Published in final edited form as:
Nature. ; 487(7407): 330–337. doi:10.1038/nature11252.

Comprehensive Molecular Characterization of Human Colon and Rectal Cancer

The Cancer Genome Atlas Network

Summary

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

Background

The Cancer Genome Atlas (TCGA) project plans to profile genomic changes in 20 different cancer types and has published results on two cancer types[1,2]. We now present results from multidimensional analyses of human colorectal cancer (CRC).

CRC is an important contributor to cancer mortality and morbidity. The distinction between colon and rectum is largely anatomical, but it impacts both surgical and radiotherapeutic management and it may impact 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 (MSS) but chromosomally unstable (CIN).

A rich history of investigations (for a review see[3]) has revealed several critical genes and pathways important to the initiation and progression of CRC[3]. These include the WNT, RAS-MAPK, PI3K, TGF-β, P53 and DNA mismatch repair pathways. Large-scale sequencing analyses[4–6] have identified numerous recurrently mutated genes and a recurrent chromosomal translocation. Despite that background, we have not had a fully integrated view of the genetic and genomic changes and their significance for colorectal tumorigenesis. Further insight into those changes may enable deeper understanding of the pathophysiology of CRC and may identify potential therapeutic targets.

Communicating authors: Raju Kucherlapati and David Wheeler, Address Correspondence to: Raju Kucherlapati, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, 617-525-4445, rkucherlapati@partners.org.

Results

Tumor/normal pairs were analyzed by different platforms. The specific numbers of samples analyzed 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 tumor/normal pairs (Supplementary Table 2 lists all mutations). Sequencing achieved >20X coverage of at least 80% of targeted exons. The somatic mutation rates varied considerably among the samples. Some had mutation rates $<1/10^6$ bases, whereas a few had mutations rates $>100/10^6$. We separated those cases (84%) with a mutation rate $<8.24/10^6$ (median number of non-synonymous mutations: 58) and those with mutations rates $>12/10^6$ (median number of mutations: 728), which we designated as hypermutated (Figure 1).

To assess the basis for the strikingly different mutation rates, we evaluated microsatellite instability (MSI)[7] and mutations in the DNA mismatch repair pathway[8–10] genes *MLH1, MLH3, MSH2, MSH3, MSH6* and *PMS2*. Among the 30 hypermutated tumors with a complete data set, 23 (77%) had high levels of MSI (MSI-H). Included were 19 with MLH1 methylation, 17 of which had high CpG island methylation phenotype (CIMP). By comparison, the remaining seven hypermutated tumors, 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 Polε aberrations rarely seen in the non-hypermutated tumors (Figure 1).

Gene mutations

Overall, we identified 32 somatic recurrently mutated genes (defined by MutSig[11] and manual curation) in the hypermutated and non-hypermutated cancers (Figure 1B). After removal of non-expressed genes, there were 15 and 17, respectively, in the hypermutated and non-hypermutated cancers (Figure 1B, see Supplementary Table 3 for complete list). Among the non-hypermutated tumors, 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 12/13 or 61 mutations, whereas the remaining genes had inactivating mutations. *CTNNB1, SMAD2, FAM123B* and *SOX9* were also mutated frequently. *FAM123B (WTX)* is an X-linked negative regulator of *WNT* signaling[12], and virtually all its mutations were loss-of-function. Mutations in *SOX9*, a gene important in cell differentiation in the intestinal stem cell niche[13,14], have not been associated previously with human cancer, but all nine mutated alleles in the non-hypermutated CRCs were frameshift or nonsense mutations. Tumor suppressors *ATM* and *ARID1A* also had a disproportionately high number of frameshift or nonsense mutations. *ARID1A* mutations have recently been reported in CRC and many other cancers[15,16].

In the hypermutated tumors, *ACVR2A, APC*, TGFBR2, MSH3, MSH6, SLC9A9 and TCF7L2 were frequent targets of mutation (Figure 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 tumors: *TP53* (60 *vs* 20%, $p < 0.0001$), and *APC* (81% *vs* 51%, $p = 0.0023$, both Fisher's exact test). Other genes, including *TGFBR2*, were recurrently mutated in the hypermutated cancers, but not in the non-hypermutated samples. These findings suggest that hypermutated and non-hypermutated tumors progress through different sequences of genetic events.

As expected, hypermutated tumors with MLH1 silencing and MSI-H exhibited additional differences in 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 mutation in hypermutated tumors without MLH1 silencing, and 50-fold higher than in non-hypermethylated tumors (Supplementary Table 2).

Classification of tumors based on mutation rate and methylation pattern

As mentioned above, patients with colon and rectal tumors are managed differently[17], and epidemiology also shows differences between the two[17]. An initial integrative analysis of MSI status, somatic copy number alterations (SCNAs), CIMP status and gene expression profiles of 132 colonic and 62 rectal tumors enabled us to examine possible biological differences between tumors in the two locations. Among the non-hypermutated tumors, however, the overall patterns of changes in copy number, CIMP, mRNA and miRNA were indistinguishable between colon and rectal carcinomas (Figure 2). Based on that result, we merged the two for all subsequent analyses.

Unsupervised clustering of the promoter DNA methylation profiles of 236 colorectal tumors revealed four subgroups (Supplementary Methods; Supplementary Figure 1). Two of the clusters contained tumors with elevated rates of methylation and were classified as CIMP-high (CIMP-H) and CIMP-low (CIMP-L), as previously described[18]. The two non-CIMP clusters were predominantly from tumors that were non-hypermutated and derived from different anatomic locations. mRNA expression profiles separated the colorectal tumors into three distinct clusters (Supplementary Figure 2). One significantly overlapped with CIMP-H tumors ($p=3\times10^{-12}$) and was enriched with hypermutated tumors; the other two clusters did not correspond with any group in the methylation data. Analysis of miRNA expression by unsupervised clustering (Supplementary Figure 3) identified no clear distinctions between rectal cancers and non-hypermethylated colon cancers.

Chromosomal and sub-chromosomal changes

257 tumors were profiled for somatic copy-number alterations (SCNAs) with Affymetrix SNP 6.0 arrays. Of those tumors, 97 were also analyzed by low depth-of-coverage (low-pass) whole-genome sequencing (WGS). As expected, the hypermutated tumors had far fewer SCNAs (Figure 2). No difference was found between MSI and MSS hypermutated tumors (Supplementary Figure 4). We used the GISTIC algorithm[19] to identify likely gene targets of focal alterations. There were several previously well-defined arm-level changes, including gains of 1q, 7p/q, 8p/q, 12q, 13q, 19q, and 20p/q[6] (Supplementary Figure 4; Supplementary Table 4). Significantly deleted chromosome arms were 18p/q (including *SMAD4*) in 66% of the tumors and 17p/q (including *TP53*) in 56%. Also significantly deleted genes were 1p, 4q, 5q, 8p, 14q, 15q, 20p, and 22q.

We identified 28 recurrent deletion peaks (Supplementary Table 4; Supplementary Figure 4), including genes like *FHIT, A2BP1* and *WWOX* with large genomic footprints located in potentially fragile sites of the genome, in near-diploid hypermutated tumors. Other focal deletions involved tumor 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 dataset. A gene fusion between adjacent genes *VTI1A* and *TCF7L2* through an interstitial deletion was found in 3% of CRCs and is required for survival of CRC cells bearing the translocation[4].

There were 17 regions of significant focal amplification (Supplementary Table 4). Some of them were superimposed on broad gains of chromosome arms. Included were a peak at 13q12.13 near the peptidase gene *USP12* and ~500kB 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 *WHSC1L1*, adjacent to *FGFR1*) and 8q24 (which contains *MYC*). An amplicon at 17q21.1, found in 4% of the tumors, contains seven genes, including the tyrosine kinase *ERBB2*. *ERBB2* amplifications have been described in colon, breast and gastric/esophageal tumors, and breast and gastric cancers bearing these amplifications have been treated effectively with the anti-ERBB2 antibody trastuzumab[20–22].

One of the most common focal amplifications, found in 7% of the tumors, is gain of a 100–150 kb region of chromosome arm 11p15.5. It contains the genes encoding insulin (*INS*), insulin-like growth factor 2 (*IGF2*), and tyrosine hydroxylase (*TH*), as well as miR-483, which is embedded within *IGF2* (Figure 3a). We found elevated expression of *IGF2* and miR-483 but not of *INS* and *TH* (Figure 3b–c). Immediately adjacent to the amplified region is *ASCL2*, a transcription factor active in specifying intestinal stem cell fate[23]. Although *ASCL2* has been implicated as a target of amplification in CRC[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[26,27]. MiR-483 may also play a role in CRC pathogenesis[28].

A subset of tumors (15%) without *IGF2* amplification also had dramatically higher levels (as much as 100X) of *IGF2* gene expression, an effect not attributable to methylation changes at the *IGF2* promoter. To assess the context of *IGF2* amplification/overexpression, we systematically looked for mutually exclusive genomic events using the MEMo method[29]. We found a pattern of near exclusivity (corrected $p < 0.01$) of *IGF2* overexpression with genomic events known to activate the PI3-K pathway (mutations of *PIK3CA/PIK3R1* or deletion/mutation of *PTEN*, Figure 3c, and Supplementary Table 5). The *IRS2* gene, whose product links IGF1R, the receptor for IGF2, with PI3-K, is on chromosome 13, which is frequently gained in colorectal cancer. Those cases with the highest *IRS2* expression were mutually exclusive of the cases with *IGF2* overexpression ($p= 0.04$) and also lacked mutations in the PI3-K pathway ($p= 0.0001$)(Figure 3c). Those results strongly suggest that the IGF2/IGF1R/IRS2 axis signals to PI3-K in CRC and imply that therapeutic targeting of the pathway could act to block PI3-K activity in this subset of patients.

Translocations

To identify novel chromosomal translocations, we performed low-pass, paired-end, whole-genome sequencing on 97 tumors with matched normals. In each case we achieved sequence coverage of ~3–4X and a corresponding physical coverage of 7.5–10X. Despite the low genome coverage, we detected 250 candidate inter-chromosomal translocation events (range 0–10/tumor). Among those 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 Figure 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 that is encoded by *TCF7L2*, deleted or mutated in 12% of non-hypermutated tumors and a homolog of *TCF7L1*[4]. We also observed 21 cases of translocation involving TTC28 located on Chromosome 22 (Supplementary Table 6). In all cases the fusions predict inactivation of TTC28, which has been identified as a target of p53 and an inhibitor of tumor cell growth[30]. Eleven of the 19 (58%) gene-gene translocations are

validated by either obtaining PCR products and in some case sequencing the junction fragments (Supplementary Figure 5).

Altered pathways in CRC

Integrated analysis of mutations, copy-number, and mRNA expression changes in 195 tumors 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 (Figure 4, Supplementary Figure 6, Supplementary Table 1).

We found that the WNT signaling pathway was altered in 93% of all tumors, 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 Figure 7), *ARID1A and FAM123B/WTX* (the latter a negative regulator of WNT/β-catenin signaling[12] found mutated in Wilm's tumor[31]). A few mutations in *FAM123B/WTX* have been described in colorectal cancer[32]. SOX9 has been suggested to play 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 100X normal. Altogether, we found 16 different altered WNT pathway genes, confirming the importance of that pathway in CRC. Interestingly, many of those alterations were found in tumors that harbor *APC* mutations, suggesting that multiple lesions affecting the WNT signaling 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 tumors, respectively. We found that 55% of non-hypermutated tumors have alterations in *KRAS, NRAS* or *BRAF*, with a significant pattern of mutual exclusivity (Supplementary Figure 6, Supplementary Table 1). We also evaluated mutations in the ERBB family of receptors because of the translational relevance of such mutations. Mutations or amplifications in one of the four genes are present in 22/165 (13%) non-hypermutated and 16/30 (53%) hypermutated cases. Some of the mutations are listed in the COSMIC database[33], suggesting a functional role. Intriguingly, recurrent V842I ERBB2 and V104M ERBB3 mutations were found in four and two non-hypermutated cases, respectively. Mutations and focal amplifications of ERBB2 (Supplementary Figure 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 a third of tumors (Figure 4; Fisher's exact test $p = 0.039$). These results suggest that simultaneous inhibition of the RAS and PI3K pathways may be required to achieve therapeutic benefit.

The TGF-β signaling pathway is known to be deregulated in colorectal and other cancers[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 tumors. 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 following DNA damage, in 7%. Alterations in those two genes showed a trend towards mutual exclusivity ($p = 0.016$) (Figure 4, Supplementary Figure 6, Supplementary Table 1).

We integrated copy number, gene expression, methylation and pathway data using the PARADIGM software platform[35]. The analysis revealed a number of novel characteristics of CRC (Figure 5A). For example, despite the diversity in anatomical origin or mutation

levels, nearly 100% of these tumors 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 (Figure 4) and suggest an important role for *MYC* in the CRC. The analysis also identified several gene networks altered across all tumor samples and those with differential alterations in hypermutated vs. non-hypermutated samples (Supplementary Table 7, Supplemental Data on the TCGA publication webpage).

Since most of the tumors used in this study were derived from prospective collection, survival data are not available. However, the tumors can be classified as aggressive or non-aggressive on the basis of tumor stage, lymph node status, distant metastasis and vascular invasion at the time of surgery. We found numerous molecular signatures associated with tumor aggressiveness, a subset of which is shown in Figure 5B. They include specific focal amplifications and deletions, and altered gene expression levels, including those of SCN5A[36], a reported regulator of colon cancer invasion (full list: Supplementary Tables 10–11). Association with tumor aggressiveness is also observed in altered expression of miRNAs and specific somatic mutations (APC, TP53, PIK3CA, BRAF, and FBXW7; Supplementary Figure 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 tumor 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 tumor aggression, and the region 22q12.3, containing APOL6[37] (Supplementary Figures 8–9).

Discussion

This comprehensive integrative analysis of 224 colorectal tumor/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 tumors we found, in the non-hypermutated tumors, irrespective of their anatomical origin, the same type of copy number, expression profile, DNA methylation and miRNA changes. Over 94% of them had a mutation in one or more members of the WNT signaling pathway, predominantly in *APC*. However, there were some differences between tumors from the right colon and the remaining 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 (Figure 2). Why most of the hypermutated samples come from the right colon and why there are two classes of tumors at this site is not known. The origins of the colon from embryonic midgut and hindgut may provide an explanation. Since the survival of patients with high MSI cancers are better and these cancers have hypermutation, 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 hypermutated cases had deregulated WNT signaling pathway. Novel findings included recurrent mutations in *FAM123B, ARID1A* and *SOX9* and very high levels of overexpression of the WNT ligand receptor Frizzled 10. To our knowledge, SOX9 has not previously been described as frequently mutated in any human cancer. SOX9 is transcriptionally repressed by WNT signaling, and the SOX9 protein has been shown to facilitate β-catenin degradation[38]. ARID1A is frequently mutated in gynecological cancers and has been shown to suppress Myc transcription[39]. Activation of WNT signaling and inactivation of the TGF-β signaling 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 analyses[6] and find many similarities and few differences in mutated genes (Supplementary Table 3).

Our integrated analysis revealed a diverse set of changes in TCF/LEF encoding genes suggesting additional roles for TCF/LEF factors in CRC beyond being passive partners for β-catenin.

Our data suggest a number of therapeutic approaches to CRC. Included are WNT signaling inhibitors and small-molecule β-catenin inhibitors that are showing initial promise[40–42]. We find that several proteins in the RTK/RAS and PI3K pathways including IGF2, IGFR, ERBB2, ERBB3, MEK, AKT and mTOR could be targets for inhibition.

Our analyses show that non-hypermutated adenocarcinomas of the colon and rectum are not distinguishable at the genomic level. However, tumors 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 signaling pathway and inactivation of the TGF-β signaling pathway, resulting in increased activity of MYC, are nearly ubiquitous events in CRC. Genomic aberrations frequently target the MAPK and PI3-K pathways but less frequently target receptor tyrosine kinases. In conclusion, the data presented here provide an unprecedented resource for understanding this deadly disease and identifying possibilities for treating it in a targeted way.

Methods Summary

Tumor and normal samples were processed by either of two Biospecimen Core Resources (BCRs), and aliquots of purified nucleic acids were shipped to the genome characterization and sequencing centers (Supplementary Methods). The BCRs provided sample sets in several different batches. To assess any batch effects we examined the mRNA expression, miRNA expression and DNA methylation data sets using a combination of cluster analysis, enhanced principal component analysis, and analysis of variance (Supplementary Methods). Although some differences among batches were detected, we did not correct them computationally because the differences were generally modest and because some of them may reflect biological phenomena (Supplementary Methods).

We used Affymetrix SNP 6.0 microarrays to detect copy-number alterations. A subset of samples was subjected to low pass (2–5X) whole genome sequencing (Illumina HiSeq), in part for detection of SCNA and chromosomal translocations[43,44]. Gene expression profiles were generated using Agilent microarrays and RNA-Seq. DNA methylation data were obtained using Illumina Infinium (HumanMethylation27) arrays. DNA sequencing of coding regions was performed by exome capture followed by sequencing on the SOLiD or Illumina HiSeq platforms. Details of the analytical methods used are described in Supplementary Methods.

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 via the ISB Regulome Explorer (http://explorer.cancerregulome.org/) 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Charles Fuchs for reviewing the manuscript prior to submission. This work was supported by the following grants from the USA National Institutes of Health: U24CA143799, U24CA143835, U24CA143840, U24CA143843, U24CA143845, U24CA143848, U24CA143858, U24CA143866, U24CA143867, U24CA143882, U24CA143883, U24CA144025, U54HG003067, U54HG003079, U54HG003273.

Appendix

Authors

Author Contributions: The TCGA research network contributed collectively to this study. Biospecimens were provided by the Tissue Source Sites and processed by the Biospecimen Core Resource. Data generation and analyses were performed by the Genome Sequencing Centers, Cancer Genome Characterization Centers, and Genome Data Analysis Centers. All data were released through the Data Coordinating Center. Project activities were coordinated by the NCI and NHGRI Project Teams. Project Leaders: Raju Kucherlapati and David A. Wheeler; **Writing Team:** Todd Auman, Adam J. Bass, Timothy A. Chan, Lawrence Donehower, Angela Hadjipanayis, Stanley R. Hamilton, Raju Kucherlapati, Peter W. Laird, Matthew Meyerson, Nikolaus Schultz, Ilya Shmulevich, Joshua M. Stuart, Joel Tepper, Vesteinn Thorsson, David A. Wheeler. **Mutations:** Michael S. Lawrence, Lisa R. Trevino, David A. Wheeler, Gad Getz; **Copy-Number and Structural Aberrations:** Alex H. Ramos, Adam J. Bass, Angela Hadjipanayis, Peng-Chieh Chen; **DNA Methylation:** Toshinori Hinoue; **Expression:** J. Todd Auman; **miRNA:** Gordon Robertson, Andy Chu; **Pathways**: Chad J. Creighton, Lawrence Donehower, Ted Goldstein, Sam Ng, Jorma de Ronde, Chris Sander, Nikolaus Schultz, Joshua M. Stuart, & Vesteinn Thorsson.

Genome Sequencing Centers

Baylor College of Medicine – Donna M. Muzny(1), Matthew N. Bainbridge(1), Kyle Chang(1), Huyen H. Dinh(1), Jennifer A. Drummond(1), Gerald Fowler(1), Christie L. Kovar(1), Lora R. Lewis(1), Margaret B. Morgan(1), Irene F. Newsham(1), Jeffrey G. Reid(1), Jireh Santibanez(1), Eve Shinbrot(1), Lisa R. Trevino(1), Yuan-Qing Wu(1), Min Wang(1), Preethi Gunaratne(1,2), Lawrence A. Donehower(1,3), Chad J. Creighton(1,3), David A. Wheeler(1), Richard A. Gibbs(1), *Broad Institute* – Michael S. Lawrence(4), Douglas Voet(4), Rui Jing(4), Kristian Cibulskis(5), Andrey Sivachenko(3), Petar Stojanov(4), Aaron McKenna(4), Eric S. Lander(4,6,7), Stacey Gabriel(8), Gad Getz(4), *Washington University in St. Louis* – Li Ding(9,10), Robert S. Fulton(9), Daniel C. Koboldt(9), Todd Wylie(9), Jason Walker(9), David J. Dooling(9,10), Lucinda Fulton(9), Kim D. Delehaunty(9), Catrina C. Fronick(9), Ryan Demeter(9), Elaine R. Mardis(9,10,11), & Richard K. Wilson(9,10,11)

Genome Characterization Centers

BC Cancer Agency – Andy Chu(12), Hye-Jung E. Chun(12), Andrew J. Mungall(12), Erin Pleasance(12), A. Gordon Robertson(12), Dominik Stoll(12), Miruna Balasundaram(12), Inanc Birol(12), Yaron S.N. Butterfield(12), Eric Chuah(12), Robin J.N. Coope(12), Noreen Dhalla(12), Ranabir Guin(12), Carrie Hirst(12), Martin Hirst(12), Robert A. Holt(12), Darlene Lee(12), Haiyan I. Li(12), Michael Mayo(12), Richard A. Moore(12), Jacqueline E. Schein(12), Jared R. Slobodan(12), Angela Tam(12), Nina Thiessen(12), Richard Varhol(12), Thomas Zeng(12), Yongjun Zhao(12), Steven J.M. Jones(12), Marco A. Marra(12), *Broad Institute* – Adam J. Bass(4,13), Alex H. Ramos(4,13), Gordon Saksena(4), Andrew D. Cherniack(4), Stephen E. Schumacher(4,13), Barbara Tabak(4,13), Scott L. Carter(4,13), Nam H. Pho(4), Huy Nguyen(4), Robert C. Onofrio(4), Andrew Crenshaw(4),

Kristin Ardlie(4), Rameen Beroukhim(4,13), Wendy Winckler(4), Gad Getz(4), Matthew Meyerson(4,13,14), *Brigham and Women's Hospital and Harvard Medical School* – Alexei Protopopov (15), Juinhua Zhang (15), Angela Hadjipanayis (16,17), Eunjung Lee (17,18), Ruibin Xi (18), Lixing Yang(18), Xiaojia Ren(15), Hailei Zhang (15), Narayanan Sathiamoorthy (19), Sachet Shukla (15), Peng-Chieh Chen (16,17), Psalm Haseley (17,18), Yonghong Xiao (15), Semin Lee (18), Jonathan Seidman (16), Lynda Chin (4,15,20), Peter J. Park (17,18,19), Raju Kucherlapati (16,17), *University of North Carolina, Chapel Hill*: J. Todd Auman(21,22), Katherine A. Hoadley(23,24,25), Ying Du(25), Matthew D. Wilkerson(25), Yan Shi(25), Christina Liquori (25), Shaowu Meng(25), Ling Li(25), Yidi J. Turman(25), Michael D. Topal(24,25), Donghui Tan(26), Scot Waring(25), Elizabeth Buda(25), Jesse Walsh(25), Corbin D. Jones(27), Piotr A. Mieczkowski(23), Darshan Singh(28), Junyuan Wu(25), Anisha Gulabani(25), Peter Dolina(25), Tom Bodenheimer(25), Alan P. Hoyle(25), Janae V. Simons(25), Matthew Soloway(25), Lisle E. Mose(24), Stuart R. Jefferys(24), Saianand Balu(25), Brian D. O'Connor(25), Jan F. Prins(28), Derek Y. Chiang(23,25), D. Neil Hayes(25,29), Charles M. Perou(23,24,25), *University of Southern California / Johns Hopkins* – Toshinori Hinoue(30), Daniel J. Weisenberger(30), Dennis T. Maglinte(30), Fei Pan(30), Benjamin P. Berman(30), David J. Van Den Berg(30), Hui Shen(30), Timothy Triche Jr(30), Stephen B. Baylin(31), & Peter W. Laird(30)

Genome Data Analysis Centers

Broad Institute – Gad Getz(4), Michael Noble(4), Doug Voet(4), Gordon Saksena(4), Nils Gehlenborg (18,4), Daniel DiCara (4), Juinhua Zhang(4,15), Hailei Zhang(4,15), Chang-Jiun Wu(4,15), Spring Yingchun Liu(4,15), Sachet Shukla(4,15), Michael S. Lawrence(4), Lihua Zhou(4), Andrey Sivachenko(4), Pei Lin(4), Petar Stojanov(4), Rui Jing(4), Richard W. Park(18), Marc-Danie Nazaire(4), Jim Robinson(4), Helga Thorvaldsdottir(4), Jill Mesirov(4), Peter J.Park(17,18,19), Lynda Chin(4,15,20), *Institute for Systems Biology* – Vesteinn Thorsson(32), Sheila M. Reynolds(32), Brady Bernard(32), Richard Kreisberg(32), Jake Lin(32), Lisa Iype(32), Ryan Bressler(32), Timo Erkkilä(32), Madhumati Gundapuneni(32), Yuexin Liu(33), Adam Norberg(32), Tom Robinson (32), Da Yang(33), Wei Zhang (33), Ilya Shmulevich(32), *Memorial Sloan-Kettering Cancer Center* – Jorma J. de Ronde(34,35), Nikolaus Schultz(34), Ethan Cerami(34), Giovanni Ciriello(34), Arthur P. Goldberg(34), Benjamin Gross(34), Anders Jacobsen(34), Jianjiong Gao(34), Bogumil Kaczkowski(34), Rileen Sinha(34), B. Arman Aksoy(34), Yevgeniy Antipin(34), Boris Reva(34), Ronglai Shen(36), Barry S. Taylor(34), Timothy A. Chan(37), Marc Ladanyi(38), Chris Sander(34), *The University of Texas MD Anderson Cancer Center* – Rehan Akbani(39), Nianxiang Zhang(39), Bradley M. Broom(39), Tod Casasent(39), Anna Unruh(39), Chris Wakefield(39), Stanley R. Hamilton(33), R. Craig Cason (33), Keith A. Baggerly(39), John N. Weinstein(39,40), *University of California, Santa Cruz / Buck Institute* – David Haussler(41,42), Christopher C. Benz(43), Joshua M. Stuart(41), Stephen C. Benz(41), J. Zachary Sanborn(41), Charles J. Vaske(41), Jingchun Zhu(41), Christopher Szeto(41), Gary K. Scott(43), & Christina Yau(43). Sam Ng(41), Ted Goldstein(41), Kyle Ellrott(41), Eric Collisson(44), Aaron E. Cozen(41), Daniel Zerbino(41), Christopher Wilks(41), Brian Craft(41) & Paul Spellman (45)

Biospecimen Core Resources

International Genomics Consortium – Robert Penny(46), Troy Shelton(46), Martha Hatfield(46), Scott Morris(46), Peggy Yena(46), Candace Shelton(46), Mark Sherman(46), & Joseph Paulauskis(46)

Nationwide Children's Hospital Biospecimen Core Resource – Julie M. Gastier-Foster(47,48,49), Jay Bowen(47), Nilsa C. Ramirez(47,48), Aaron Black(47), Robert Pyatt(47,48), Lisa Wise(47), & Peter White(47,49)

Tissue Source Sites and Disease working group

Monica Bertagnolli(50), Jen Brown(51), Timothy A. Chan(52), Gerald C. Chu(53), Christine Czerwinski(51), Fred Denstman(54), Rajiv Dhir(55), Arnulf Dörner(56), Charles S. Fuchs(57,58), Jose G. Guillem(59), Mary Iacocca(51), Hartmut Juhl(60), Andrew Kaufman(52), Bernard Kohl III(61), Xuan Van Le(61), Maria C. Mariano(62), Elizabeth N. Medina(62), Michael Meyers(63), Garrett M. Nash(59), Phillip B. Paty(59), Nicholas Petrelli(54), Brenda Rabeno(51), William G. Richards(64), David Solit(66), Pat Swanson(51), Larissa Temple(52), Joel E. Tepper(65), Richard Thorp(61), Efsevia Vakiani(62), Martin R. Weiser(59), Joseph E. Willis (67), Gary Witkin(51), Zhaoshi Zeng(59), Michael J. Zinner(63), & Carsten Zornig(68)

Data Coordination Center

Mark A. Jensen(69), Robert Sfeir(69), Ari B. Kahn(69), Anna L. Chu(69), Prachi Kothiyal(69), Zhining Wang(69), Eric E. Snyder(69), Joan Pontius(69), Todd D. Pihl(69), Brenda Ayala(69), Mark Backus(69), Jessica Walton(69), Jon Whitmore(69), Julien Baboud(69), Dominique L. Berton(69), Matthew C. Nicholls(69), Deepak Srinivasan(69), Rohini Raman(69), Stanley Girshik(69), Peter A. Kigonya(69), Shelley Alonso(69), Rashmi N. Sanbhadti(69), Sean P. Barletta(69), John M. Greene(69) & David A. Pot(69)

Project Team

National Cancer Institute – Kenna R. Mills Shaw(70), Laura A. L. Dillon(70), Ken Buetow(71), Tanja Davidsen(71), John A. Demchok(70), Greg Eley(72), Martin Ferguson(73), Peter Fielding(70), Carl Schaefer(71), Margi Sheth(70), & Liming Yang(70)

National Human Genome Research Institute: Mark S. Guyer(74), Bradley A. Ozenberger(74), Jacqueline D. Palchik(74), Jane Peterson(74), Heidi J. Sofia(74), & Elizabeth Thomson(74)

1. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030
2. Department of Biology and Biochemistry, University of Houston, Houston TX 77204
3. Dan L. Duncan Cancer Center, Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030 USA
4. The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University Cambridge, MA 02142 USA
5. Medical Sequencing Analysis and Informatics, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge MA 02142 USA
6. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142 USA
7. Department of Systems Biology, Harvard University, Boston, MA 02115 USA

8. Genetic Analysis Platform, The Eli and Edythe L. Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142 USA
9. The Genome Institute, Washington University School of Medicine, St. Louis, MO 63108 USA
10. Department of Genetics, Washington University School of Medicine, St. Louis, MO 63108 USA
11. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO 63108 USA
12. Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, BC V5Z, Canada
13. Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115 USA
14. Department of Pathology, Harvard Medical School, Boston, MA 02115 USA
15. Belfer Institute for Applied Cancer Science, Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115 USA
16. Department of Genetics, Harvard Medical School, Boston, MA 02115 USA
17. Division of Genetics, Brigham and Women's Hospital, Boston, MA 02115 USA
18. The Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115 USA
19. Informatics Program, Children's Hospital, Boston, MA 02115 USA
20. Department of Dermatology, Harvard Medical School, Boston, MA 02115 USA
21. Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
22. Institute for Pharmacogenetics and Individualized Therapy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
23. Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
24. Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, Chapel Hill, NC 27599 USA
25. Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
26. Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
27. Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
28. Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
29. Department of Internal Medicine, Division of Medical Oncology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA
30. University of Southern California Epigenome Center, University of Southern California, Los Angeles, CA 90089 USA

31. Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, MD 21231 USA

32. Institute for Systems Biology, Seattle, WA 98109 USA

33. Division of Pathology and Laboratory Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA

34. Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

35. Divisions of Experimental Therapy, Molecular Biology, Surgical Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

36. Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

37. Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

38. Department of Pathology, Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

39. Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA

40. Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030 USA

41. Department of Biomolecular Engineering and Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, CA 95064 USA

42. Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, CA 95064 USA

43. Buck Institute for Age Research, Novato, CA 94945 USA

44. Division of Hematology/Oncology, University of California - San Francisco, San Francisco, CA 94143 USA

45. Oregon Health and Science University, Department of Molecular and Medical Genetics, Portland OR 97239 USA

46. International Genomics Consortium, Phoenix, AZ 85004 USA

47. Nationwide Children's Hospital Biospecimen Core Resource, The Research Institute at Nationwide Children's Hospital, Columbus, OH 43205 USA

48. The Ohio State University College of Medicine, Department of Pathology, Columbus, OH 43205 USA

49. The Ohio State University College of Medicine, Department Pediatrics, Columbus, OH 43205 USA

50. Department of Surgery, Brigham and Women’s Hospital, Harvard Medical School, Brookline, MA 02115 USA

51. Department of Pathology, Christiana Care Health Services, Newark DE 19718, USA

52. Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

53. Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Brookline, MA 02115 USA

54. Department of Surgery, Helen F. Graham Cancer Center at Christiana Care, Newark, DE 19718 USA

55. Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15213 USA

56. Klinik **für** Chirurgie, Krachenhaus Alten Eichen, Hamburg, Germany

57. Department of Medical Oncology, Dana-Farber Cancer Institute, Brookline, MA 02115 USA

58. Department of Medicine, Brigham and Women's Hospital, Brookline, MA 02115 USA

59. Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

60. Indivumed Inc., Kensington, MD 20895 USA

61. ILSbio, LLC, Chestertown, MD 21620 USA

62. Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

63. Department of Surgery, Brigham and Women's Hospital, Brookline, MA 02115 USA

64. Tissue and Blood Repository, Brigham and Women's Hospital, Brookline, MA 02115 USA

65. Dept of Radiation Oncology, University of North Carolina School of Medicine. Chapel Hill, NC 27599

66. Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY 10065 USA

67. Department of Pathology, Case Medical Center, Cleveland, OH 44106 USA

68. Chirugische Klinik, Israelitisches Krankenhaus, Hamburg, Germany

69. SRA International, Fairfax, VA 22033 USA

70. The Cancer Genome Atlas Program Office, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892 USA

71. Center for Biomedical Informatics and Information Technology (CBIIT), National Cancer Institute, National Institutes of Health, Rockville, MD 20852 USA

72. Scimentis, LLC, Statham, GA 30666 USA

73. MLF Consulting, Arlington, MA 02474 USA

74. National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892 USA

References

1. TCGA. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008; 455:1061–1068. [PubMed: 18772890]
2. TCGA. Integrated genomic analyses of ovarian carcinoma. Nature. 2011; 474:609–615. [PubMed: 21720365]

3. Fearon ER. Molecular genetics of colorectal cancer. Annual review of pathology. 2011; 6:479–507.
4. Bass AJ, et al. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. Nature genetics. 2011
5. Sjoblom T, et al. The consensus coding sequences of human breast and colorectal cancers. Science (New York, N.Y. 2006; 314:268–274.
6. Wood LD, et al. The genomic landscapes of human breast and colorectal cancers. Science (New York, N.Y. 2007; 318:1108–1113.
7. Umar A, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. Journal of the National Cancer Institute. 2004; 96:261–268. [PubMed: 14970275]
8. Aaltonen LA, et al. Clues to the pathogenesis of familial colorectal cancer. Science (New York, N.Y. 1993; 260:812–816.
9. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature. 1993; 363:558–561. [PubMed: 8505985]
10. Parsons R, et al. Hypermutability and mismatch repair deficiency in RER+ tumor cells. Cell. 1993; 75:1227–1236. [PubMed: 8261516]
11. Dooley AL, et al. Nuclear factor I/B is an oncogene in small cell lung cancer. Genes & development. 25:1470–1475. [PubMed: 21764851]
12. Major MB, et al. Wilms tumor suppressor WTX negatively regulates WNT/beta-catenin signaling. Science. 2007; 316:1043–1046. [PubMed: 17510365]
13. Mori-Akiyama Y, et al. SOX9 is required for the differentiation of paneth cells in the intestinal epithelium. Gastroenterology. 2007; 133:539–546. [PubMed: 17681175]
14. Bastide P, et al. Sox9 regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. The Journal of cell biology. 2007; 178:635–648. [PubMed: 17698607]
15. Jones S, et al. Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. Human mutation. 2011
16. Wilson BG, Roberts CW. SWI/SNF nucleosome remodellers and cancer. Nature reviews. 2011; 11:481–492.
17. Minsky BD. Unique considerations in the patient with rectal cancer. Seminars in oncology. 2011; 38:542–551. [PubMed: 21810513]
18. Hinoue T, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. Genome research. 2012; 22:271–282. [PubMed: 21659424]
19. Beroukhim R, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proceedings of the National Academy of Sciences of the United States of America. 2007; 104:20007–20012. [PubMed: 18077431]
20. Camps J, et al. Integrative genomics reveals mechanisms of copy number alterations responsible for transcriptional deregulation in colorectal cancer. Genes, chromosomes & cancer. 2009; 48:1002–1017. [PubMed: 19691111]
21. Varley JM, Swallow JE, Brammar WJ, Whittaker JL, Walker RA. Alterations to either c-erbB-2(neu) or c-myc proto-oncogenes in breast carcinomas correlate with poor short-term prognosis. Oncogene. 1987; 1:423–430. [PubMed: 3330785]
22. Yokota J, et al. Amplification of c-erbB-2 oncogene in human adenocarcinomas in vivo. Lancet. 1986; 1:765–767. [PubMed: 2870269]
23. van der Flier LG, et al. Transcription factor achaete scute-like 2 controls intestinal stem cell fate. Cell. 2009; 136:903–912. [PubMed: 19269367]
24. Jubb AM, Hoeflich KP, Haverty PM, Wang J, Koeppen H. Ascl2 and 11p15.5 amplification in colorectal cancer. Gut. 2010
25. Stange DE, et al. Expression of an ASCL2 related stem cell signature and IGF2 in colorectal cancer liver metastases with 11p15.5 gain. Gut. 2010; 59:1236–1244. [PubMed: 20479215]
26. Cui H, et al. Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. Science (New York, N.Y. 2003; 299:1753–1755.

27. Nakagawa H, et al. Loss of imprinting of the insulin-like growth factor II gene occurs by biallelic methylation in a core region of H19-associated CTCF-binding sites in colorectal cancer. Proceedings of the National Academy of Sciences of the United States of America. 2001; 98:591–596. [PubMed: 11120891]
28. Veronese A, et al. Oncogenic role of miR-483-3p at the IGF2/483 locus. Cancer research. 2010; 70:3140–3149. [PubMed: 20388800]
29. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. Genome research. 2011; 22:398–406. [PubMed: 21908773]
30. Brady CA, et al. Distinct p53 transcriptional programs dictate acute DNA-damage responses and tumor suppression. Cell. 2011; 145:571–583. [PubMed: 21565614]
31. Rivera MN, et al. An X chromosome gene, WTX, is commonly inactivated in Wilms tumor. Science (New York, N.Y. 2007; 315:642–645.
32. Scheel SK, et al. Mutations in the WTX-gene are found in some high-grade microsatellite instable (MSI-H) colorectal cancers. BMC cancer. 2010; 10:413. [PubMed: 20696052]
33. Forbes SA, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). **Chapter 10**, Unit 10 11. Current protocols in human genetics / editorial board, Jonathan L. Haines … [et al. 2008
34. Massague J, Blain SW, Lo RS. TGFbeta signaling in growth control, cancer, and heritable disorders. Cell. 2000; 103:295–309. [PubMed: 11057902]
35. Vaske CJ, et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. Bioinformatics (Oxford, England). 2010; 26:i237–i245.
36. House CD, et al. Voltage-gated Na+ channel SCN5A is a key regulator of a gene transcriptional network that controls colon cancer invasion. Cancer research. 2010; 70:6957–6967. [PubMed: 20651255]
37. Liu Z, Lu H, Jiang Z, Pastuszyn A, Hu CA. Apolipoprotein l6, a novel proapoptotic Bcl-2 homology 3-only protein, induces mitochondria-mediated apoptosis in cancer cells. Molecular cancer research : MCR. 2005; 3:21–31. [PubMed: 15671246]
38. Topol L, Chen W, Song H, Day TF, Yang Y. Sox9 inhibits Wnt signaling by promoting beta-catenin phosphorylation in the nucleus. The Journal of biological chemistry. 2009; 284:3323–3333. [PubMed: 19047045]
39. Nagl NG Jr, Zweitzig DR, Thimmapaya B, Beck GR Jr, Moran E. The c-myc gene is a direct target of mammalian SWI/SNF-related complexes during differentiation-associated cell cycle arrest. Cancer research. 2006; 66:1289–1293. [PubMed: 16452181]
40. Chen B, et al. Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. Nature chemical biology. 2009; 5:100–107.
41. Ewan K, et al. A useful approach to identify novel small-molecule inhibitors of Wnt-dependent transcription. Cancer research. 2010; 70:5963–5973. [PubMed: 20610623]
42. Sack U, et al. S100A4-induced cell motility and metastasis is restricted by the Wnt/{beta}-catenin pathway inhibitor calcimycin in colon cancer cells. Molecular biology of the cell. 2011; 22:3344–3354. [PubMed: 21795396]
43. Chen K, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nature methods. 2009; 6:677–681. [PubMed: 19668202]
44. Xi R, et al. Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108:E1128–E1136. [PubMed: 22065754]

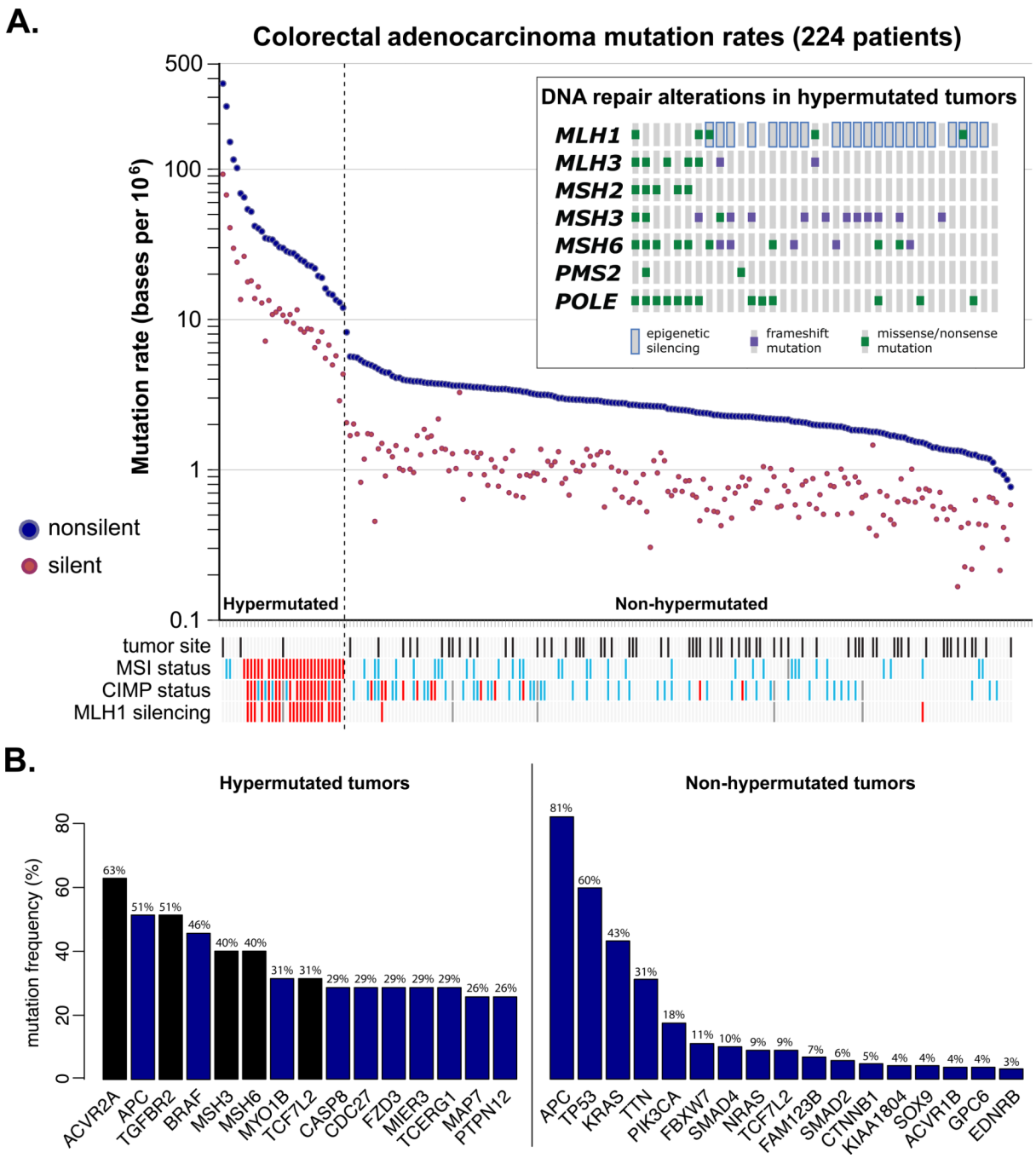

Figure 1. Mutation frequencies in human CRC

A. Mutation frequencies in each of the tumors. Note a clear separation of hypermutated and non-hypermutated samples. **Inset:** Mutations in mismatch repair genes and *POLE* among the hypermutated samples. The order of the samples is the same as in Figure 1A. **B.** Significantly mutated genes in non-hypermutated and hypermutated tumors. Blue bars represent genes identified by MutSig and genes in black bars are identified by manual examination of sequence data.

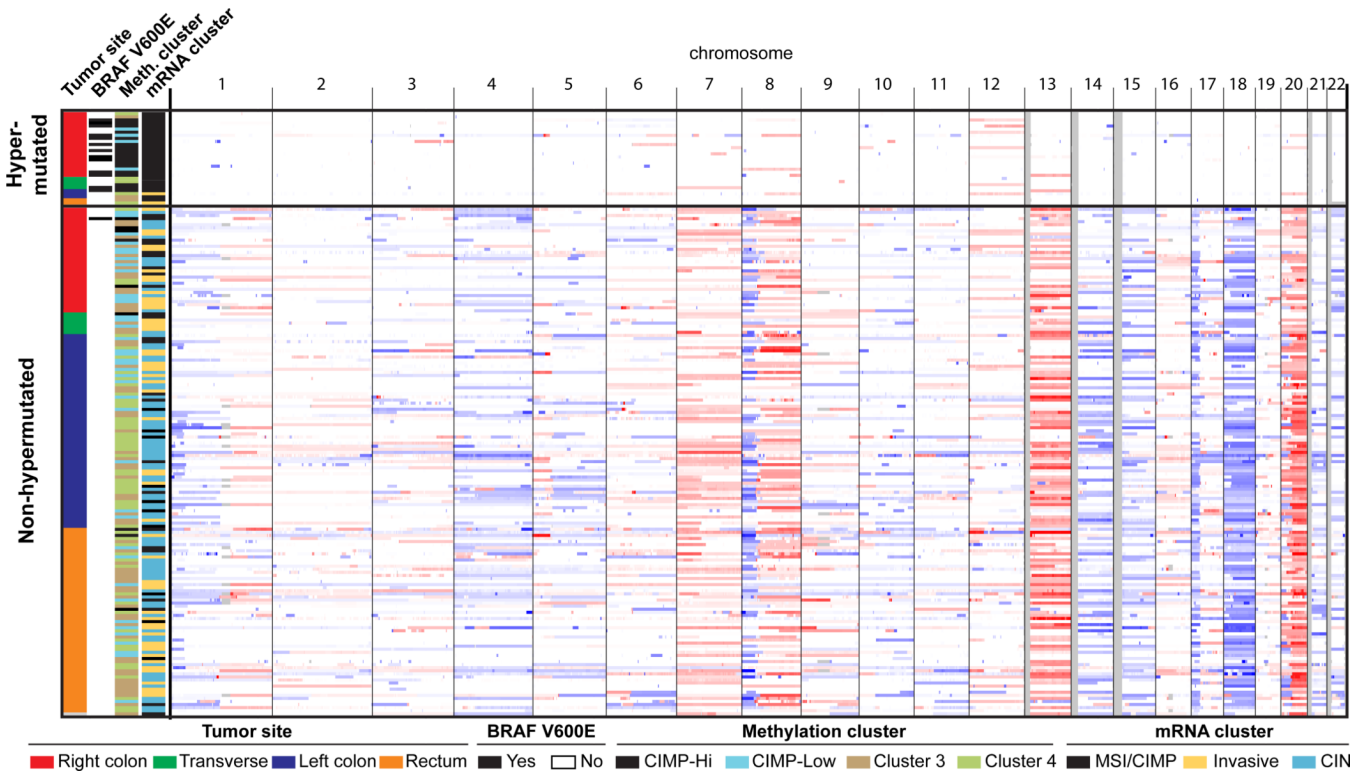

Figure 2. Integrative analysis of genomic changes in 195 CRC tumors

Hypermutated tumors have near diploid genomes and are highly enriched for hypermethylation, CIMP expression phenotype, and *BRAF* V600E mutations. Non-hypermutated tumors originating from different sites are virtually indistinguishable from each other based on their copy-number alteration patterns, DNA methylation, or gene expression patterns. Copy-number changes of the 22 autosomes are shown in shades of red for copy-number gains and shades of blue for copy-number losses.

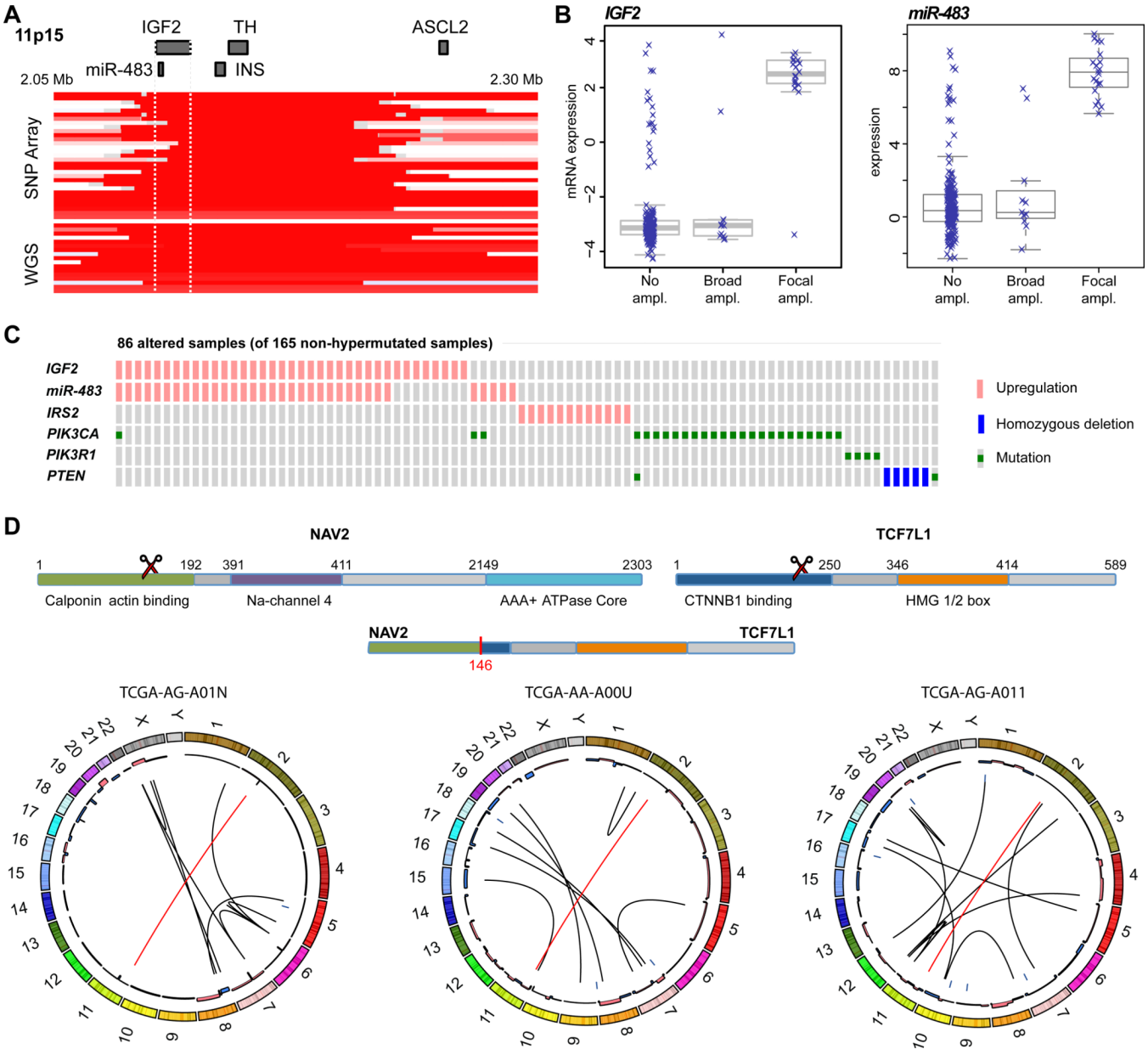

Figure 3. Copy number changes and structural aberrations in CRC

A. Focal amplification of 11p15.5. Segmented DNA copy-number data from SNP arrays and low pass whole genome sequencing are shown. Each row represents a patient; amplified regions are shown in red. **B.** Correlation of expression levels with copy number changes for IGF2 and miR-483. **C.** IGF2 amplification and over-expression are mutually exclusive of alterations in PI3K signaling genes. **D.** Recurrent NAV2-TCF7L2 fusions. The structure of the two genes, locations of the breakpoints leading to the translocation and circular representations of all rearrangements in tumors with a fusion are shown. The red line lines represent the NAV2-TCF7L2 fusions, black lines indicate other rearrangements. The inner ring represents copy-number changes (blue = loss, pink = gain).

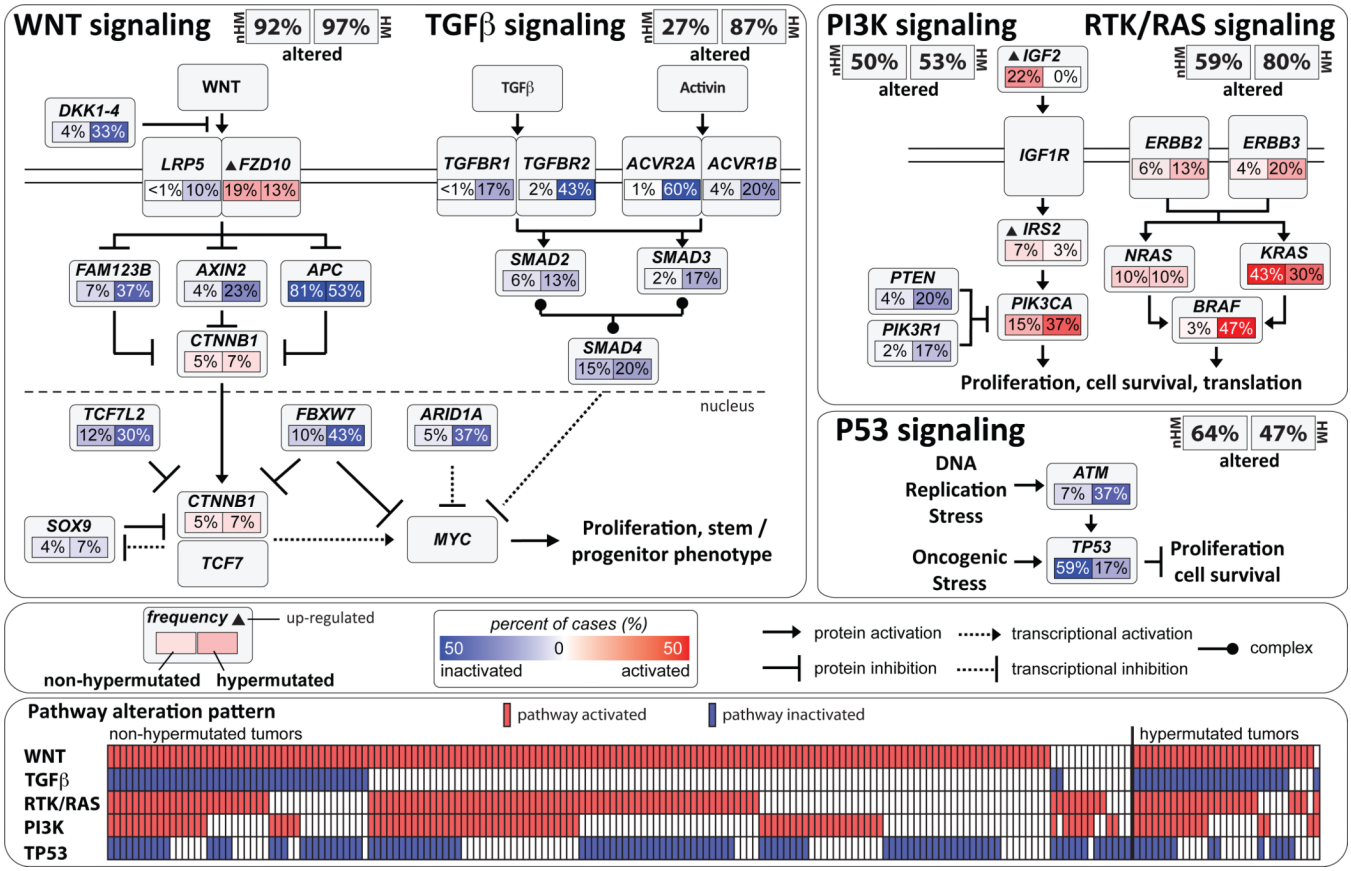

Figure 4. Diversity and frequency of genetic changes leading to deregulation of signaling pathways in CRC

Non-hypermuated (n = 165) and hypermutated (n = 30) samples with complete data were analyzed separately. Alterations are defined by somatic mutations, homozygous deletions, high-level, focal amplifications, and, in some cases, by significant up- or down-regulation of gene expression (*IGF2, FZD10, SMAD4*). Alteration frequencies are expressed as a percentage of all cases; activated genes are red and inactivated genes are blue. The bottom panel shows for each sample if at least one gene in each of the five pathways is altered.

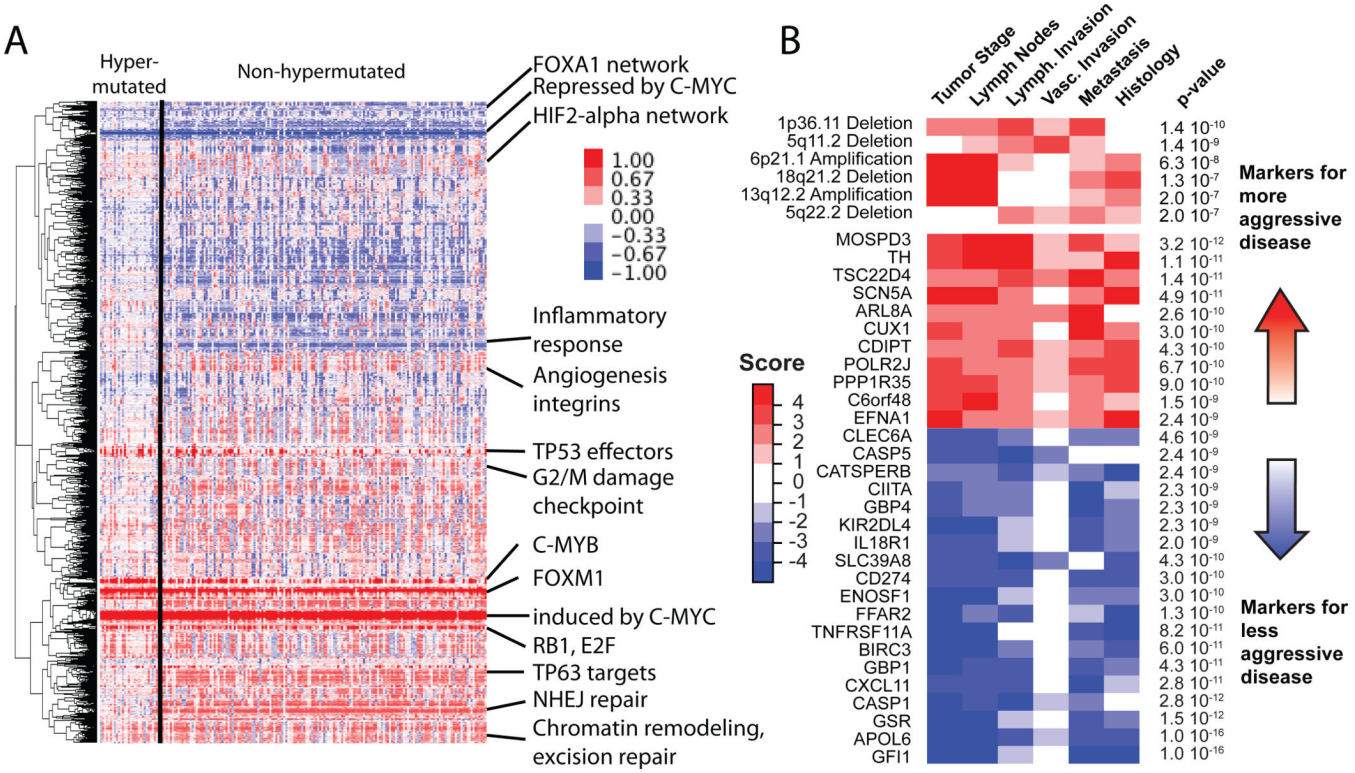

Figure 5. Integrative analyses of multiple data sets

A. Clustering of genes and pathways affected in colon and rectum tumors deduced by PARADIGM analysis. Blue = under-expressed relative to normal and red = overexpressed relative to normal. Some of the pathways deduced by this method are shown on the right. **B.** Gene expression signatures and SCNAs associated with tumor aggression. Molecular signatures (rows) that show statistically significant association with tumor aggressiveness according to selected clinical assays (columns) are displayed in color, with red indicating markers of tumor aggressiveness, and blue the markers of less aggressive tumors. Significance is based on the combined *p*-value from the weighted Fisher's method, corrected for multiple testing. Color intensity and score is in accordance with the strength of an individual clinical-molecular association, and is proportional to $log_{10}(p)$, where p is p-value for that association. To limit the vertical extent of the figure, gene expression signatures are restricted to combined p-value $p<10^{-9}$, SCNAs to $p<10^{-7}$ and features are shown only if they are also significant in the subset of non-MSI-H samples (the analysis was performed separately on the full data as well as on the MSI-H and non-MSI-H subgroups).