

# Immunomic, genomic and transcriptomic characterization of CT26 colorectal carcinoma

Castle *et al.*

RESEARCH ARTICLE

Open Access

# Immunomic, genomic and transcriptomic characterization of CT26 colorectal carcinoma

John C Castle<sup>1\*</sup>, Martin Loewer<sup>1</sup>, Sebastian Boegel<sup>1,2</sup>, Jos de Graaf<sup>1</sup>, Christian Bender<sup>1</sup>, Arbel D Tadmor<sup>1</sup>, Valesca Boisguerin<sup>1,3</sup>, Thomas Bukur<sup>1</sup>, Patrick Sorn<sup>1</sup>, Claudia Paret<sup>1</sup>, Mustafa Diken<sup>1</sup>, Sebastian Kreiter<sup>1</sup>, Özlem Türeci<sup>2</sup> and Ugur Sahin<sup>1,2,3\*</sup>

## Abstract

**Background:** Tumor models are critical for our understanding of cancer and the development of cancer therapeutics. Here, we present an integrated map of the genome, transcriptome and immunome of an epithelial mouse tumor, the CT26 colon carcinoma cell line.

**Results:** We found that *Kras* is homozygously mutated at p.G12D, *Apc* and *Tp53* are not mutated, and *Cdkn2a* is homozygously deleted. Proliferation and stem-cell markers, including *Top2a*, *Birc5* (*Survivin*), *Cldn6* and *Mki67*, are highly expressed while differentiation and top-crypt markers *Muc2*, *Ms4a8a* (*MS4A8B*) and *Epcam* are not. *Myc*, *Trp53* (*tp53*), *Mdm2*, *Hif1a*, and *Nras* are highly expressed while *Egfr* and *Flt1* are not. MHC class I but not MHC class II is expressed. Several known cancer-testis antigens are expressed, including *Atad2*, *Cep55*, and *Pbk*. The highest expressed gene is a mutated form of the mouse tumor antigen gp70. Of the 1,688 non-synonymous point variations, 154 are both in expressed genes and in peptides predicted to bind MHC and thus potential targets for immunotherapy development. Based on its molecular signature, we predicted that CT26 is refractory to anti-EGFR mAbs and sensitive to MEK and MET inhibitors, as have been previously reported.

**Conclusions:** CT26 cells share molecular features with aggressive, undifferentiated, refractory human colorectal carcinoma cells. As CT26 is one of the most extensively used syngeneic mouse tumor models, our data provide a map for the rationale design of mode-of-action studies for pre-clinical evaluation of targeted- and immunotherapies.

**Keywords:** Immunotherapy, Cancer models, Computational immunology, Colorectal cancer

## Background

Murine CT26 (Colon Tumor #26) cells were developed in 1975 by exposing BALB/c mice to N-nitroso-N-methylurethane (NMU), resulting in a rapid-growing grade IV carcinoma that is easily implanted and readily metastasizes [1]. Used in over 500 published studies, the CT26 colon carcinoma is one of the most commonly used cell lines in drug development. Numerous cytotoxic agents as well as therapeutics targeting specific signaling pathways have been studied with these cells [2-4]. Moreover, as the CT26 model in BALB/c mice provides a syngeneic in vivo test system, it is frequently used for developing and testing immunotherapeutic concepts.

In sharp contrast to its frequent use in drug development, there have been no comprehensive studies of the genome and transcriptome of CT26. *Kras* is mutated in CT26 [5] but other mutations are not known. Mutations in *Cdkn2a*, *Mek*, *Braf* and *Pi3k* in combination with *Egfr* and *Vegf* expression, for instance, may influence the results of pre-clinical investigations of treatment modalities. Moreover, while gp70, the product of the envelope gene of murine leukemia virus (MuLV)-related cell surface antigen, is a known model antigen for studying antigen-specific immune responses in the CT26 system, there is no comprehensive knowledge of potential tumor antigens in this cell line system.

Further, the lack of comprehensive data on the murine CT26 colon cancer data sharply contrasts to the extensive molecular characterization of human colorectal cancer (CRC). As a group, human CRC is highly heterogeneous with multiple evolutionary paths, with molecular

\* Correspondence: john.castle@tron-mainz.de; sahin@uni-mainz.de  
<sup>1</sup>TRON gGmbH - Translational Oncology, Johannes Gutenberg-University Medical Center gGmbH, Langenbeckstr. 1, Building 708, 55131 Mainz, Germany  
<sup>2</sup>University Medical Center of the Johannes Gutenberg-University Mainz, 55131 Mainz, Germany  
Full list of author information is available at the end of the article

signatures classifying subtypes and steps from adenoma to carcinoma. Many human CRC genomes are now known and multiple molecular signatures, classifications and biomarker concepts are published [6-9]. As comprehensive genomic and transcriptomic data of CT26 has not been available, it is unclear how CT26, a chemically-induced tumor, molecularly correlates to human CRC subtypes and to what extent it may be used as model.

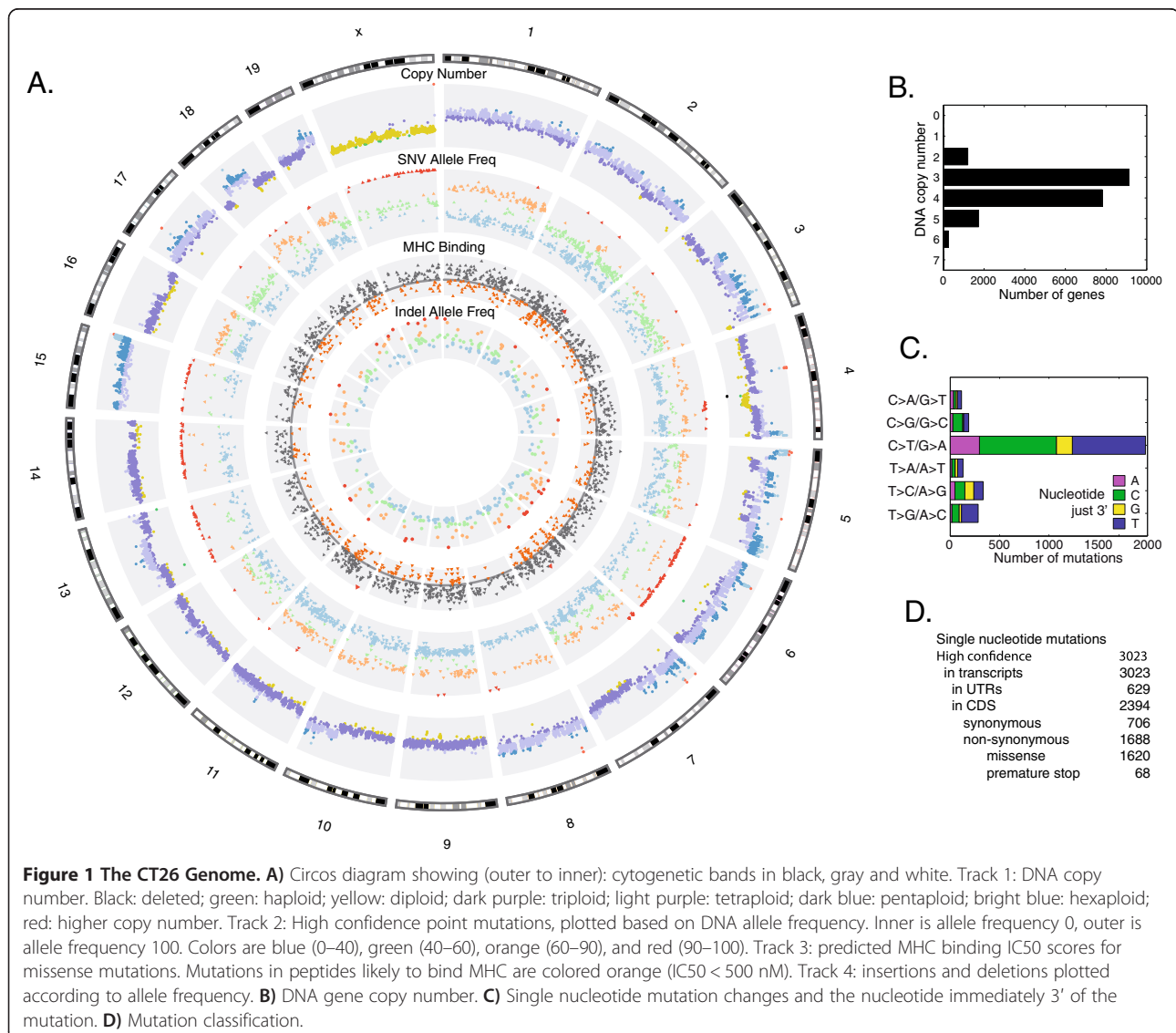
To answer these questions, we utilized next-generation sequencing, bioinformatics and immuno-informatics to create an integrated mouse solid tumor mutanome, transcriptome and immunome, providing an overdue analysis of the CT26 cancer cell line.

## Results and discussion

*The CT26 tumor genome:* using the NGS reads, we assessed copy number and nucleotide variations by

comparing CT26 to BALB/cJ DNA. We determined absolute DNA copy number using the ratio of exome-seq reads mapping to each gene from CT26 versus those from BALB/cJ, and integrating variant allele fraction (Figure 1A, outer ring). We found that the ploidy of CT26 is strikingly large with large regions of triploidy and tetraploidy, in agreement with previous karyotyping results [10]. The median and mean copy number in average across all genes is 3 and 3.5, respectively, with 8,686 genes in triploid regions (45% of the genes) and 7,448 (39%) in tetraploid regions (Figure 1B). No reads map to the Y chromosome (DNA or RNA), suggesting that CT26 cells originated from a female mouse. Only one homozygous deletion was found, which contains the tumor suppressor *Cdkn2a* (cyclin-dependent kinase inhibitor 2A; *Ink4a*) locus on mouse chromosome 4.

We identified 3,023 high-confidence single nucleotide variations (SNVs; Figure 1A, 2nd ring) and 362 short



insertions and deletions (indels; Figure 1A, inner ring). Indels are dominated by A/T deletions (44%). We selected high confidence SNVs in exons (3,023; Figure 1D), the majority of which are localized in coding regions (2,394; 79%). Of the SNVs in coding regions, the majority (1,688; 71%) cause non-synonymous protein changes, including 1,620 missense and 68 nonsense variants. The CCDS database identifies 32 million protein-encoding nucleotides in the mouse genome. Relative to a 2011 BALB/cJ genome, the CT26 variation rate in coding regions is 53 non-synonymous and 22 silent mutations per Mb. This is significantly more than the average found in spontaneous human tumors (4 mutations per Mb) but still within the range observed for primary human CRC tumors, which ranges from less than 1 per Mb to over 100 mutations per Mb [11].

The identified SNVs represent variations between the CT26 genome, derived from a BALB/c mouse in 1975, and a BALB/cJ mouse in 2011. As such, the SNVs include both somatic mutations associated with the CT26 onco-transformation and genetic drift in the BALB/c genome. We found 40,000 mouse SNPs that distinguish the BALB/cJ and mm9 (C57BL/6) exomes. Of these, only 1.6% show a discrepancy between the CT26 and 2011 BALB/cJ genomes. Thus, while this does not eliminate genetic drift or conclusively identify the substrain that gave rise to CT26 cells, it demonstrates that the genome of the mouse that originally created the CT26 cells is similar to that of the current BALB/cJ mouse.

Spontaneous human CRC tumors contain primarily C > T/G > A SNVs [12]. Of the 3,023 SNVs in the CT26 genome, 2,313 (77%) are transitions, of which most (1,980, 66%) are C > T/G > A mutations (Figure 1C), similar to the human CRC mutation profile. Based on data from over 7,000 human tumors, G is the dominant nucleotide immediately 3' of the mutated nucleotide in human CRC tumors (CG > TG mutations) [11]. Conversely, we found that CT26 SNVs are depleted in CG > TG and CA > TA mutations and enriched in CT > TT and CC > TC mutations. This pattern, a C > T mutation followed by a pyrimidine, is found in tumor samples from human patients pretreated with temozolomide, an alkylating anticancer drug [11]. CT26 was originally induced by the alkylating agent NMU. That temozolomide and NMU are both associated with tumors enriched in C > T mutations at positions followed by a pyrimidine suggest a similar mutagenic pattern for these two alkylating agents.

Of the 3,023 CT26 SNVs, 296 (10%) are homozygous or heterozygous (100% allele frequency, Figure 1A, 2nd ring), even in amplified regions with high copy number. Homozygous variants cluster across chromosomes 6, 13, 14, 15, and X. These regions could be the result of either a loss of heterozygosity (LOH) onco-transformation or genetic drift in a BALB/c mouse followed by inbreeding. If the result of

an onco-transformation, that the regions experienced LOH, followed by mutations and copy number amplification suggests that resulting individual alleles were amplified 2-fold (chr X), 3-fold (chr 14), 4-fold (chr 6), and 5-fold (chr 15).

We further investigated chromosome X. Mutations occur on chromosome X with 100% and 50% DNA allele frequency, suggesting that chromosome X is diploid in CT26 cells. Female cells typically express XIST and inactivate one X allele. In CT26, the RNA-Seq data show that XIST is not expressed and, examining the allele expression of heterozygous mutations, that transcription occurs from both chromosome X alleles. These findings are concordant with a scenario where the chromosome X experienced both a loss of the inactivated allele and an amplification of the non-inactivated allele (occurring in either order).

In summary, the data imply that the CT26 has a complex genome of high ploidy which underwent several amplification events. Relative to a 2011 BALB/c genome, the number of mutations is higher than average, with many non-synonymous mutations. The mutation pattern reflects the treatment with the NMU alkylating agent, a similar but distinct pattern than found in spontaneous primary CRC.

*CT26 SNVs in onco-relevant genes:* we investigated whether mutations associated with CRC [12-14] are also prevalent in CT26. APC, KRAS and TP53 are frequent drivers of the linear and uniform evolution of spontaneous human CRC; of these, only Kras is mutated in CT26. The CT26 Kras genomic locus is triploid and all alleles contain V8M (located in a small molecule binding site [15]) and G12D (known to stimulate proliferation) mutations.

Several CRC subtypes are linked to syndromes based on inherited gene defects and mutations. Genes associated with familial CRC (e.g., HNPCC, Lynch Syndrome, FAP, Peutz-Jeghers) include mismatch repair genes Mlh1, Mlh2, Mlh6, Msh2, Myh, Pms1, Stk1, Mutyh and Ctnnb1. None are mutated in CT26. The lack of mutations in mismatch repair genes Mlh1 and Msh2, which are associated with CRC microsatellite instability (MSI), agrees with the lack of mutation in Braf, which is frequently associated with the MSI-high phenotype [16].

Further, the tumor suppressor Cdkn2a is homozygously deleted and the genomic Mapk1 (MEK) and Met loci are amplified in CT26. CRC-associated genes Fbxw7, Pik2ca, Pten, Smad2, Smad4, Tcf7l2 are not mutated. Non-synonymous point mutations occur in other CRC genes Brca2 (R2066K), Pdgfra (V103I), Nav3 (V154I, S334N), Atr (H792Q), Cdk8 (S87F), and Rel (A406T). Mutations in cancer-related genes include mTor (V971M), Birc2 (E395K), Casp4 (H84Y), Cenpe (A834V), Esr1 (P508S), Hdac2 (P228S), Ins1 (Y40C), Insr (A493V), Muc1

(L555F), Pik3c3 (S282A), Pik3cg (D120N), Fgfr1 (S107F), Ddr2 (A161V), Notch1 (R365S) and Rhoj (L137F). Frameshift-causing indels occur in oncogenes Ewsr1 (at amino acid 629) and Mpp3 (at amino acid 91).

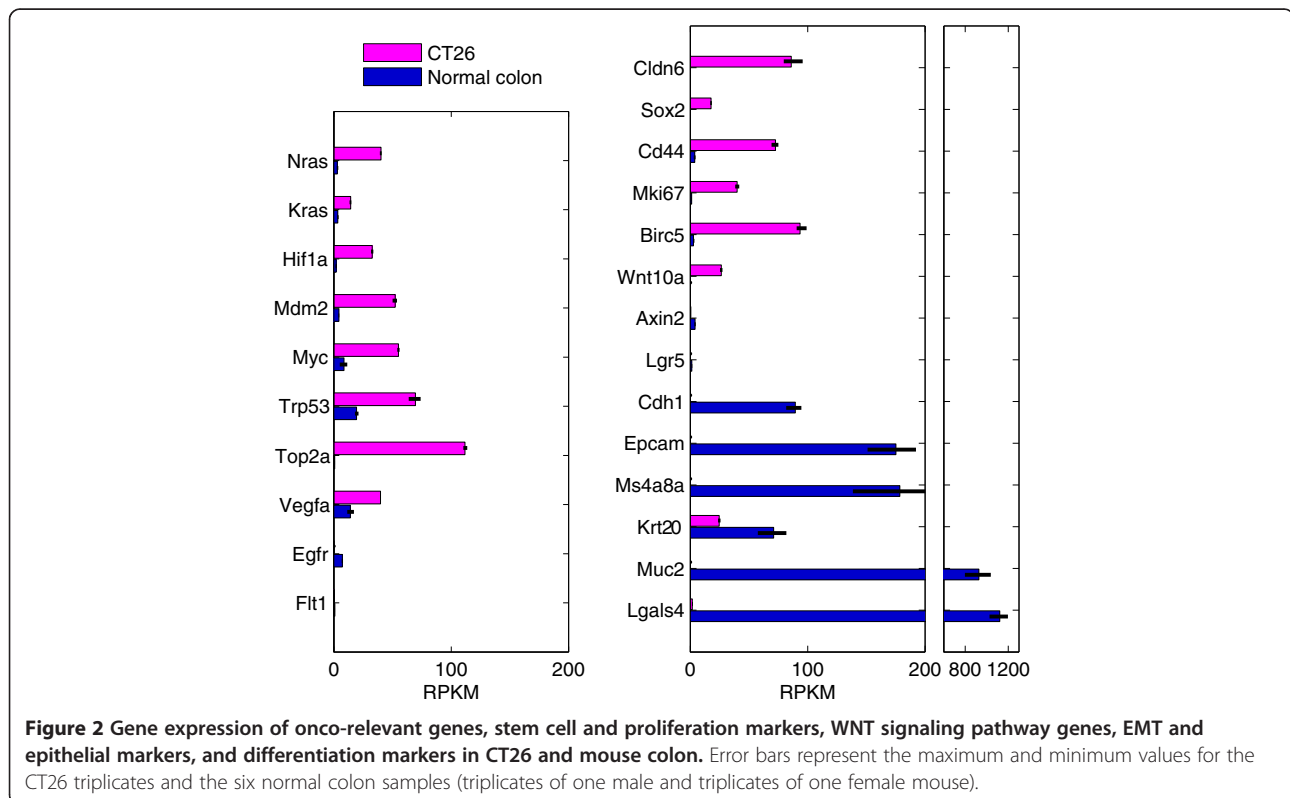
**CT26 gene expression:** we generated gene expression profiles from CT26 cells. Cancer-relevant genes such Nras, Vegfa, Trp53 (TP53), Myc, Mdm2, and Hif1a are expressed at high levels in CT26 (Figure 2, left). Egfr and Flt1 are not expressed. Gene expression in CT26 relative to normal colon was used for pathway enrichment analysis in order to identify broadly enriched pathways (Figure 3). Not surprisingly, the identified pathways relate to cell proliferation (cell cycle phases and transitions, DNA replication) and increased translation (protein and RNA metabolism). We examined individual gene sets enriched in CT26 (Figure 4). Most enriched is “CELL\_CYCLE\_RB1\_TARGETS”, a gene set curated from a study examining RB1 target genes involved in cell cycle regulation [17], reflecting over-expression of all Rb1 target genes (Figure 4B). Rb1 mRNA is itself 8-fold up-regulated. Ezh2, downstream of the Egfr-ras-raf pathway, impacts DNA methylation, promotes EMT and is associated with poor prognosis in CRC [18,19]. Together with its target genes, Ezh2 is over-expressed in CT26 cells. Mechanistically, that Rb1, Ezh2, Lin9, and E2f mRNAs and their target genes are over-expressed suggests that the Rb1, Ezh2, Lin9, and E2f mRNA levels, in

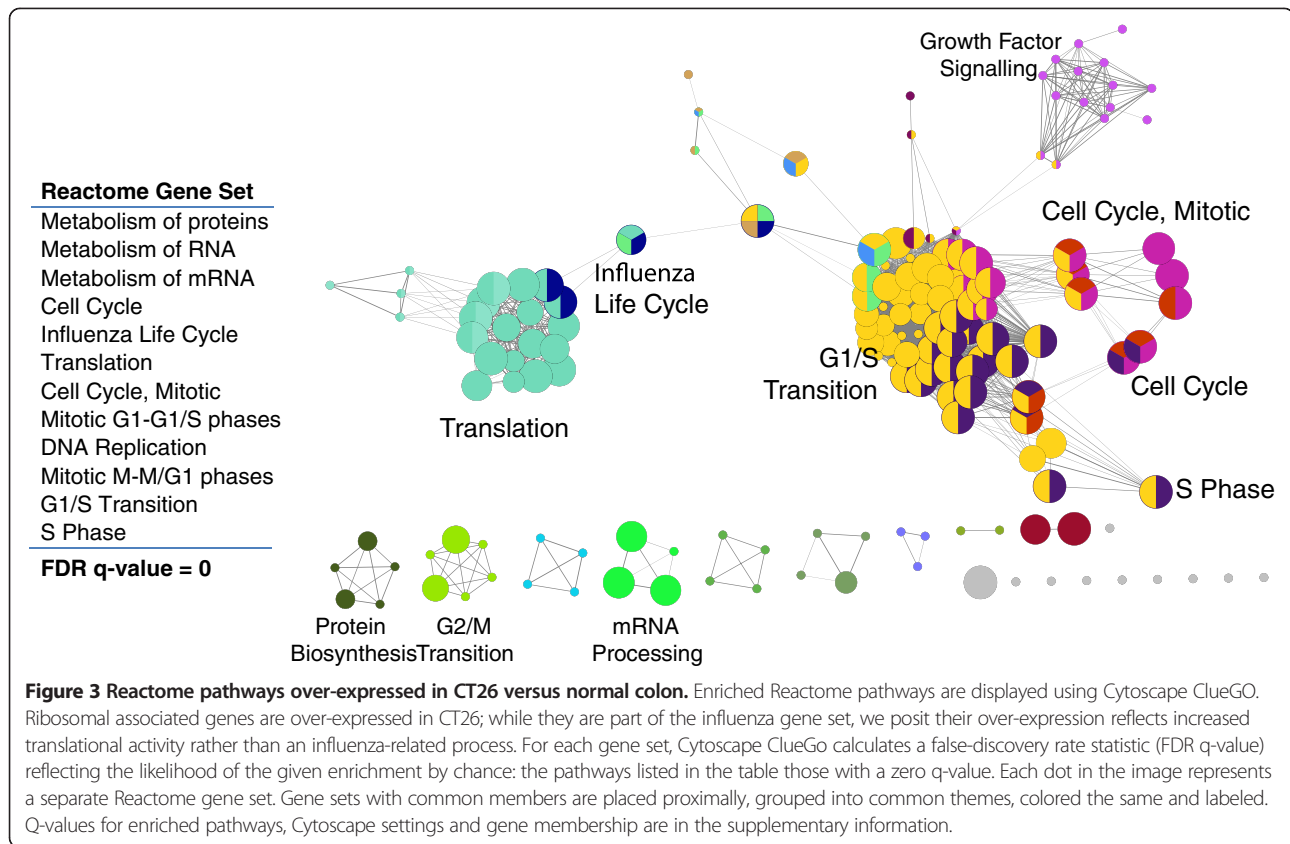
addition to post-translational modifications, play a critical role controlling activation of each pathway.

The gene set associated with genes down-regulated after Foxo3 up-regulation was found to be up-regulated (Figure 4E). In agreement with this, Foxo3 is significantly down-regulated in CT26 cells. Foxo3 expression has been identified as a potential biomarker for CRC outcome [20], with low Foxo3 associated with 2-fold shorter survival. The low Foxo3 expression, the high Ezh2 expression and the enrichment of the “melanoma metastasis” gene set [21] are all in line with the aggressive and high metastatic activity of CT26 cells.

Differentiation markers further corroborate that CT26 cells are in a highly proliferative, undifferentiated state. The “undifferentiated cancer” gene set is highly up-regulated in the CT26 cells (Figure 4A). Stem cell markers Cldn6 and Sox2 are highly expressed while differentiation markers Muc2 and Ms4a8a (human MS4A8B) [22] markers are not expressed (Figure 2, right). Whereas Lgals1 (Galectin-1) is over 30-fold up-regulated in CT26 cells, the orthologous gene Lgals4 (Galectin-4), a differentiation marker, is over 500-fold down-regulated in CT26 cells. The proliferation markers Top2a (DNA topoisomerase 2-alpha), Mki67 and Birc5 (Survivin) are all highly expressed in CT26 cells.

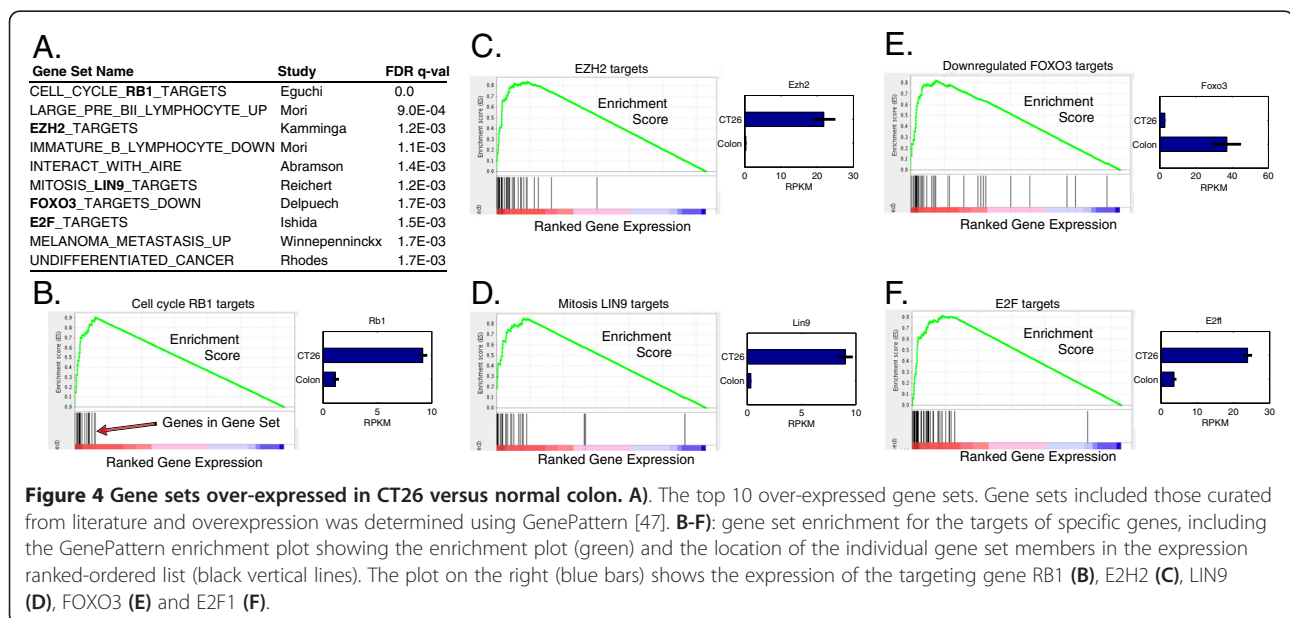
Epcam marks epithelial cells and colon crypt tops [23] and is not expressed in CT26 cells. Cdh1 (e-cadherin)





marks the epithelial-mesenchymal transition [24] and is highly expressed in normal colon but not expressed in CT26. CD44 marks the crypt bottoms and is 18-fold up-regulated. Silencing of WNT targets such as ASCL2, AXIN2 and LGR5 is often accomplished through CpG

promoter methylation and associated with poor prognosis and increase metastatic spread [25]. In CT26, Wnt10a is highly up-regulated but WNT target genes, with the exception of Birc5, are not expressed. These markers classify CT26 as cells that originated in the



lower-crypt and are in an undifferentiated state prone to metastasize [26,27].

CRC cohort studies have identified markers for classifying patient CRC tumors (Additional file 1: Table S8). The three-group CRC classification platform using differentiation marker KRT20 and “top crypt” markers CA, MS4 A12 and CD177 [6] classifies CT26 as a tumor with a less mature phenotype and worse progression. The classification platform using genes FRMD6, ZEB1, HTR2B and CDX2 [7] classifies CT26 as the “CCS3” sub-type, with poor prognosis, low therapy response and resistance to cetuximab. The 7 gene “CRCAssigner-7” platform [8] classifies CT26 cells as either “stem like” or “CR-TA” (cetuximab-resistant transit-amplifying).

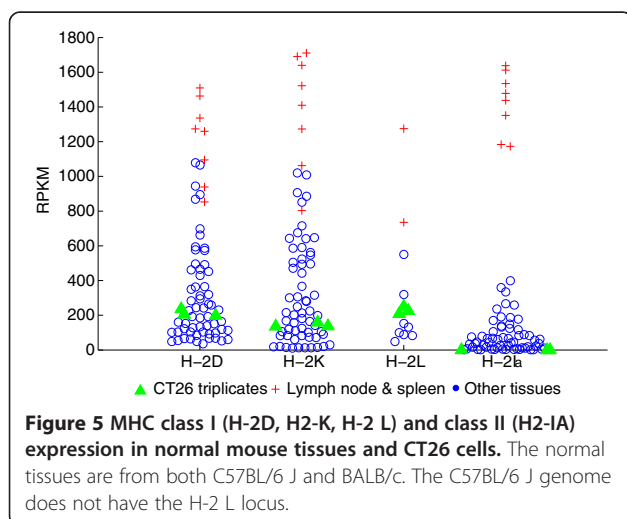
*The CT26 cancer immunome:* immunotherapy concepts include targeting tumor-specific antigens presented on MHC molecules. We determined that CT26 cells have the same MHC types as the parental BALB/c mice: H-2D<sup>d</sup>, H-2K<sup>d</sup> and H-2L<sup>d</sup> (class I) and H-2Ia<sup>d</sup> (class II). This is expected and a useful confirmation of the BALB/c-CT26 lineage, given on-going reports of cell line mis-identifications. Class I loci H-2D<sup>d</sup> and H-2K<sup>d</sup> are expressed at levels comparable to normal tissues (Figure 5), lower than lymph node and spleen but higher than non-immune tissues (e.g., heart, kidney, brain). B2m, part of the MHC class I complex, is highly expressed. Both suggest that MHC class I is functional. Normal tissues show variable expression of MHC class II (e.g., lymph node and spleen are high, colon expresses at 150 RPKM and brain is low but non-zero). CT26 cells express neither MHC class II (0 RPKM) nor the MHC class II transactivator Ciita, suggesting that CT26 cells do not have functional MHC class II antigen presentation.

Genes with tumor-associated expression as well as genes with somatic mutations may act as tumor-associated antigens (TAAs) (Table 1). Gp70 (an endogenous envelope

protein of from a MuLV-related retrovirus) is a classical model tumor antigen frequently exploited when using CT26 system to investigate CD8 T cell immunity [28]. Expression of gp70 in normal mouse tissues has been observed in mice over 8 months old [29,30]; however, gp70 levels are strikingly high levels in murine tumor cell lines including CT26 [31]. Indeed, our data show that gp70 has the highest expression of all CT26 genes. While gp70 DNA was not captured by the NGS exome-capture, we were able to determine the gp70 sequence using the RNA-Seq reads, averaging over 5,000x coverage due to the high expression (Table 1, Additional file 1). Relative to the gp70 sequence in the mm9 genome, the CT26 gp70 sequence falls in a CT26 tetraploid region and has 9 non-synonymous mutations, including 3 homozygous and 6 heterozygous variants. Two variants are in dbSNP while three are found in Genbank mRNAs from other mouse tumor cell lines, suggesting that four could be unique to CT26 cells. Three variants introduce stop codons; however all are heterozygous such that a full length gp70 can likely be translated.

The family of cancer testes (CT) antigens has high tumor cell selectivity. We found that CT antigens with the highest expression in CT26 cells are known colorectal CT antigens Casc5, Cep55 and Pbk (Table 2). These three, along with Atad2 and Ttk, have very low expression in the normal colon samples. Low expression of the human homologs of Casc5, Ctage5, Pbk and Spag9 has been observed in multiple tissues, such that these are cancer testes-selective antigens and they may be subject to tolerance [32]. Conversely, while expressed at 5-fold higher levels in CT26 cells, Rqcd1is also expressed at significant levels in normal colon and is thus not an ideal immunotherapy target.

In addition to tissue specific and over-expressed tumor antigens, somatic mutations provide tumor-specific immunotherapy T-cell targets [33] that may be used for truly individualized cancer therapeutics and vaccines [34]. A mutation for a cancer vaccine target must be expressed and presented on MHC molecules. Of the 3,023 CT26 point mutations, 1,172 are in expressed genes and, of these, 154 are in epitopes predicted to strongly bind to MHC molecules (highest 1% consensus percentile) (Figure 1, 3rd ring). 73 occur in highly expressed genes (at least 10 RPKM). Table 3 shows eight such point mutations that meet these criteria. For each SNV, Additional file 1: Table S2 lists the mutation-containing epitope and MHC allele predicted to have the strongest MHC binding by the IEDB algorithm [35]. Previous work by us and others [36] finds that roughly 30% of these mutations are antigenic and capable of generating a T cell response when used in immunizations. Thus, these mutations provide a broad portfolio of potentially exploitable TSAs for future studies.



**Table 1 The CT26 gp70 SNVs relative to the mm9 reference genome (negative strand)**

| Mm9 coordinate | Reference | Mutation | Zygoty | AA change | DbSNP 128  | Observed in Genbank mRNAs?  |
|----------------|-----------|----------|--------|-----------|------------|---|
| chr8:125952138 | T         | A        | Homo   | S > T     | rs30558843 | Many, including CT26 [mRNA GU441834]                                  |
| chr8:125951873 | G         | A        | Hetero | W > *     |            | No  |
| chr8:125951822 | A         | G        | Hetero | Y > C     |            | CT26 [mRNA GU441834]  |
| chr8:125951717 | G         | A        | Hetero | W > *     |            | No  |
| chr8:125951634 | G         | A        | Homo   | E > K     |            | CT26, B16 (melanoma), RCB0527-Jyg-MC(B) & RCB0526-Jyg-MC(A) (mammary) |
| chr8:125951556 | G         | T        | Hetero | G > *     |            | RCB0526-Jyg-MC(A) (mammary tumor)                                     |
| chr8:125951208 | G         | A        | Homo   | G > S     |            | CT26 [mRNA GU441834]  |
| chr8:125950710 | G         | A        | Hetero | E > L     |            | RCB0526-Jyg-MC(A) (mammary tumor)                                     |
| chr8:125950284 | G         | A        | Hetero | G > R     | rs30722372 | No  |

AA, amino acid; \*, stop codon.

### Conclusion

This is the first integrated genome, transcriptome and immunome map of a mouse epithelial tumor. We found that the patterns of mutations in onco-relevant genes, the gene expression signatures and the regulated pathways in CT26 cells are in agreement with their origin in colon epithelia and share features with human primary CRCs. The mutations and expression profiles are similar to those reported for sporadic, undifferentiated, therapy-refractory, metastasis-prone human CRC. Moreover, we identified non-synonymous SNVs with predicted MHC class I binding capability which, together with the robust MHC class I expression of CT26 cells, provide a valuable resource for use of the CT26 model system to develop immunotherapeutic approaches.

The integrated use of mutation allele fraction and DNA copy number allowed us to determine the absolute copy number and zygoty for each mutation. The CT26 cells have extensive triploidy and tetraploidy and a high mutation rate (53 non-synonymous mutations per Mb). While Trp53, Braf, and Pik3ca are not mutated, Kras is mutated at G12D. Similar to human CRC samples, there

is a preference for C > T/G > A transitions. However, the CT26 mutation pattern shows a preference for C > T mutations at sites that are followed by a pyrimidine, a pattern that is more similar to that found in tumors from patients pre-treated with temozolomide than to that found in most human CRC tumors.

Clinically-approved patient selection biomarkers for anti-EGFR treatments cetuximab and panitumumab include assessment of EGFR levels and KRAS G12D mutation status. In CT26, we found the Kras G12D mutation and no expression of Egfr. Consistent with this, CT26 cells have been shown to be refractory to the rodent Egfr-targeting mAbs [2]. Similarly, KRAS G12D mutations and MAPK1 (MEK) and MET amplification are published biomarkers for colorectal tumor sensitivity to both MEK and MET inhibitors [3,4]. The homozygous Kras G12D mutation and Mapk1 and Met amplifications in CT26 suggest sensitivity to MEK and MET inhibition. In concordance with this, CT26 cells have been shown to be sensitive to MEK and MET inhibitors [2,37]. Further, the expression of markers such as Top2a and Cldn6 and lack of expression of Muc2, Epcam and Lgals4 show that CT26 cells are in an undifferentiated, proliferative state.

**Table 2 Potential CT26 T-cell tumor expression antigens**

| Gene   | CT Antigen | CT26   | Colon |
|--------|------------|--------|-------|
| Atad2  | CT137      | 42.4   | 0.5   |
| Casc5  | CT29       | 10.2   | 0.1   |
| Cep55  | CT111      | 25.4   | 0.4   |
| Ctage5 | CT21       | 17.7   | 1.8   |
| Dcaf12 | CT102      | 11.4   | 7.8   |
| Pbk    | CT84       | 39.6   | 1.0   |
| Rqcd1  | CT129      | 30.4   | 5.4   |
| Spag9  | CT89       | 14.7   | 1.2   |
| Ttk    | CT96       | 13.1   | 0.3   |
| Gp70   | -          | 7225.4 | 0.0   |

Known CT antigens with expression above 10 RPKM in CT26 and below 10 RPKM in normal mouse colon are shown. CT26 and mouse colon expression values are in RPKM units.

**Table 3 Eight potential CT26 mutation antigens that are in genes expressed in CT26 and in epitopes predicted to bind MHC class I molecules based on IEDB consensus ranks**

| Gene    | Mutation | Epitope              | MHC allele        |
|---------|----------|----------------------|-------------------|
| Csnk1g3 | N42K     | VGP <b>K</b> FRVGKK  | H-2D <sup>d</sup> |
| E2f8    | I522T    | TYLQPAQAQM           | H-2K <sup>d</sup> |
| Fam111a | G213E    | CVYGF <b>K</b> EETI  | H-2D <sup>d</sup> |
| Hdac2   | P228S    | KYYAVNFM             | H-2K <sup>d</sup> |
| Nudt19  | L335F    | IYMT <b>F</b> PSSENK | H-2K <sup>d</sup> |
| Phf3    | G1814E   | FPPQNM <b>F</b> EF   | H-2D <sup>d</sup> |
| Smc3    | D733A    | KFKAS <b>R</b> ASI   | H-2K <sup>d</sup> |

The bold letters signify the mutated amino acid.

<sup>d</sup>The official names of mouse MHC alleles.



Our study provides an overdue genomic and transcriptomic analysis of one of the most frequently used cell lines for drug development. Further, the results form the basis for the rationale design of pre-clinical studies using this model for drug development based on detailed molecular knowledge.

## Methods

**Samples:** BALB/c mice (Charles River) were kept in accordance with legal policies on animal research at the University of Mainz. In 2011, Germline BALB/cJ DNA was extracted from mouse tail. CT26.WT colon carcinoma cells were purchased from the American Type Culture Collection (Product: ATCC CRL-2638, Lot Number: 58494154). 3rd and 4th passages of cells were used for tumor experiments.

**NGS sequencing and data processing:** exome capture from CT26 and BALB/c mice were sequenced in triplicate using the Agilent Sure-Select solution-based mouse protein coding exome capture assay. CT26 oligo(dT)-isolated RNA for gene expression profiling was prepared in triplicate. Libraries were sequenced on an Illumina HiSeq2000. Protocol details are found in the Additional file 1. DNA-derived sequence reads were aligned to the mm9 genome using bwa [38] (default options, version 0.5.8c). Ambiguous reads mapping to multiple locations of the genome were removed. RNA-derived sequence reads were aligned using bowtie [39] to the mm9 genome and RefSeq exon-exon junctions. Default and “-v2 -best” parameters were used for transcriptome and genome alignments, respectively.

For the exome reads, there was an average of 103 million read pairs per sample. As each sample was sequenced in triplicate, this resulted in over 300 million 50 nt paired-end reads for the CT26 and BALB/cJ exomes. 83% of the reads mapped to the mm9 reference genome, with 51% of the nucleotides on target, resulting in a mean coverage of 170x. The CT26 transcriptome was sequenced in triplicate with an average of 27 million reads and total of 81 million reads, of which 94% could be aligned. NGS read statistics are in Additional file 1.

**DNA copy number:** absolute allele copy number, and mutation allele fraction were simultaneously determined using a novel algorithm that assumes a) that mutation allele fraction can take only discrete values in tumor cells based on allele copy number and b) that the relative tumor to germline number of exome-seq reads mapping to a gene locus is proportional to locus copy number [40]. Copy number estimations are in Additional file 2.

**Mutation identification:** single nucleotide mutations (SNVs) that were identified by all algorithms samtools [18], Mutect [41], and SomaticSniper [42] and in the replicates were further filtered using binomial filters that eliminate erroneous tumor observations and decrease the likelihood that a mutation is classified as somatic

due to lack of coverage in the germline sample. Insertions and deletions (indels) were identified using samtools and Varscan2 with at least 10 DNA reads support and further filtered by removing indels with germline support after realigning the reads to an integrated wild-type and mutated reference genome. SNVs and indels are in Additional files 3 and 4.

**SNP detection:** SNPs were detected by running the samtools mpileup command (version 0.1.19) on sites defined by dbSNP (version 128 for mm9), using the BALB/c and CT26 exome alignments as input and binning the results by the phred scaled SNP quality as returned by samtools/bcftools.

**Gene expression:** expression values were determined by counting reads overlapping transcript exons and junctions, and normalizing to RPKM expression units (Reads which map Per Kilobase of transcript length per Million mapped reads). 10 RPKM is roughly the 80th percentile (80% of the gene expression values fall below 10 RPKM). Gene expression values are in Additional file 5.

**Pathway enrichment:** the ENCODE Consortium profiled two normal mouse colons in triplicate using RNA-Seq [43]; raw data were downloaded and processed through the computational workflow used for the CT26 RNA-Seq reads. Gene expression profiles from the triplicate CT26 and six normal mouse colon RNA-Seq runs were statistically compared using a t-test. Enriched Reactome [44] gene sets were identified using GSEA [45] and Cytoscape ClueGO [46] and over-expressed genes (t-test > 20). Enriched Reactome pathways are in Additional file 6. Gene set enrichment was performed using GenePattern [47], the Molecular Signatures Database [48], and the expression ranked gene list. Enriched GenePattern gene sets are listed in Additional file 7 and gene membership is listed in Additional file 8. All identifiers were translated from mouse to human using Homologene [49]. The list of cancer testes (CT) antigens was from the CTdatabase [50].

**MHC typing and expression:** typing and expression were determined using RNA-Seq reads and the seq2HLA algorithm [51] using the parameter setting “-best” rather than “-a”. All mouse tissue samples were sequenced (RNA-Seq) by us except the normal colon dataset, which was retrieved from the ENCODE project. RNA-Seq fastq reads were mapped according to the parameters described in Boegel et al. [51]. Two distinct reference files were created for BALB/c, containing reference sequences for H-2D<sup>d</sup>, H-2K<sup>d</sup>, H-2L<sup>d</sup> and H-2Ia, and for C57BL/6 containing reference sequences for H-2D<sup>b</sup>, H-2K<sup>b</sup>, H-2Ia<sup>b</sup>. Expression was determined by the total number of unique sequence reads mapping to class I or class II genes and normalized according to reads per kilobase of exon model per million mapped reads (RPKM) using the length of the allele transcripts contained in the

reference dataset: H-2D<sup>b</sup> = 1567 nt, H-2K<sup>b</sup> = 1564 nt, H-2Ia<sup>b</sup> = 932 nt, H-2D<sup>d</sup> = 1586 nt, H-2K<sup>d</sup> = 1540 nt, H-2L<sup>d</sup> = 1102 nt, H-2Ia<sup>d</sup> = 978 nt.

**MHC binding:** MHC binding predictions were performed using the IEDB algorithm v2.5 [35], “consensus” setting, the CT26 cell-line specific MHC type and the identified somatic point mutations. The best neo-epitope for a mutation was calculated as follows: all possible 8-, 9-, 10-, 11-mer peptides containing the mutated amino acids were input to the IEDB algorithm, which predicts the binding affinity (IC<sub>50</sub> in nM and the consensus percentile rank) of the peptide to the cell line HLA alleles. The best neo-epitope-MHC pair was defined as the peptide which has the strongest predicted binding affinity to the respective MHC allele. Epitopes with a consensus percentile rank of less than or equal to 1% are reported as likely immunogenic.

#### Availability of supplementary information

CT26 and BALB/cJ NGS fastq reads are available from ENA as PRJEB5320 (RNA-Seq) and PRJEB5321 (Exome).

#### Additional files

**Additional file 1:** Contains supplementary methods, supplementary tables, NGS read statistics, and the gp70 mutations and protein sequence.

**Additional file 2:** Absolute copy number for each gene determined using the number of exome-seq reads mapping to each gene from CT26 and balb/c samples and using the allele fraction to determine ploidy. Columns include gene, copy number, normalized ratio, chromosome and gene start coordinate in the mm9 assembly.

**Additional file 3:** The 3,023 high confidence point mutations found in CT26 transcripts. Columns include chromosomal position of the mutation, reference and observed nt, classification non/synonymous, classification UTR/CDS, amino acid substitution, gene symbol, transcript ID and affected exon, mean gene expression (combined of gene expression and exon expression), possible repeat region, dbSNP ID and dbSNP validation source, MHC prediction for MHC class I and class II alleles for the mutated neo-epitope and corresponding wild type peptide, allele to which it is binding, percentile rank, neo-epitope sequence, IC<sub>50</sub> [nM]. The mutation-containing epitope and associated MHC allele were selected using the IEDB algorithm v2.5 [3], with consensus setting, with the listed epitope and MHC being the pair with the predicted strongest binding.

**Additional file 4:** The 363 insertion and deletion mutations. Columns include chromosomal position of the mutation, reference and observed nt(s), location of indel, frameshift mutation, classification non/synonymous, in UTR, gene symbol, UCSC transcript ID (if the mutation can occur in more than one transcript of this gene they are separated by a space), transcript ID and affected exon, mean expression (combined of gene expression and exon expression), possible repeat region, dbSNP ID and dbSNP validation source, allele frequency for CT26 and Balb/c replicates, number of reads per sample supporting this indel.

**Additional file 5:** Gene expression values for CT26 and ENCODE normal mouse colon samples in RPKM values and raw read counts.

**Additional file 6:** Results of the GSEA gene set Reactome pathway enrichment, including Reactome pathway name, gene membership and FDR q-values.

**Additional file 7:** Results of the GenePattern enrichment using ranked ordered expression. Gene sets included those curated from

literature and overexpression was determined using GenePattern [4]. Gene membership and enrichment values are in the file.

**Additional file 8:** Contains the Gene Pattern gene set membership and enrichment values in an html format. The file index.html is the entry point.

#### Competing interests

The authors disclose no competing interests. The authors are employed by organizations developing cancer immunotherapies.

#### Authors' contributions

JG, CD and VB performed NGS and Sanger sequencing; TB and PS processed all NGS reads; ML, SB, CB and AT identified mutations, gene expression pathways and predicted immunogenicity; MD and SK generated samples; JC, SK and UG conceived of the experiment; JC wrote the manuscript; OT edited the manuscript. All authors approved the final manuscript.

#### Acknowledgements

We thank Corina Cosma-Busch and Goran Martic for project management; Julia Beckerle and Meike Wagner for lab work; and Ludmila Schemarow, Bernhard Renard, Marius Byl, Jelle Scholtalbers, Thorsten Litzenger, André Brinkman, Tim Süß and Markus Tacke for the computational infrastructure.

#### Author details

<sup>1</sup>TRON gGmbH - Translational Oncology, Johannes Gutenberg-University Medical Center gGmbH, Langenbeckstr. 1, Building 708, 55131 Mainz, Germany. <sup>2</sup>University Medical Center of the Johannes Gutenberg-University Mainz, 55131 Mainz, Germany. <sup>3</sup>BioNTech AG, Kupferbergterrasse 17-19, 55131 Mainz, Germany.

Received: 2 November 2013 Accepted: 21 February 2014

Published: 13 March 2014

#### References

1. Griswold DP, Corbett TH: A colon tumor model for anticancer agent evaluation. *Cancer* 1975, **36**:2441–2444.
2. van Houdt WJ, Hoogwater FJ, de Bruijn MT, Emmink BL, Nijkamp MW, Raats DA, van der Groep P, van Diest P, Borel Rinkes IH, Kranenburg O: Oncogenic KRAS desensitizes colorectal tumor cells to epidermal growth factor receptor inhibition and activation. *Neoplasia* 2010, **12**:443–452.
3. Yeh JJ, Routh ED, Rubinas T, Peacock J, Martin TD, Shen XJ, Sandler RS, Kim HJ, Keku TO, Der CJ: KRAS/BRAF mutation status and ERK1/2 activation as biomarkers for MEK1/2 inhibitor therapy in colorectal cancer. *Mol Cancer Ther* 2009, **8**:834–843.
4. Ma PC, Schaefer E, Christensen JG, Salgia R: A selective small molecule c-MET Inhibitor, PHA665752, cooperates with rapamycin. *Clin Cancer Res* 2005, **11**:2312–2319.
5. Zhang B, Halder SK, Zhang S, Datta PK: Targeting transforming growth factor-beta signaling in liver metastasis of colon cancer. *Cancer Lett* 2009, **277**:114–120.
6. Dalerba P, Kalisky T, Sahoo D, Rajendran PS, Rothenberg ME, Leyrat AA, Sim S, Okamoto J, Johnston DM, Qian D, Zabala M, Bueno J, Neff NF, Wang J, Shelton AA, Visser B, Hisamori S, Shimono Y, van de Wetering M, Clevers H, Clarke MF, Quake SR: Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* 2011, **29**:1120–1127.
7. De Sousa EMF, Wang X, Jansen M, Fessler E, Trinh A, de Rooij LP, de Jong JH, de Boer OJ, van Leersum R, Bijlsma MF, Rodermond H, van der Heijden M, van Noesel CJ, Tuynman JB, Dekker E, Markowitz F, Medema JP, Vermeulen L: Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. *Nat Med* 2013, **19**:614–618.
8. Sadanandam A, Lyssiotis CA, Homicsko K, Collisson EA, Gibb WJ, Wullschlegel S, Ostos LC, Lannon WA, Grotzinger C, Del Rio M, Lhermitte B, Olshen AB, Wiedenmann B, Cantley LC, Gray JW, Hanahan D: A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat Med* 2013, **19**:619–625.
9. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, Etienne-Grimaldi MC, Schiappa R, Guenot D, Ayadi M, Kirzin S, Chazal M, Fléjou JF, Benchimol D, Berger A, Lagarde A, Pencreach E, Piard F, Elias D, Parc Y, Olschwang S, Milano G, Laurent-Puig P, Boige V: Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 2013, **10**:e1001453.

10. Senovilla L, Vitale I, Martins I, Tailler M, Pailleret C, Michaud M, Galluzzi L, Adjemian S, Kepp O, Niso-Santano M, Shen S, Mariño G, Criollo A, Boilève A, Job B, Ladoire S, Ghiringhelli F, Sistigu A, Yamazaki T, Rello-Varona S, Locher C, Poirier-Colame V, Talbot M, Valent A, Berardinelli F, Antocchia A, Ciccocanti F, Fimia GM, Piacentini M, Fueyo A, et al: **An immunosurveillance mechanism controls cancer cell ploidy.** *Science* 2012, **337**:1678–1684.
11. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Börresen-Dale AL, Bouyalt S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Jäger N, Jones DT, Jones D, Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MML-Seq Consortium; ICGC PedBrain, et al: **Signatures of mutational processes in human cancer.** *Nature* 2013, **500**:415–421.
12. Cancer Genome Atlas N: **Comprehensive molecular characterization of human colon and rectal cancer.** *Nature* 2012, **487**:330–337.
13. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, Szabo S, Buckhaults P, Farrell C, Meeh P, Markowitz SD, Willis J, Dawson D, Willson JK, Gazdar AF, Hartigan J, Wu L, Liu C, Parmigiani G, Park BH, Bachman KE, Papadopoulos N, Vogelstein B, Kinzler KW, Velculescu VE: **The consensus coding sequences of human breast and colorectal cancers.** *Science* 2006, **314**:268–274.
14. Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J, Silliman N, Szabo S, Dezso Z, Ustyanksky V, Nikolskaya T, Nikolsky Y, Karchin R, Wilson PA, Kaminker JS, Zhang Z, Croshaw R, Willis J, Dawson D, Shipitsin M, Willson JK, Sukumar S, Polyak K, Park BH, Pethiyagoda CL, Pant PV, et al: **The genomic landscapes of human breast and colorectal cancers.** *Science* 2007, **318**:1108–1113.
15. Maurer T, Garrenton LS, Oh A, Pitts K, Anderson DJ, Skelton NJ, Fauber BP, Pan B, Malek S, Stokoe D, Ludlam MJ, Bowman KK, Wu J, Giannetti AM, Starovashnik MA, Mellman I, Jackson PK, Rudolph J, Wang W, Fang G: **Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity.** *Proc Natl Acad Sci USA* 2012, **109**:5299–5304.
16. Kambara T, Simms LA, Whitehall VL, Spring KJ, Wynter CV, Walsh MD, Barker MA, Arnold S, McGovern A, Matsubara N, Tanaka N, Higuchi T, Young J, Jass JR, Leggett BA: **BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum.** *Gut* 2004, **53**:1137–1144.
17. Eguchi T, Takaki T, Itadani H, Kotani H: **RB silencing compromises the DNA damage-induced G2/M checkpoint and causes deregulated expression of the ECT2 oncogene.** *Oncogene* 2007, **26**:509–520.
18. Tiwari N, Tiwari VK, Waldmeier L, Balwiercz PJ, Arnold P, Pachkov M, Meyer-Schaller N, Schubeler D, van Nimwegen E, Christofori G: **Sox4 is a master regulator of epithelial-mesenchymal transition by controlling ezh2 expression and epigenetic reprogramming.** *Cancer Cell* 2013, **23**:768–783.
19. Tong ZT, Cai MY, Wang XG, Kong LL, Mai SJ, Liu YH, Zhang HB, Liao YJ, Zheng F, Zhu W, Liu TH, Bian XW, Guan XY, Lin MC, Zeng MS, Zeng YX, Kung HF, Xie D: **EZH2 supports nasopharyngeal carcinoma cell aggressiveness by forming a co-repressor complex with HDAC1/HDAC2 and Snail to inhibit E-cadherin.** *Oncogene* 2012, **31**:583–594.
20. Bullock MD, Bruce A, Sreekumar R, Curtis N, Cheung T, Reading I, Primrose JN, Ottensmeier C, Packham GK, Thomas G, Mirnezami AH: **FOXO3 expression during colorectal cancer progression: biomarker potential reflects a tumour suppressor role.** *Br J Cancer* 2013, **109**:387–394.
21. Winnepenninckx V, Lazar V, Michiels S, Dessen P, Stas M, Alonso SR, Avril MF, Ortiz Romero PL, Robert T, Balacescu O, Eggermont AM, Lenoir G, Sarasin A, Tursz T, van den Oord JJ, Spatz A, Melanoma Group of the European Organization for Research and Treatment of Cancer: **Gene expression profiling of primary cutaneous melanoma and clinical outcome.** *J Natl Cancer Inst* 2006, **98**:472–482.
22. Michel J, Schonhaar K, Schledzewski K, Gkaniatsou C, Sticht C, Kellert B, Lasitschka F, Geraud C, Goerd S, Schmieder A: **Identification of the novel differentiation marker MS4A8B and its murine homolog MS4A8A in colonic epithelial cells lost during neoplastic transformation in human colon.** *Cell Death Dis* 2013, **4**:e469.
23. Furth EE, Li J, Purev E, Solomon AC, Rogler G, Mick R, Putt M, Zhang T, Somasundaram R, Swoboda R, Herlyn D: **Serum antibodies to EpCAM in healthy donors but not ulcerative colitis patients.** *Cancer Immunol Immunother* 2006, **55**:528–537.
24. Thiery JP: **Epithelial-mesenchymal transitions in tumour progression.** *Nature reviews Cancer* 2002, **2**:442–454.
25. de Sousa EMF, Colak S, Buikhuizen J, Koster J, Cameron K, de Jong JH, Tuynman JB, Prasetyanti PR, Fessler E, van den Bergh SP, Rodermond H, Dekker E, van der Loos CM, Pals ST, van de Vijver MJ, Versteeg R, Richel DJ, Vermeulen L, Medema JP: **Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients.** *Cell Stem Cell* 2011, **9**:476–485.
26. Gregorieff A, Clevers H: **Wnt signaling in the intestinal epithelium: from endoderm to cancer.** *Genes Dev* 2005, **19**:877–890.
27. Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ, Clevers H: **Identification of stem cells in small intestine and colon by marker gene Lgr5.** *Nature* 2007, **449**:1003–1007.
28. Jordan KR, McMahan RH, Kemmler CB, Kappler JW, Slansky JE: **Peptide vaccines prevent tumor growth by activating T cells that respond to native tumor antigens.** *Proc Natl Acad Sci USA* 2010, **107**:4652–4657.
29. McCubrey J, Risser R: **Genetic interactions in the spontaneous production of endogenous murine leukemia virus in low leukemic mouse strains.** *J Exp Med* 1982, **156**:337–349.
30. McWilliams JA, Sullivan RT, Jordan KR, McMahan RH, Kemmler CB, McDuffie M, Slansky JE: **Age-dependent tolerance to an endogenous tumor-associated antigen.** *Vaccine* 2008, **26**:1863–1873.
31. DeLeo AB, Shiku H, Takahashi T, John M, Old LJ: **Cell surface antigens of chemically induced sarcomas of the mouse. I. Murine leukemia virus-related antigens and alloantigens on cultured fibroblasts and sarcoma cells: description of a unique antigen on BALB/c Meth A sarcoma.** *J Exp Med* 1977, **146**:720–734.
32. Hofmann O, Caballero OL, Stevenson BJ, Chen YT, Chua R, Maher CA, Panji S, Schaefer U, Kruger A, Lehtvaslaihho M, Carninci P, Hayashizaki Y, Jongeneel CV, Simpson AJ, Old LJ, Hide W: **Genome-wide analysis of cancer/testis gene expression.** *Proc Natl Acad Sci USA* 2008, **105**:20422–20427.
33. Wolfel T, Hauer M, Schneider J, Serrano M, Wolfel C, Klehmann-Hieb E, De PE, Hankeln T, Buschenfelde KH M z, Beach D: **A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma.** *Science* 1995, **269**:1281–1284.
34. Castle JC, Kreiter S, Diekmann J, Lower M, van de Roemer N, de Graaf J, Selmi A, Diken M, Boegel S, Paret C, Koslowski M, Kuhn AN, Britten CM, Huber C, Türeci O, Sahin U: **Exploiting the mutanome for tumor vaccination.** *Cancer Res* 2012, **72**:1081–1091.
35. Kim Y, Sette A, Peters B: **Applications for T-cell epitope queries and tools in the immune epitope database and analysis resource.** *J Immunol Methods* 2011, **374**:62–69.
36. Segal NH, Parsons DW, Peggs KS, Velculescu V, Kinzler KW, Vogelstein B, Allison JP: **Epitope landscape in breast and colorectal cancer.** *Cancer Res* 2008, **68**:889–892.
37. Bellon SF, Kaplan-Lefko P, Yang Y, Zhang Y, Moriguchi J, Rex K, Johnson CW, Rose PE, Long AM, O'Connor AB, Gu Y, Coxon A, Kim TS, Tasker A, Burgess TL, Dussault I: **c-Met inhibitors with novel binding mode show activity against several hereditary papillary renal cell carcinoma-related mutations.** *J Biol Chem* 2008, **283**:2675–2683.
38. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**:1754–1760.
39. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.
40. Castle JC, Biery M, Bouzek H, Xie T, Chen R, Misura K, Jackson S, Armour CD, Johnson JM, Rohl CA, Raymond CK: **DNA copy number, including telomeres and mitochondria, assayed using next-generation sequencing.** *BMC Genomics* 2010, **11**:244.
41. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G: **Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.** *Nat Biotechnol* 2013, **31**:213–219.
42. Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK, Ding L: **SomaticSniper: identification of somatic point mutations in whole genome sequencing data.** *Bioinformatics* 2012, **28**:311–317.
43. Consortium EP: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9**:e1001046.
44. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L: **Reactome: a knowledgebase of biological pathways.** *Nucleic Acids Res* 2005, **33**:D428–D432.
45. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP: **GSEA-P: a desktop application for gene set enrichment analysis.** *Bioinformatics* 2007, **23**:3251–3253.

46. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z, Galon J: **ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.** *Bioinformatics* 2009, **25**:1091–1093.
47. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP: **GenePattern 2.0.** *Nat Genet* 2006, **38**:500–501.
48. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci USA* 2005, **102**:15545–15550.
49. Coordinators NR: **Database resources of the national center for biotechnology information.** *Nucleic Acids Res* 2013, **41**:D8–D20.
50. Almeida LG, Sakabe NJ: **CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens.** *Nucleic Acids Res* 2009, **37**:D816–D819.
51. Boegel S, Lower M, Schafer M, Bukur T, de Graaf J, Boisguerin V, Tureci O, Diken M, Castle JC, Sahin U: **HLA typing from RNA-Seq sequence reads.** *Genome Med* 2013, **4**:102.

doi:10.1186/1471-2164-15-190

**Cite this article as:** Castle *et al.*: Immunomic, genomic and transcriptomic characterization of CT26 colorectal carcinoma. *BMC Genomics* 2014 **15**:190.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

