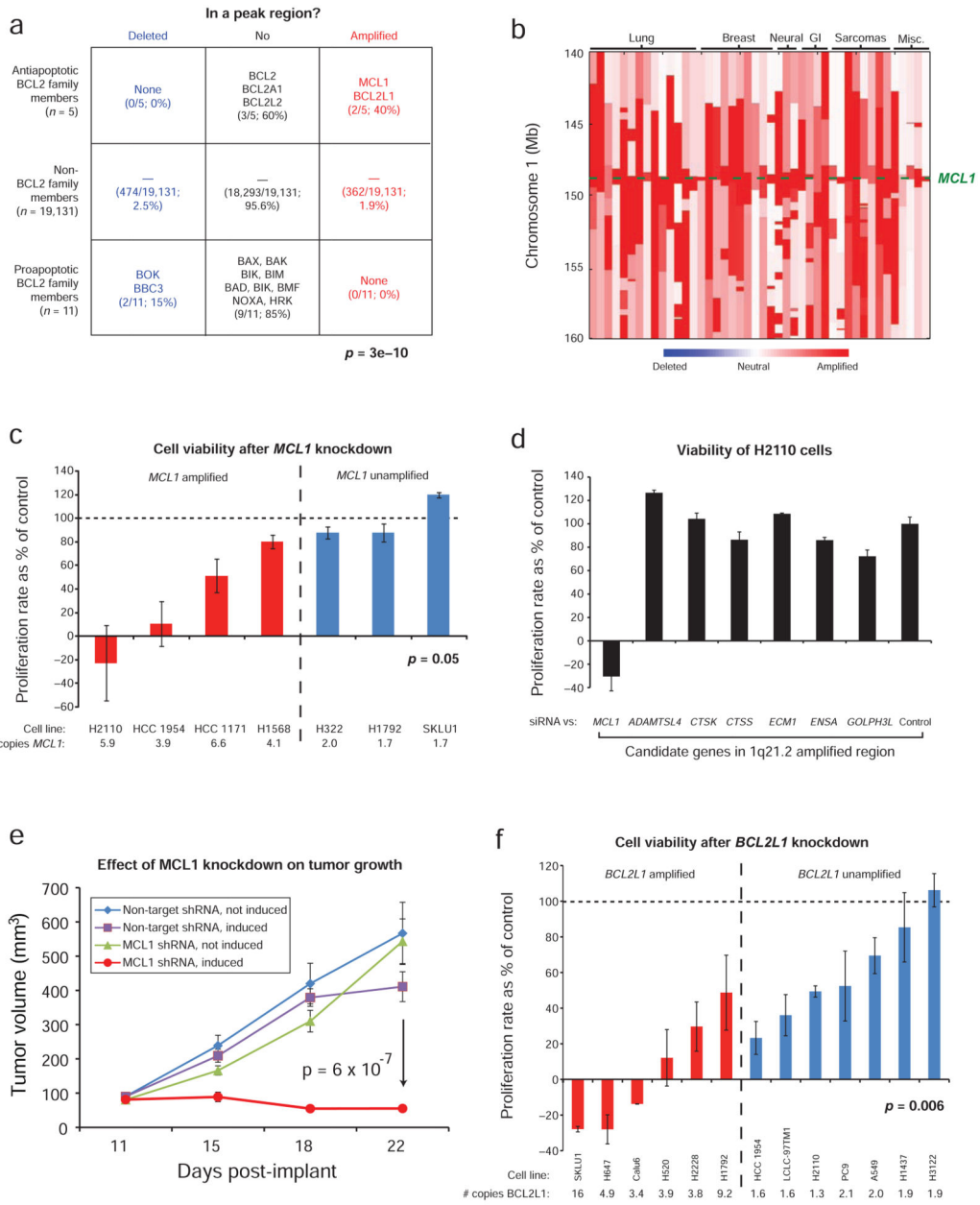
Figure 3.
Dependency of cancer cell lines on the amplified *BCL2* family members, *MCL1* and *BCL2L1*. a) Enrichment of pro- and anti-apoptotic *BCL2* family members deletion and amplification peaks. b) Copy-number profiles among 50 tumors around *MCL1* (lineages are across the top; genomic locations 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* knock-down on growth of NCI-H2110 xenografts. f) Changes in cell number after *BCL2L1* knockdown relative to controls. Error bars represent s.e.m. across 3 experiments.

Figure 4.
The majority of significant focal SCNA peaks identified in any one tumor type are also identified in the rest of the dataset (its complement). The top Venn diagram represents median results across the 17 tumor types represented by >40 samples. Venn diagrams representing the specific examples of non-small cell lung cancer, esophageal adenocarcinoma, and acute lymphoblastic leukemia are displayed along the bottom. Diagrams are not drawn to scale.