Comprehensive molecular characterization of human colon and rectal cancer

The Cancer Genome Atlas Network

To characterize somatic alterations in colorectal carcinoma, we conducted a genome-scale analysis of 276 samples, analysing exome sequence, DNA copy number, promoter methylation and messenger RNA and microRNA expression. A subset of these samples (97) underwent low-depth-of-coverage whole-genome sequencing. In total, 16% of colorectal carcinomas were found to be hypermutated: three-quarters of these had the expected high microsatellite instability, usually with hypermethylation and MLH1 silencing, and one-quarter had somatic mismatch-repair gene and polymerase ε (POLE) mutations. Excluding the hypermutated cancers, colon and rectum cancers were found to have considerably similar patterns of genomic alteration. Twenty-four genes were significantly mutated, and in addition to the expected APC, TP53, SMAD4, PIK3CA and KRAS mutations, we found frequent mutations in ARID1A, SOX9 and FAM123B. Recurrent copy-number alterations include potentially drug-targetable amplifications of ERBB2 and newly discovered amplification of IGF2. Recurrent chromosomal translocations include the fusion of NAV2 and WNT pathway member TCF7L1. Integrative analyses suggest new markers for aggressive colorectal carcinoma and an important role for MYC-directed transcriptional activation and repression.

The Cancer Genome Atlas project plans to profile genomic changes in 20 different cancer types and has so far published results on two cancer types12. We now present results from multidimensional analyses of human colorectal carcinoma (CRC).

CRC is an important contributor to cancer mortality and morbidity. The distinction between the colon and the rectum is largely anatomical, but it has both surgical and radiotherapeutic management implications and it may have an impact on prognosis. Most investigators divide CRC biologically into those with microsatellite instability (MSI; located primarily in the right colon and frequently associated with the CpG island methylator phenotype (CIMP) and hyper-mutation) and those that are microsatellite stable but chromosomally unstable.

A rich history of investigations (for a review see ref. 3) has uncovered several critical genes and pathways important in the initiation and progression of CRC (ref. 3). These include the WNT, RAS–MAPK, PI3K, TGF-β, P53 and DNA mismatch-repair pathways. Large-scale sequencing analyses4–6 have identified numerous recurrently mutated genes and a recurrent chromosomal translocation. Despite this background, we have not had a fully integrated view of the genetic and genomic changes and their significance for colorectal tumorigenesis. Further insight into these changes may enable deeper understanding of the pathophysiology of CRC and may identify potential therapeutic targets.

Results

Tumour and normal pairs were analysed by different platforms. The specific numbers of samples analysed by each platform are shown in Supplementary Table 1.

Exome-sequence analysis

To define the mutational spectrum, we performed exome capture DNA sequencing on 224 tumour and normal pairs (all mutations are listed in Supplementary Table 2). Sequencing achieved >20-fold coverage of at least 80% of targeted exons. The somatic mutation rates varied considerably among the samples. Some had mutation rates of <1 per 10⁶ bases, whereas a few had mutations rates of >100 per 10⁶. We separated cases (84%) with a mutation rate of <8.24 per 10⁶ (median number of non-silent mutations, 58) and those with mutation rates of >12 per 10⁶ (median number of total mutations, 728), which we designated as hypermutated (Fig. 1).

To assess the basis for the considerably different mutation rates, we evaluated MSI and mutations in the DNA mismatch-repair pathway4–8 genes MLH1, MLH3, MSH2, MSH3, MSH6 and PMS2. Among the 30 hypermutated tumours with a complete data set, 23 (77%) had high levels of MSI (MSI-H). Included in this group were 19 tumours with MLH1 methylation, 17 of which had CIMP. By comparison, the remaining seven hypermutated tumours, including the six with the highest mutation rates, lacked MSI-H, CIMP or MLH1 methylation but usually had somatic mutations in one or more mismatch-repair genes or POLE aberrations seen rarely in the non-hypermutated tumours (Fig. 1).

Gene mutations

Overall, we identified 32 somatic recurrently mutated genes (defined by MutSig11 and manual curation) in the hypermutated and non-hypermutated cancers (Fig. 1b). After removal of non-expressed genes, there were 15 and 17 in the hypermutated and non-hypermutated cancers, respectively (Fig. 1b; for a complete list see Supplementary Table 3). Among the non-hypermutated tumours, the eight most frequently mutated genes were APC, TP53, KRAS, PIK3CA, FBXW7, SMAD4, TCF7L2 and NRAS. As expected, the mutated KRAS and NRAS genes usually had oncogenic codon 12 and 13 or codon 61 mutations, whereas the remaining genes had inactivating mutations. CTNNB1, SMAD2, FAM123B (also known as WTX) and SOX9 were also mutated frequently. FAM123B is an X-linked negative regulator of WNT signalling6, and virtually all of its mutations were loss of function. Mutations in SOX9, a gene important for cell differentiation in the intestinal stem cell niche10,11, have not been associated previously with human cancer, but all nine mutated alleles in the non-hypermutated CRCs were frameshift or nonsense mutations. Tumour-suppressor genes listed in Supplementary Table 1.

©2012 Macmillan Publishers Limited. All rights reserved
two genes that were frequently mutated in the non-hypermutated
BRAF(V600E)
MSH6
Figure 1| Mutation frequencies in human CRC. a
black bars represent genes identified by manual examination of sequence data.
tumours. Blue bars represent genes identified by the MutSig algorithm and
Figure 2| Integrative analysis of genomic changes in 195 CRCs.
ACVR2A
Non-hypermutated tumours originating from different sites are virtually
TGFBR2
MSI-H showed additional differences in the mutational profile. When we specifically examined 28 genes with long mononucleotide
mutations. However, hypermutated tumours with MLH1 silencing and
mutations. BRAF(V600E) mutations have recently
TUMOR SITE
MSI STATUS
MLH1 SILENCING
Non-silent Silent Hypermutated Non-hypermutated
ACVR2A
MLH1 silencing
frameshift or nonsense mutations. ARID1A mutations have recently
been reported in CRC and many other cancers15,16.
In the hypermutated tumours, ACVR2A, APC, TGFBR2, MSH3, MSH6, SLC9A9 and TCF7L2 were frequent targets of mutation (Fig. 1b), along with mostly BRAF(V600E) mutations. However, two genes that were frequently mutated in the non-hypermutated
cancers were significantly less frequently mutated in hypermutated
tumours: TP53 (60 versus 20%, P < 0.0001) and APC (81% versus
51%, P = 0.0023; both Fisher’s exact test). Other genes, including
TGFB2R2, were mutated recurrently in the hypermutated cancers, but not in the non-hypermutated samples. These findings indicate that hypermutated and non-hypermutated tumours progress through
different sequences of genetic events.
As expected, hypermutated tumours with MLH1 silencing and
MSI-H showed additional differences in the mutational profile. When we specifically examined 28 genes with long mononucleotide repeats in their coding sequences, we found that the rate of frameshift
mutation was 3.6-fold higher than the rate of such mutations in
hypermutated tumours without MLH1 silencing and 50-fold higher
than that in non-hypermethylated tumours (Supplementary Table 2).

Mutation rate and methylation patterns
As mentioned above, patients with colon and rectal tumours are
managed differently17, and epidemiology also highlights differences between the two17. An initial integrative analysis of MSI status,
somatic copy-number alterations (SCNAs), CIMP status and gene-
expression profiles of 132 colonic and 62 rectal tumours enabled us to examine possible biological differences between tumours in the two
locations. Among the non-hypermutated tumours, however, the
overall patterns of changes in copy number, CIMP, mRNA and
miRNA were indistinguishable between colon and rectal carcinomas (Fig. 2). On the basis of this result, we merged the two for all
subsequent analyses.
Unsupervised clustering of the promoter DNA methylation
profiles of 236 colorectal tumours identified four subgroups (Supple-
mentary Fig. 1 and Supplementary Methods). Two of the clusters
contained tumours with elevated rates of methylation and were
classified as CIMP high and CIMP low, as previously described18.
The two non-CIMP clusters were predominantly from tumours that
were non-hypermutated and derived from different anatomic locations. mRNA expression profiles separated the colorectal tumours
into three distinct clusters (Supplementary Fig. 2). One significantly
overlapped with CIMP-high tumours (P = 3 × 10−12) and was
enriched with hypermutated tumours, and the other two clusters
did not correspond with any group in the methylation data.
Analysis of miRNA expression by unsupervised clustering (Supple-
mentary Fig. 3) identified no clear distinctions between rectal cancers and non-hypermethylated colon cancers.
Chromosomal and sub-chromosomal changes

In total, 257 tumours were profiled for SCNs with Affymetrix SNP 6.0 arrays. Of these tumours, 97 were also analysed by low-depth-of-coverage (low-pass) whole-genome sequencing. As expected, the hypermutated tumours had far fewer SCNs (Fig. 2). No difference was found between microsatellite-stable and -unstable hypermutated tumours (Supplementary Fig. 4). We used the GISTIC algorithm\textsuperscript{19} to identify probable gene targets of focal alterations. There were several previously well-defined arm-level changes, including gains of 1q, 7p and q, 8p and q, 12q, 13q, 19q, and 20p and q (ref. 6). (Supplementary Fig. 4 and Supplementary Table 4). Significantly deleted chromosome arms were 18p and q (including SMAD4) in 66% of the tumours and 17p and q (including TP53) in 56%. Other significantly deleted chromosome arms were 1p, 4q, 5q, 8p, 14q, 15q, 20p and 22q.

We identified 28 recurrent deletion peaks (Supplementary Fig. 4 and Supplementary Table 4), including the genes FHIT, RBFOX1 and WWOX with large genomic footprints located in potentially fragile sites of the genome, in near-diploid hypermutated tumours. Other focal deletions involved tumour-suppressor genes such as SMAD4, APC, PTEN and SMAD3. A significant focal deletion of 10p25.2 spanned four genes, including TCF7L2, which was also frequently mutated in our data set. A gene fusion between adjacent genes VTTIA and TCF7L2 through an interstitial deletion was found in 3% of CRCs and is required for survival of CRC cells bearing the translocation\textsuperscript{4}.

There were 17 regions of significant focal amplification (Supplementary Table 4). Some of these were superimposed on broad gains of chromosome arms, and included a peak at 13q12.13 near the peptidase-coding gene USP12 and at ~500 kb distal to the CRC candidate oncogene CDK8; an adjacent peak at 13q12; a peak containing KLF5 at 13q22.1; and a peak at 20q13.12 adjacent to HNF4A. Peaks on chromosome 8 included 8p12 (which contains the histone methyl-transferase-coding gene WHSC1L1, adjacent to FGFR1) and 8q24 (which contains MYC). An ampiclon at 17q21.1, found in 4% of the tumours, contains seven genes, including the tyrosine kinase ERBB2. ERBB2 amplifications have been described in colon, breast and gastro–oesophageal tumours, and breast and gastric cancers bearing these amplifications have been treated effectively with the anti-ERBB2 antibody trastuzumab\textsuperscript{20–22}.

One of the most common focal amplifications, found in 7% of the tumours, is the gain of a 100–150-kb region of the chromosome arm 11p15.5. It contains genes encoding insulin (INS), insulin-like growth factor 2 (IGF2) and tyrosine hydroxylase (TH), as well as miR-483, which is embedded within IGF2 (Fig. 3a). We found elevated expression of IGF2 and miR-483 but not of INS and TH (Fig. 3b, c). Immediately adjacent to the amplified region is ASCL2, a transcription factor active in specifying intestinal stem-cell fate\textsuperscript{23}. Although ASCL2 has been implicated as a target of amplification in CRC\textsuperscript{23–25}, it was consistently outside the region of amplification and its expression was not correlated with copy-number changes. These observations suggest that IGF2 and miR-483 are candidate functional targets of 11p15.5 amplification. IGF2 overexpression through loss of imprinting has been implicated in the promotion of CRC\textsuperscript{26, 27}. MiR-483 may also have a role in CRC pathogenesis\textsuperscript{26}.

A subset of tumours without IGF2 amplification (15%) also had considerably higher levels of IGF2 gene expression (as much as a 100-fold increase), an effect not attributable to methylation changes at the IGF2 promoter. To assess the context of IGF2 amplification/
overexpression, we systematically searched for mutually exclusive genomic events using the MEMo method. We found a pattern of near exclusivity (corrected P < 0.01) of IGF2 overexpression with genomic events known to activate the PI3K pathway (mutations of PIK3CA and PIK3R1 or deletion/mutation of PTEN; Fig. 3c and Supplementary Table 5). The IRS2 gene, encoding a protein linking IGFR1 (the receptor for IGF2) with PI3K, is on chromosome 13, which is frequently gained in CRC. The cases with the highest IRS2 expression were mutually exclusive of the cases with IGF2 overexpression (P = 0.04) and also lacked mutations in the PI3K pathway (P = 0.0001; Fig. 3c). These results strongly suggest that the IGF2–IGF1R–IRS2 axis signals to PI3K in CRC and imply that therapeutic targeting of the pathway could act to block PI3K activity in this subset of patients.

**Translocations**

To identify new chromosomal translocations, we performed low-pass, paired-end, whole-genome sequencing on 97 tumours with matched normal samples. In each case we achieved sequence coverage of ~3–4-fold and a corresponding physical coverage of 7.5–10-fold. Despite the low genome coverage, we detected 250 candidate interchromosomal translocation events (range, 0–10 per tumour). Among these events, 212 had one or both breakpoints in an intergenic region, whereas the remaining 38 juxtaposed coding regions of two genes in putative fusion events, of which 18 were predicted to code for in-frame events (Supplementary Table 6). We found three separate cases in which the first two exons of the NAV2 gene on chromosome 11 are joined with the 3’ coding portion of TCF7L1 on chromosome 2 (Supplementary Fig. 5). TCF7L1 encodes TCF3, a member of the TCF/LEF class of transcription factors that heterodimerize with nuclear β-catenin to enable β-catenin-mediated transcriptional regulation. Intriguingly, in all three cases, the predicted structure of the NAV2–TCF7L1 fusion protein lacks the TCF3 β-catenin-binding domain. This translocation is similar to another recurrent translocation identified in CRC, a fusion in which the amino terminus of VTI1A is joined to TCF4, which is encoded by TCF7L2 on chromosome 2 (Supplementary Fig. 6). Eleven of the 19 (58%) gene–gene translocations were validated by obtaining PCR products or, in some cases, sequencing the junction fragments (Supplementary Fig. 5).

**Figure 4 | Diversity and frequency of genetic changes leading to deregulation of signalling pathways in CRC.** Non-hypermutated (nHM; n = 165) and hypermutated (HM; n = 30) samples with complete data were analysed separately. Alterations are defined by somatic mutations, homozygous deletions, high-level focal amplifications, and, in some cases, by significant up- or downregulation of gene expression (IGF2, FZD10, SMAD4). Alteration frequencies are expressed as a percentage of all cases. Red denotes activated genes and blue denotes inactivated genes. Bottom panel shows for each sample if at least one gene in each of the five pathways described in this figure is altered.

**Altered pathways in CRC**

Integrated analysis of mutations, copy number and mRNA expression changes in 195 tumours with complete data enriched our understanding of how some well-defined pathways are deregulated. We grouped samples by hypermutation status and identified recurrent alterations in the WNT, MAPK, PI3K, TGF-β and p53 pathways (Fig. 4, Supplementary Fig. 6 and Supplementary Table 1).

We found that the WNT signalling pathway was altered in 93% of all tumours, including biallelic inactivation of APC (Supplementary Table 7) or activating mutations of CTNNB1 in ~80% of cases. There were also mutations in SOX9 and mutations and deletions in TCF7L2, as well as the DKK family members and AXIN2, FBXW7 (Supplementary Fig. 7), ARID1A and FAM123B (the latter is a negative regulator of WNT–β-catenin signalling found mutated in Wilms’ tumour). A few mutations in FAM123B have previously been described in CRC. SOX9 has been suggested to have a role in cancer, but no mutations have previously been described. The WNT receptor frizzled (FZD10) was overexpressed in ~17% of samples, in some instances at levels of 100× normal. Altogether, we found 16 different altered WNT pathway genes, confirming the importance of this pathway in CRC. Interestingly, many of these alterations were found in tumours that harbour APC mutations, suggesting that multiple lesions affecting the WNT signalling pathway confer selective advantage.

Genetic alterations in the PI3K and RAS–MAPK pathways are common in CRC. In addition to IGF2 and IRS2 overexpression, we found mutually exclusive mutations in PIK3R1 and PIK3CA as well as deletions in PTEN in 2%, 15% and 4% of non-hypermutated tumours, respectively. We found that 55% of non-hypermutated tumours have alterations in KRAS, NRAS or BRAF, with a significant pattern of mutual exclusivity (Supplementary Fig. 6 and Supplementary Table 1). We also evaluated mutations in the erythropoietic leukemia viral oncogene homolog (ERBB) family of receptors because of the translational relevance of such mutations. Mutations or amplifications in
one of the four ERBB family genes are present in 22 out of 165 (13%) non-hypermutated and 16 out of 30 (53%) hypermutated cases. Some of the mutations are listed in the COSMIC database\textsuperscript{33}, suggesting a functional role. Intriguingly, recurrent ERBB2(V842I) and ERBB3(V104M) mutations were found in four and two non-hypermutated cases, respectively. Mutations and focal amplifications of ERBB2 (Supplementary Fig. 6) should be evaluated as predictors of response to agents that target those receptors. We observed co-occurrence of alterations involving the RAS and PI3K pathways in one-third of tumours (Fig. 4; \(P = 0.039\), Fisher’s exact test). These results indicate that simultaneous inhibition of the RAS and PI3K pathways may be required to achieve therapeutic benefit.

The TGF-\(\beta\) signalling pathway is known to be deregulated in CRC and other cancers\textsuperscript{34}. We found genomic alterations in TGFBR1, TGFBR2, ACVR2A, ACVR1B, SMAD2, SMAD3 and SMAD4 in 27% of the non-hypermutated and 87% of the hypermutated tumours. We also evaluated the p53 pathway, finding alterations in TP53 in 59% of non-hypermutated cases (mostly biallelic; Supplementary Table 8) and alterations in ATM, a kinase that phosphorylates and activates p53 after DNA damage, in 7%. Alterations in these two genes showed a trend towards mutual exclusivity (\(P = 0.016\)) (Fig. 4, Supplementary Fig. 6 and Supplementary Table 1).

We integrated copy number, gene expression, methylation and pathway data using the PARADIGM software platform\textsuperscript{35}. The analysis showed a number of new characteristics of CRC (Fig. 5a). For example, despite the diversity in anatomical origin or mutation levels, nearly 100% of these tumours have changes in MYC transcriptional targets, both those promoted by and those inhibited by MYC. These findings are consistent with patterns deduced from genetic alterations (Fig. 4) and suggest an important role for MYC in CRC. The analysis also identified several gene networks altered across all tumour samples and those with differential alterations in hypermutated versus non-hypermutated samples (Supplementary Table 7, Supplementary Data on the Cancer Genome Atlas publication webpage).

Because most of the tumours used in this study were derived from a prospective collection, survival data are not available. However, the tumours can be classified as aggressive or non-aggressive on the basis of tumour stage, lymph node status, distant metastasis and vascular invasion at the time of surgery. We found numerous molecular signatures associated with tumour aggressiveness, a subset of which is shown in Fig. 5b. They include specific focal amplifications and deletions, and altered gene-expression levels, including those of miRNAs and specific somatic mutations (APC, TP53, PIK3CA, BRAF and FBXW7; Supplementary Fig. 8b). Mutations in FBXW7 (38 cases) and distant metastasis (32 cases) never co-occurred (\(P = 0.0019\)). Interestingly, a number of genomic regions have multiple molecular associations with tumour aggressiveness that manifest as clinically related genomic hotspots. Examples of this are the region 20q13.12, which includes a focal amplification and multiple genes correlating with tumour aggression, and the region 22q12.3, containing APOL6 (ref. 37) (Supplementary Figures 8 and 9).

**Discussion**

This comprehensive integrative analysis of 224 colorectal tumour and normal pairs provides a number of insights into the biology of CRC and identifies potential therapeutic targets. To identify possible biological differences in colon and rectum tumours, we found, in the non-hypermutated tumours irrespective of their anatomical origin, the same type of copy number, expression profile, DNA methylation and miRNA changes. Over 94% had a mutation in one or more members of the WNT signalling pathway, predominantly in APC. However, there were some differences between tumours from the right colon and all other sites. Hypermethylation was more common in the right colon, and three-quarters of hypermutated samples came from the same site, although not all of them had MSI (Fig. 2). Why most of the hypermutated samples came from the right colon and why there are two classes of tumours at this site is not known. The origins of the colon from embryonic midgut and hindgut may provide an explanation. As the survival rate of patients with high MSI-related cancers is better and these cancers are hypermutated, mutation rate may be a better prognostic indicator.
Whole-exome sequencing and integrative analysis of genomic data provided further insights into the pathways that are dysregulated in CRC. We found that 93% of non-hypermutated and 97% of hyper-mutated cases had a deregulated WNT signalling pathway. New findings included recurrent mutations in FAM123B, ARID1A and SOX9 and very high levels of overexpression of the WNT ligand receptor gene FZD10. To our knowledge, SOX9 has not previously been described as frequently mutated in any human cancer. SOX9 is transcriptionally repressed by WNT signalling, and the SOX9 protein has been shown to facilitate β-catenin degradation38. ARID1A is frequently mutated in gynaecological cancers and has been shown to suppress MYC transcription39. Activation of WNT signalling and inactivation of the TGF-β signalling pathway are known to result in activation of MYC. Our mutational and integrative analyses emphasize the critical role of MYC in CRC. We also compared our results with other large-scale analyses43,44. Gene-expression profiles were generated using Agilent microarrays and a subset of samples was subjected to low-pass (2–5 x) denoising. Although some differences among batches were detected, we did not correct them and DNA methylation data sets using a combination of cluster analysis, enhanced characterization and sequencing centres (Supplementary Methods). The bio-resources, and aliquots of purified nucleic acids were shipped to the genome specimen core resources provided sample sets in several different batches. To assess any batch effects we examined the miRNA expression, miRNA expression and DNA methylation data sets using a combination of cluster analysis, enhanced principal component analysis and analysis of variance (Supplementary Methods).

Our analyses show that non-hypermutated adenocarcinomas of the colon and rectum are not distinguishable at the genomic level. However, tumours from the right/ascending colon were more likely to be hypermethylated and to have elevated mutation rates than were other CRCs. As has been recognized previously, activation of the WNT signalling pathway and inactivation of the TGF-β signalling pathway, resulting in increased activity of MYC, are nearly ubiquitous events in CRC. Genetic aberrations frequently target the MAPK and PI3K pathways but less frequently target receptor tyrosine kinases. In conclusion, the data presented here provide a useful resource for understanding this deadly disease and identifying possibilities for treating it in a targeted way.

METHODS SUMMARY

Tumour and normal samples were processed by either of two biospecimen core resources, and aliquots of purified nucleic acids were shipped to the genome specimen core resources provided sample sets in several different batches.

All of the primary sequence files are deposited in dbGap and all other data are deposited at the Data Coordinating Center (DCC) for public access (http://cancergenome.nih.gov/). Data matrices and supporting data can be found at http://tcga-data.nci.nih.gov/docs/publications/coadread_2012. The data can also be explored through the ISB Regulome Explorer (http://explorer.cancerregulome.org/). Next Generation Clustered Heatmaps (http://bioinformatics.mdanderson.org/main/TGCA/Supplements/NCGHM-CRC) and the cBio Cancer Genomics Portal (http://cbioportal.org). Descriptions of the data can be found at https://wiki.nci.nih.gov/x/j5dXAg and in Supplementary Methods.

Received 15 November 2011; accepted 22 May 2012.

1. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455, 1061–1068 (2008).
2. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature 474, 609–615 (2011).
3. Fearon, E. R. Molecular genetics of colorectal cancer. Annu. Rev. Pathol. 6, 479–507 (2011).
4. Bass, A. J. et al. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent V71A–TCF7L2 fusion. Nature Genet. 43, 964–968 (2011).
5. Sjoblom, T. et al. The consensus coding sequences of human breast and colorectal cancers. Science 314, 268–274 (2006).
6. Wood, L. D. et al. The genomic landscapes of human breast and colorectal cancers. Science 318, 1108–1113 (2007).
7. Umar, A. et al. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J. Natl Cancer Inst 95, 261–268 (2004).
8. Aaltonen, L. A. et al. Clues to the pathogenesis of familial colorectal cancer. 260, Science 812–816 (1993).
9. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 363, 558–561 (1993).
10. Parsons, R. et al. Hypermethylation and mismatch repair deficiency in RER+ tumour cells. Cell 75, 1227–1236 (1993).
11. Dooley, A. L. et al. Nuclear factor I/B is an oncogene in small cell lung cancer. Genes Dev. 25, 1470–1475 (2011).
12. Major, M. B. et al. Wilms tumour suppressor WTX negatively regulates WNT/β-catenin signalling. Science 316, 1043–1046 (2007).
13. Mori-Akiyama, Y. et al. SOX9 is required for the differentiation of paneth cells in the intestinal epithelium. Gastroenterology 133, 539–546 (2007).
14. Bastide, P. et al. Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. J. Cell Biol. 178, 635–648 (2007).
15. O’Reilly, E. et al. Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. Hum. Mutat. 33, 100–103 (2012).
16. Wilson, B. G. & Roberts, C. W. SWI/SNF nucleosome remodelers and cancer. Nat. Rev. Cancer 11, 481–492 (2011).
17. Moysky, B. D. Unique considerations in the patient with rectal cancer. Semin. Oncol. 38, 542–551 (2011).
18. Hinoue, T. et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome Res. 22, 271–282 (2012).
19. 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).
20. Camps, J. et al. Integrative genomics reveals mechanisms of copy number alterations responsible for transcriptional deregulation in colorectal cancer. Genes Chromosomes Cancer 48, 452–468 (2009).
21. Varley, J. M., Swallow, J. E., Brammar, J. L. & Walker, R. A. Alterations to either c-erb-B2 (neu) or c-myc proto-oncogenes in breast carcinomas correlate with poor short-term prognosis. Oncogene 1, 423–430 (1987).
22. Koizumi, J. et al. Amplification of c-erb-B-2 oncogene in human adenocarcinomas in vivo. Lancet 327, 765–767 (1986).
23. van der Flier, L. G. et al. The genomic landscapes of human breast and colorectal cancers. Proc. Natl Acad. Sci. USA 104, 591–596 (2007).
24. Veronese, A. et al. Oncogenic role of miR-483-3p at the GSF2/G63 locus. Cancer Res. 70, 3140–3149 (2010).
25. Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Res. 22, 398–406 (2012).
26. Brady, C. A. et al. Distinct pathways of transcriptional programs dictate acute DNA damage responses and tumor suppression. Cell 145, 571–583 (2011).
27. Rivera, M. N. et al. An X chromosome gene, WTX, is commonly inactivated in Wilms tumor. Science 315, 642–645 (2007).
28. Scheier, S. K. et al. Mutations in the WTX-gene are found in some high microsatellite unstable (MSI-H) colorectal cancers. BMC Cancer 10, 413 (2010).
29. Forbes, S. A. et al. The catalogue of somatic mutations in cancer (COSMIC). Curr. Protoc. Hum. Genet. Ch. 10. Unit 10.11 (2008).
30. Massagué, J., Blain, S. W. & L. Roberts, C. W. TGF-β signaling in growth control, cancer, and heritable disorders. Cell 103, 295–300 (2001).
31. Vaske, C. J. et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomic data using PARADIGM. Bioinformatics 26, 237–245 (2010).
32. House, C. D. et al. Voltage-gated Na+ channel SCN5A is a key regulator of a gene transcriptional network that controls colon cancer invasion. Cancer Res. 70, 6957–6967 (2010).
33. Liu, Z., Lu, H., Jiang, Z., Pasztuszyn, A. & Hu, C. A. Apolipoprotein B6-only protein, induces mitochondria-mediated apoptosis in cancer cells. Mol. Cancer Res. 3, 21–31 (2005).
Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements
This work was supported by the following grants from the National Institutes of Health: U24CA137599, U24CA138358, U24CA138410, U24CA138435, U24CA138484, U24CA138388, U24CA138466, U24CA138467, U24CA138824, U24CA138838, U24CA140254, US4HG003067, US4HG003079 and US4HG003273.

Author Contributions
The Cancer Genome Atlas research network contributed collectively to this study. Biospecimens were provided by the tissue source sites and projects. Data were generated in the Cancer Genome-Characterization Centers and genome data analysis centers. All data were released through the Data Coordinating Center. Project activities were coordinated by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) project teams. Project leaders were R.K. and D.A.W. Writing team: T.A., A.J.B., T.A.C., L.D., S.J.H., R.R., P.K., L.E., J.C., J.H., J.W., F.W., R. Wheeler, R. A. Gibbs.

Genome Sequencing Center Washington University in St. Louis Li Ding10, Robert S. Fulton, Daniel C. Koboldt, Todd Wylie9, Jason Walker9, David J. Dooling9,10, Lucinda Fulton9, Preethi Gunaratne1, N. Bainbridge1, Kyle Chang1, Huyen H. Dinh1, Jennifer A. Drummond1, Gerald F. MacNeil1, Jackie Juhl60, Andrew Kaufman52, Bernard Kohl III61, Xuan Van Le61, Maria C. Mariano62, Joel Frith63, Richard W. Park18, Wu-Wen Li59, ENCODE Project Consortium64, Lee Hood58, Michael Robinson59, Robert F. Gifford58, and Charles S. Facciotti.

Genome Sequencing Center UCSC and the Buck Institute Lawrence4, Douglas Voet4, Gordon Saksena4, Gordon Sacks4, Benjamin A. J. LaDuca4, Benjamin H. Neel4, and Jennifer C. Ross4.

Genome Sequencing Center Broad Institute Michael A. Deininger66, Jennifer A. Seiler66, Kelly J. Hirst12, Robert A. Holt12, Darlene Lee12, Haiyan I. Li12, Michael Mayo12, Richard A. Horvitz12, Hye-Jung E. Chun12, Andrew J. Mungall12, Erin Pleasance12, A. Gordon Robertson12, Trevor Oldham12, David S. Lytwyn12, Jennifer L. Smith12, Dan S. Hartley12, Gregory O. Tebbs12, Matthew A. Zabel12, Andrew J. Leach12, Jacob A. Marsh12, Joshua T. Stolzenberg12, Michael R. Rosenblatt12, and Andrew D. Baker12.

Genome Sequencing Center Broad Institute Nicholas J. Coates16, Samuel J. Teich16, and J. Zachary Sanborn16.

Genome Sequencing Center Baylor College of Medicine Donna M. Muzny1, Matthew N. Bainbridge1, Kyle Chang1, Huyen H. Dinh1, Jennifer A. Drummond1, Keri Ford2, Michael B. Morgan3, Irene F. Newsham4, Jeffrey G. Reid5, Jirin Santibanez-Koref6, Eve Shinbrot6, Lisa R. Trevino6, Yuan-Qing Wu6, Mind Wang6, Preethi Gunaratne6, and Nisreen Y. Hamady6.

Genome Sequencing Center Children's Hospital Dana-Farber Cancer Institute Adam J. Pinkel7, Michael S. Lawrence7, Steven A. Montgomery7, Alex M. Lander7, Paul M. Feinberg7, and Dennis J. Weisenberger7.

Genome Sequencing Center Dana-Farber Cancer Institute and National Human Genome Research Institute Project teams. Project leaders were R.K. and D.A.W. Writing team: T.A., A.J.B., T.A.C., L.D., S.J.H., R.R., P.K., L.E., J.C., J.H., J.W., F.W., R. Wheeler, R. A. Gibbs.

Genome Sequencing Center Duke University Michael B. Love17, D. Russell Darland17, Xi Chen17, Erin E. Rehm17, Benjamin D. Brown17, L. Justin Tuck17, and Jonathan W. Young17.

Genome Sequencing Center EU-ENCODE5,6 Project Daniel H. Silver18, Benjamin I. Stoeckert Jr18, and Andrew J. Leach18.

Genome Sequencing Center EU-ENCODE Project James L. Brown19, Adam J. Pinkel20, and Michael B. Morgan20.

Genome Sequencing Center Genome-Characterization Centers University of Southern California and Johns Hopkins University Hospit300, Yutaka Shinawi300, and Robert E. Shippy300.

Genome Sequencing Center Hawaii Cancer Center and the University of Hawaii Michael A. Deininger66, Jennifer A. Seiler66, Kelly J. Hirst12, Robert A. Holt12, Darlene Lee12, Haiyan I. Li12, Michael Mayo12, Richard A. Horvitz12, Hye-Jung E. Chun12, Andrew J. Mungall12, Erin Pleasance12, A. Gordon Robertson12, Trevor Oldham12, David S. Lytwyn12, Jennifer L. Smith12, Dan S. Hartley12, Gregory O. Tebbs12, Matthew A. Zabel12, Andrew J. Leach12, Jacob A. Marsh12, Joshua T. Stolzenberg12, Michael R. Rosenblatt12, and Andrew D. Baker12.

Genome Sequencing Center Institute for Systems Biology Christine S. Thorvaldsdottir23,24, Gordon Saksena4, and Tavares D. Scheffer23,24.

Genome Sequencing Center Institute for Systems Biology Jennifer A. Demuth25, John F. Dorsky25, T.A. Chappel25, and C. Sobolevsky25.

Genome Sequencing Center Institute for Systems Biology John R. Buckler26,27,28, and Andrew S. McKeown28.

Genome Sequencing Center Institute for Systems Biology Randal D. W. Jones29,30, and James L. Thomas29,30.

Genome Sequencing Center Institute for Systems Biology Robert Dorfman31,32,33, and Michael J. Shiekhattar31,32,33.

Genome Sequencing Center Institute for Systems Biology Michelle A. Johnson34, Andrew R. Kurcok34, and Elliot S. Maskin34.

Genome Sequencing Center Institute for Systems Biology Martin J. C. Ryan35,36,37, and Walter L. Kwarn35,36,37.

Genome Sequencing Center Institute for Systems Biology Shinya Nakanishi38,39, Ting Wang40,41,42, and Elena Weissman43.

Genome Sequencing Center Institute for Systems Biology Emmanouil K. Dermitzakis44,45, Constantinos Tsikas44,45, and Artis Skliros44,45.

Genome Sequencing Center Instituto Productions46,47,48, and Raul E. de la Torre46,47,48.

Genome Sequencing Center Instituto Productions James E. Sharp49,50,51, and Kathy J. DeTar49,50,51.

Genome Sequencing Center Instituto Productions Mark D. Groudine49,50,51, and S. Robert Fanning49,50,51.

Genome Sequencing Center Instituto Productions Alex H. Shiekhattar49, and Kevin L. Gray49,50,51.

Genome Sequencing Center Instituto Productions Michael J. Price49,50,51, and James T. Peterson49,50,51.

Genome Sequencing Center Instituto Productions Michael J. Price49,50,51, and James T. Peterson49,50,51.

Genome Sequencing Center Instituto Productions Michael J. Price49,50,51, and James T. Peterson49,50,51.

Genome Sequencing Center Instituto Productions Michael J. Price49,50,51, and James T. Peterson49,50,51.

Genome Sequencing Center Instituto Productions Michael J. Price49,50,51, and James T. Peterson49,50,51.

Genome Sequencing Center Instituto Productions Michael J. Price49,50,51, and James T. Peterson49,50,51.

Genome Sequencing Center Instituto Productions Michael J. Price49,50,51, and James T. Peterson49,50,51.

Genome Sequencing Center Instituto Productions Michael J. Price49,50,51, and James T. Peterson49,50,51.
