

RESEARCH

Open Access



# Restoration of KMT2C/MLL3 in human colorectal cancer cells reinforces genome-wide H3K4me1 profiles and influences cell growth and gene expression

Chatarina Larsson<sup>1</sup>, Lina Cordeddu<sup>2</sup>, Lee Siggins<sup>2</sup>, Tatjana Pandzic<sup>1</sup>, Snehangshu Kundu<sup>1</sup>, Liquan He<sup>1</sup>, Muhammad Akhtar Ali<sup>1,3</sup>, Nuša Pristovšek<sup>1,4</sup>, Karin Hartman<sup>1</sup>, Karl Ekwall<sup>2</sup> and Tobias Sjöblom<sup>1\*</sup> 

## Abstract

**Background:** The histone 3 lysine 4 (H3K4) monomethylase KMT2C is mutated across several cancer types; however, the effects of mutations on epigenome organization, gene expression, and cell growth are not clear. A frequently recurring mutation in colorectal cancer (CRC) with microsatellite instability is a single nucleotide deletion within the exon 38 poly-A(9) repeat (c.8390delA) which results in frameshift preceding the functional carboxy-terminal SET domain. To study effects of *KMT2C* expression in CRC cells, we restored one allele to wild type *KMT2C* in the two CRC cell lines RKO and HCT116, which both are homozygous c.8390delA mutant.

**Results:** Gene editing resulted in increased *KMT2C* expression, increased H3K4me1 levels, altered gene expression profiles, and subtle negative effects on cell growth, where higher dependence and stronger effects of *KMT2C* expression were observed in RKO compared to HCT116 cells. Surprisingly, we found that the two RKO and HCT116 CRC cell lines have distinct baseline H3K4me1 epigenomic profiles. In RKO cells, a flatter genome-wide H3K4me1 profile was associated with more increased H3K4me1 deposition at enhancers, reduced cell growth, and more differential gene expression relative to HCT116 cells when *KMT2C* was restored. Profiling of H3K4me1 did not indicate a highly specific regulation of gene expression as *KMT2C*-induced H3K4me1 deposition was found globally and not at a specific enhancer sub-set in the engineered cells. Although we observed variation in differentially regulated gene sets between cell lines and individual clones, differentially expressed genes in both cell lines included genes linked to known cancer signaling pathways, estrogen response, hypoxia response, and aspects of immune system regulation.

**Conclusions:** Here, *KMT2C* restoration reduced CRC cell growth and reinforced genome-wide H3K4me1 deposition at enhancers; however, the effects varied depending upon the H3K4me1 status of *KMT2C* deficient cells. Results indicate that *KMT2C* inactivation may promote colorectal cancer development through transcriptional dysregulation in several pathways with known cancer relevance.

**Keywords:** KMT2C, MLL3, H3K4me1, Cancer

\* Correspondence: [tobias.sjoblom@igp.uu.se](mailto:tobias.sjoblom@igp.uu.se)

<sup>1</sup>Department of Immunology, Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85 Uppsala, Sweden

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Epigenetic modifiers are frequently mutated in cancer [1, 2]. Causal links to cancer development are established for some but not all of these genes. Proteins of the mixed lineage leukemia (MLL), also known as histone-lysine N-methyltransferase 2 (KMT2), family are subunits of multi-protein complexes that deposit methyl groups on lysine 4 of histone 3 through a SET-domain catalyzed histone methyltransferase activity [3]. The genes encoding the KMT2 proteins have been identified to be altered in several cancer types (reviewed in [4, 5]). The first family member to be linked to cancer development was KMT2A (MLL1), for which translocations resulting in oncogenic fusion proteins were first discovered in leukemias, and non-synonymous mutations, predominantly frameshift and nonsense, are frequently found in several solid tumor types [4, 5]. The genes encoding KMT2C (MLL3) and KMT2D (MLL4) are among the most frequently mutated in cancer [1]. These two proteins function in maintenance of histone 3 lysine 4 monomethylation (H3K4me1) levels at enhancer elements (reviewed in [6]). Analysis of clinical samples has revealed reduced *KMT2C* expression in larynx carcinoma [7], pancreatic ductal adenocarcinoma [8], and gastric cancer [9], and silencing of *KMT2C* due to promoter DNA hypermethylation has been observed in urothelial cancer [10]. The *KMT2C* gene is located on chromosome 7q36.1, which is commonly deleted in hematological malignancies [11, 12]. Deletion of *KMT2C* has also been identified in colorectal cancer (CRC) [13], and somatic mutations in *KMT2C* have been identified as potential drivers of tumorigenesis in several tumor types, including CRC [1, 14]. Missense and non-sense germline *KMT2C* variants have also been associated with cancer development in families with suspected hereditary cancer [15–18]. Of mutations present in the COSMIC database, 28.3% of *KMT2C* and 37.0% of *KMT2D* mutations, primarily frameshift and nonsense mutations, were previously found to impact the catalytic SET domain of the respective proteins [4]. A substantial proportion of mutations, notably many missense mutations, was also found in the PHD domains of *KMT2C* (17.1%) and *KMT2D* (12.9%). The mutational pattern suggests a tumor suppressor function of *KMT2C* which may be disrupted by differently localized mutations.

Several observations and experimental data further support the notion of *KMT2C* as a tumor suppressor gene. Forward genetic screens based on transposon mutagenesis have identified common insertion sites in the *Kmt2c* locus in mouse models of pancreatic adenocarcinoma and APC-deficient colorectal cancer development [19, 20]. Historically much focus has been on studying the function of the SET domain in cancer cells through knockout models. Specifically, biallelic knockout

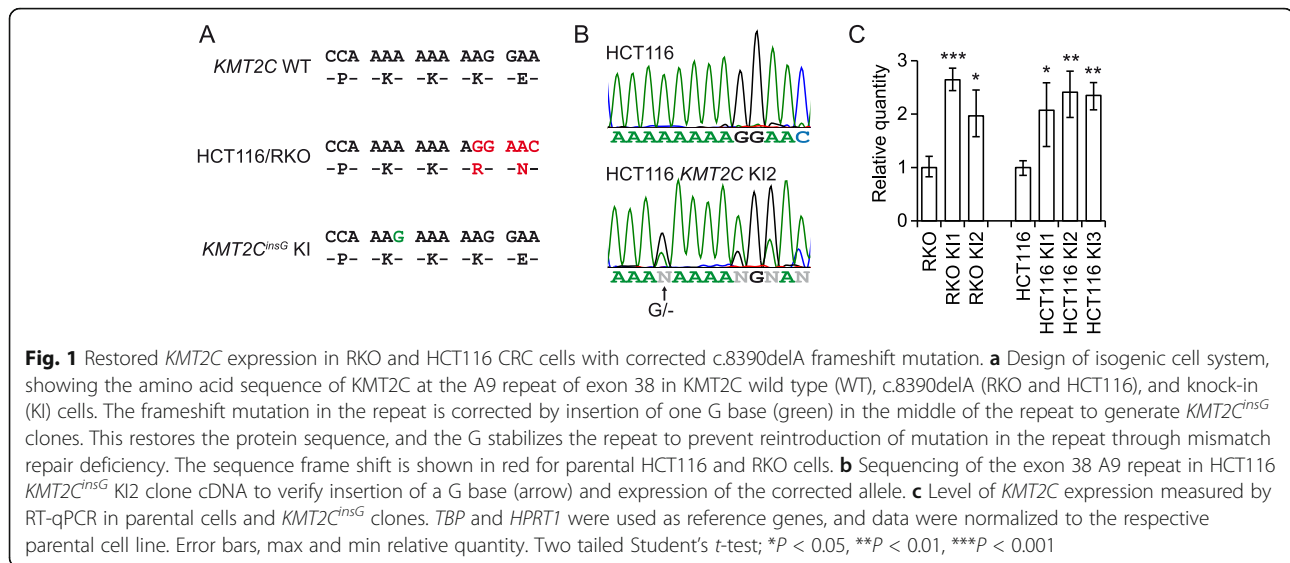
of the SET domain of *Kmt2c/Mll3* in mice led to formation of epithelial tumors, which supports a tumor suppressor role for the H3K4 monomethylase function of the gene [21]. More recently, PHD domain mutations in *KMT2C* were studied in human cells, revealing that these disrupt the interaction between *KMT2C* and the tumor suppressor BAP1. This induces transcriptional changes, possibly by reduced recruitment of MLL3 and UTX/KDM6A to enhancers [22], and demonstrates the value of specific study of cancer mutations to discern their functional impact in cells.

Around 15% of CRC are classified to have microsatellite instability (MSI) caused by deficiency in the mismatch repair (MMR) pathway. This subtype of tumors is prone to acquisition of frameshift mutations in short repeat sequences such as mononucleotide repeats [23, 24]. The *KMT2C* gene harbors a poly(A)<sub>9</sub> repeat within the coding sequence of exon 38, and a recurrent frameshift mutation in this mononucleotide repeat, c.8390delA, was previously reported in 25–48% of MSI CRC cases [25, 26] and has also been shown prevalent in MSI gastric cancer [27]. The high mutation prevalence and possible loss-of-function effect [25, 26] motivated us to examine the role of *KMT2C* in CRC cells by restoring wild type *KMT2C* in two homozygous c. 8390delA mutant human CRC cell lines by genome editing. Comparison of cells with and without functional *KMT2C* allowed profiling of changes in H3K4me1 deposition at enhancers and changes in gene expression brought about by *KMT2C* expression, revealing that *KMT2C* induced a moderate global increase in H3K4me1 which affected colorectal enhancers associated with differentially expressed genes in the two isogenic cell models.

## Results

### Restoration of *KMT2C* expression in MSI CRC cells by gene targeting

A characteristic of MSI cancers is the accumulation of frameshift mutations in microsatellite repeats. The two MSI CRC cell lines RKO and HCT116 both contain homozygous c.8390delA frameshift mutations in exon 38 that are reported to lead to nonsense-mediated decay (NMD) and loss-of-function due to loss of the SET domain at the C-terminus of the protein [25, 26]. By recombinant adeno-associated virus (rAAV) mediated gene targeting in these cell lines, we aimed to restore protein function by correcting the amino acid sequence of one of the two affected alleles in each cell line to allow for full length protein expression. Since microsatellite repeats are prone to undergo frameshift mutation in MSI cancer cells, we used a strategy where we stabilized the repeat sequence by addition of a G base in replacement of the lost A base in the affected A<sub>9</sub> repeat (Fig. 1a), as described by us previously [28]. This leads



to insertion of the correct amino acid in the protein and restoration of the reading frame, and the shortened A repeat is less likely to contract frameshift mutations. Targeted clones were identified by PCR, and insertion was validated by sequencing over the target poly-A repeat in DNA (not shown) and RNA (Fig. 1b, Additional File 1, Fig. S1), which verified that the corrected allele was expressed as mRNA in all clones. Three clones with correct targeting were confirmed for the HCT116 cell line, and two clones were confirmed for the RKO cell line. We predicted restoration of the reading frame to rescue the mRNA from the NMD previously described for c.8390delA mutant *KMT2C* [25], and by RT-qPCR we detected a 2-fold increase in *KMT2C* expression in the targeted clones (Fig. 1c). This indicates that wild type *KMT2C* expression was restored in RKO and HCT116 *KMT2C*<sup>insG</sup> cell lines and that the expression level was upregulated to the same extent in both cell lines.

#### Gene expression profiling in *KMT2C*<sup>insG</sup> clones

Monomethylation of H3K4 in combination with H3K27ac is a mark of active enhancers [29], and restoration of *KMT2C* function is therefore likely to increase gene transcription. We used RNA sequencing to determine the expression profile of all *KMT2C*<sup>insG</sup> clones and the respective parental cell lines. We set a threshold for differential expression at 1.5 log<sub>2</sub> fold up- or downregulated in at least one *KMT2C*<sup>insG</sup> clone relative to the respective parental cell line. For validation of data, up- and downregulated genes were selected for each cell line based on fold regulation and presence of data in at least two clones per respective cell line. RT-qPCR on independently collected samples from the cell lines verified the pattern of expression detected by RNA sequencing

for all of the eleven investigated genes (Additional File 1, Fig. S2). The RKO cells showed the highest number of genes (*n* = 853) with changed expression upon *KMT2C* restoration, with clone KI1 having 265 differentially regulated genes (DEGs) of which 57% were upregulated, and clone KI2 having 678 DEGs of which 65% were upregulated (Additional File 2, Table S1). Of the DEGs in clone 1, 90 (34%) were differentially regulated also in clone 2, which is more than expected by chance. If differential expression was random, 0–16 genes would be expected with 99% confidence to overlap between the two RKO clones (Additional File 2, Table S2). For HCT116 clones, 353 genes were affected in total, with 73, 137, and 209 DEGs in clones KI1, KI2, and KI3, respectively, and between 40–52% of these were upregulated (Additional File 2, Table S1). Also in HCT116, the number of DEGs overlapping between clones was greater than what would be expected by chance, although with lower numbers than for RKO clones (Additional File 2, Table S2). Because of H3K4me1 being an activating mark on enhancer elements, we separately analyzed the overlapping DEGs that were upregulated by *KMT2C* expression in the cell lines, revealing that also for those the overlap was greater than what would be expected from random differential expression (Additional File 2, Table S3). To investigate overlap between RKO and HCT116 cells, we included all genes that were differentially regulated in ≥ 1 KI clone. Of these, a total of 61 DEGs were common for both cell lines, but for only 29 the expression change of the affected *KMT2C*<sup>insG</sup> clones was regulated in the same direction for both RKO and HCT116 cell lines (Additional File 3, Tables S4–5). Investigation of the expression level of these 29 genes in all individual KI clones revealed that only five genes (*ANK1*, *PRSS23*, *SAMD9*, *TSPAN1*, *WFIKKN1*) were regulated in the same direction

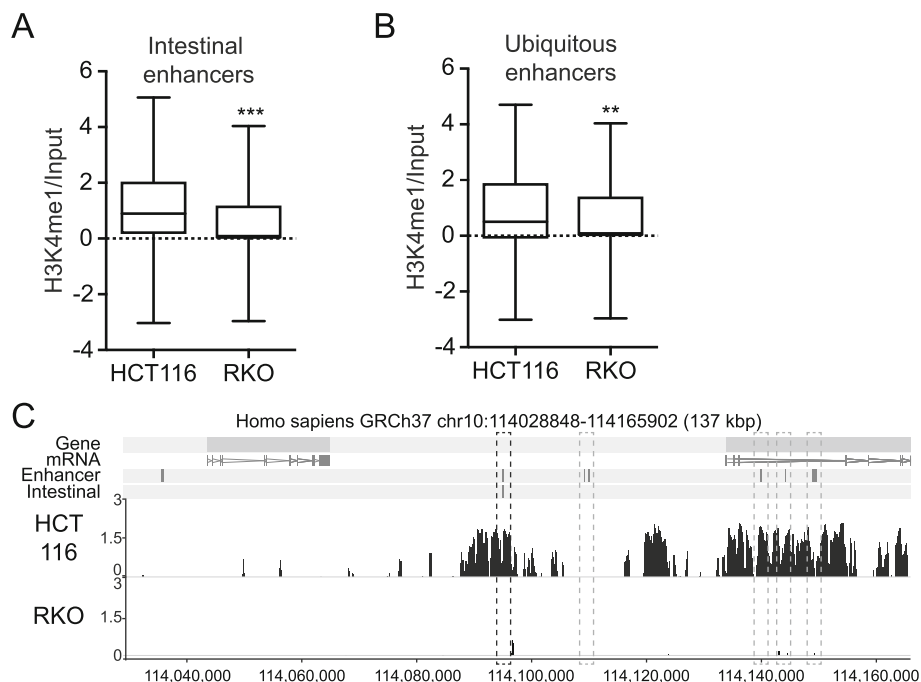
in all the five individual *KMT2C<sup>insG</sup>* clones. By RT-qPCR, we could validate upregulation of *ANK1*, *PRSS23*, *SAMD9*, and *TSPAN1*, however not to a significant level in all HCT116 clones, but we could not validate downregulation of *WFIKKN1* (Additional File 1, Fig. S3). Gene set overlap analysis for the respective sets of DEGs in HCT116 and RKO showed enrichment for genes in cancer-related signaling pathways such as KRAS, TNFA, and TGF- $\beta$  for both RKO and HCT116 cell lines (Additional File 3, Table S6). In addition, both cell lines showed enrichment of DEGs involved in hypoxia, response to estrogen, and several immune system processes. Also for the upregulated genes estrogen response, and KRAS, TNFA, and TGF- $\beta$  signaling were among the hallmark functions with significant overlap among the genes (Additional File 3, Table S7). In summary, these data showed that restoration of *KMT2C* expression in CRC cells altered the expression of hundreds of genes that, although the altered set differed between cell lines and individual *KMT2C<sup>insG</sup>* clones, converged in common functional groups with cancer relevance.

#### Profiling of H3K4 monomethylation in *KMT2C<sup>insG</sup>* cell lines

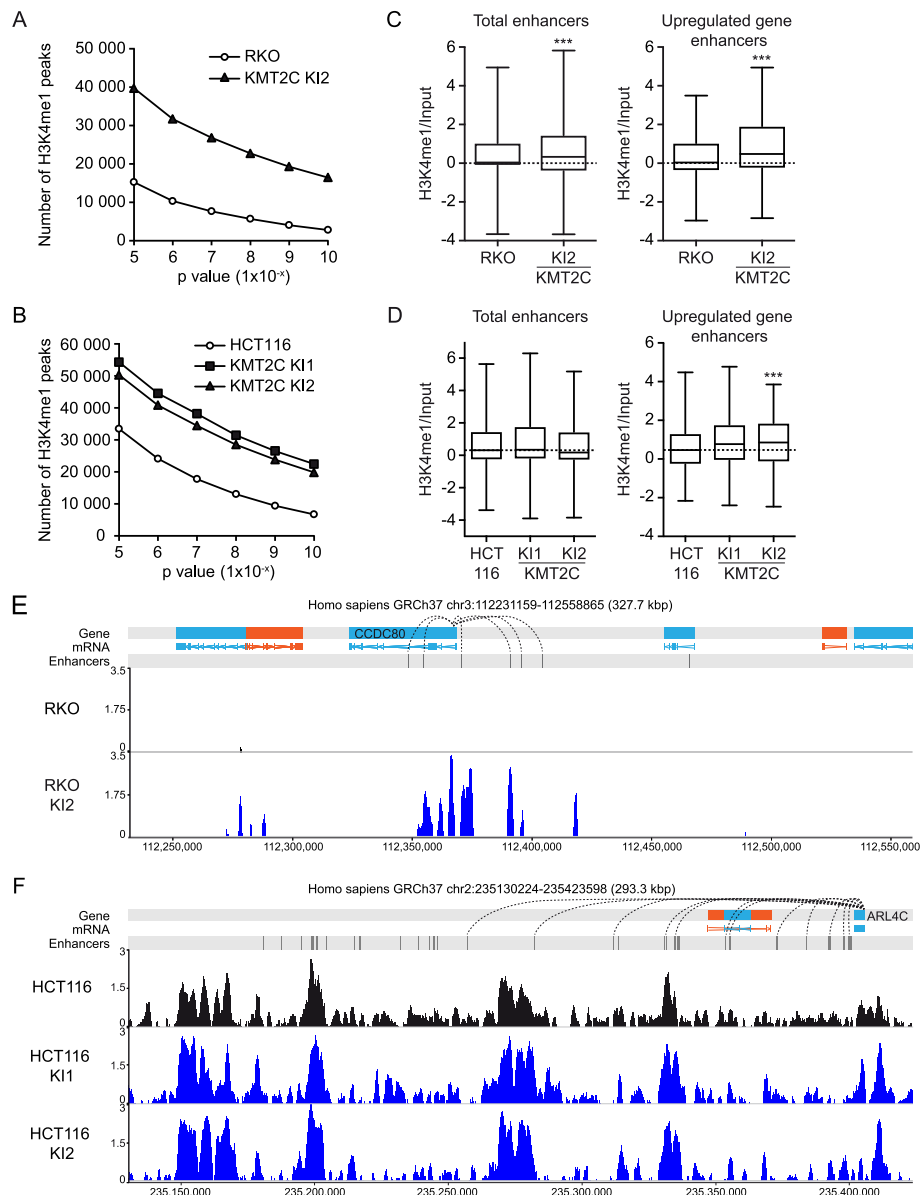
To profile changes in H3K4me1 upon restoration of *KMT2C* expression, we subjected cells to chromatin immunoprecipitation (ChIP) sequencing. We observed a robust genome-wide H3K4me1 profile in HCT116 despite inactivation of *KMT2C*, consistent with a previously

published study demonstrating *KMT2D* dependent monomethylation in these cells [30]. In contrast, RKO showed a weaker and flatter H3K4me1 ChIP-seq profile with 2.2–2.4-fold fewer peaks than HCT116, depending on peak detection stringency ( $p = 1 \times 10^{-5} - 1 \times 10^{-10}$ ) (Additional File 1, Fig. S4a). *KMT2C* is predominantly reported to affect gene regulation through H3K4me1 at gene enhancers [30–32]. The weaker H3K4me1 enrichment detected in RKO cells relative to HCT116 was a global phenomenon occurring at both intestinal enhancers and ubiquitous enhancers ( $n = 447$ ,  $p < 0.001$  and  $n = 505$ ,  $p = 0.0066$ , respectively) (Fig. 2a–c).

Upon restoration of *KMT2C*, a 2.6–5.9 and 1.5–3.4-fold increase in H3K4me1 enriched regions was detected, respectively, for RKO and HCT116 *KMT2C<sup>insG</sup>* cells (Fig. 3a–b). The *KMT2C* restoration-induced enrichment was stronger in RKO *KMT2C<sup>insG</sup>* cells, which have a weaker baseline level of H3K4me1 in parental RKO cells. Global levels of H3K4me1 in *KMT2C<sup>insG</sup>* cells and parental HCT116 and RKO cells were also estimated by immunoblotting (Additional File 1, Fig. S4b and c), independently confirming the pattern for increased H3K4me1 in *KMT2C<sup>insG</sup>* cells, however with increases of lower magnitude than estimated by ChIP sequencing, and only at a significant level in RKO *KMT2C<sup>insG</sup>* clone K12. The immunoblot also validated the lower baseline H3K4me1 level in RKO cells. Using enhancers with predicted enhancer-promoter pairs, obtained from the



**Fig. 2** Genome-wide H3K4me1 profiling reveals differences between *KMT2C* deficient MSI CRC cell lines HCT116 and RKO. **a–b** H3K4me1 enrichment at **a** intestinal and **b** ubiquitous enhancers (\*\* $P < 0.01$ , \*\*\* $P < 0.001$ ). Dashed line marks 0. **c** Genome browser shot showing H3K4me1 ChIP-seq enrichment across human enhancers (gray dashed boxes), including an intestinal enhancer (black dashed box)



**Fig. 3** Restored *KMT2C* expression increases H3K4me1 at enhancers of upregulated genes in RKO and HCT116 cells. **a–b** ChIP-seq detection of H3K4me1 enriched regions at different stringencies of peak detection in RKO *KMT2C<sup>insG</sup>* KI2 clone (**a**) and HCT116 *KMT2C<sup>insG</sup>* KI1 and KI2 clones (**b**). **c–d** H3K4me1 enrichment at global human genome enhancers and at enhancers of upregulated genes in (**c**) RKO parental and *KMT2C<sup>insG</sup>* KI2 clone and (**d**) HCT116 parental and *KMT2C<sup>insG</sup>* clones KI1 and KI2 ( $***P < 0.001$ ). Dashed lines mark the median of the parental cells in **c–d**. **e** H3K4me1 signal at the *CCDC80* gene, which was upregulated in RKO *KMT2C<sup>insG</sup>* KI2 cells. **f** Detection of H3K4me1 at the chromosomal region of the *ARL4C* gene which was upregulated in HCT116 *KMT2C<sup>insG</sup>* KI1 and KI2 clones. Enhancer-promoter associations are marked by dotted lines in **e–f**

FANTOM consortium, we measured H3K4me1 in the total FANTOM enhancer data set ( $n = 43,011$ ) and at enhancers of genes showing upregulated expression in RKO or HCT116 *KMT2C<sup>insG</sup>* cells, respectively (Fig. 3c–d). For RKO cells, *KMT2C* restoration resulted in increased H3K4me1 levels globally across all enhancers in the investigated *KMT2C<sup>insG</sup>* clone (Fig. 3c;  $p < 0.001$ ). This increase was also observed at enhancers of upregulated genes ( $n = 680$ ,  $p < 0.001$ ). For HCT116 cells, H3K4me1 was increased at

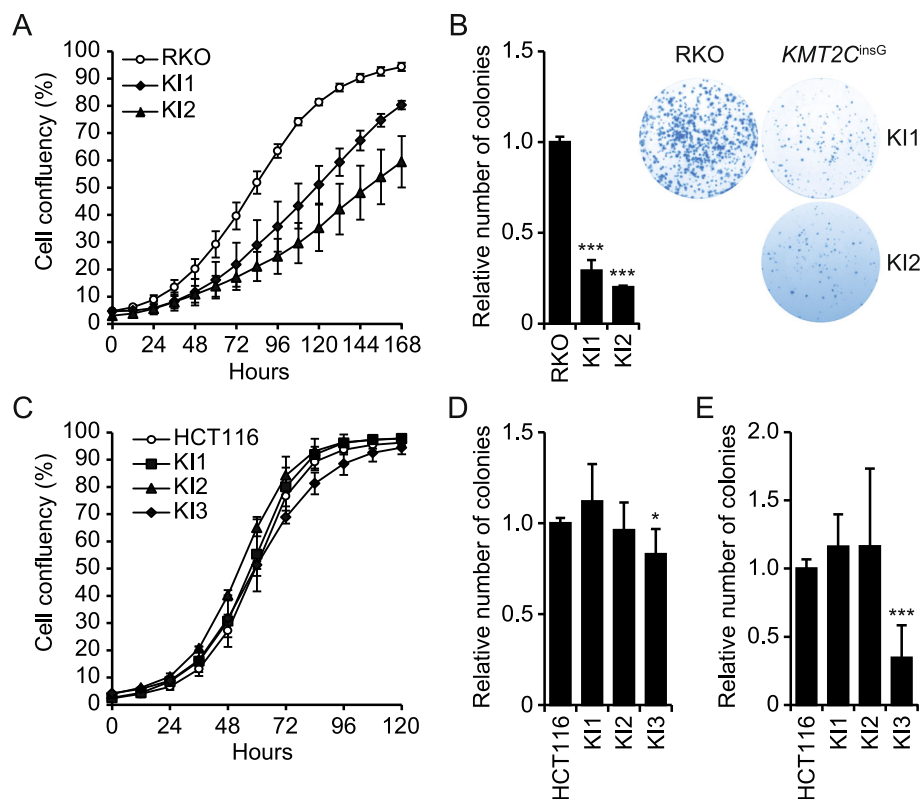
enhancers of upregulated genes upon *KMT2C* restoration ( $n = 178$ ) but was not apparent at enhancers globally (Fig. 3d). Examination of the H3K4me1 profiles showed that H3K4me1 was enriched at previously unmethylated enhancers in RKO cells whereas HCT116 cells had H3K4me1 already prior to *KMT2C* restoration, exemplified by enhancers of the *CCDC80* gene, which was 7-fold upregulated in RKO *KMT2C<sup>insG</sup>* clone KI2 (Fig. 3e), and *ARL4C*, which



was 4-fold upregulated in HCT116 *KMT2C<sup>insG</sup>* clone KI3 (Fig. 3f). We examined if the additional H3K4me1 peaks detected by restoration of *KMT2C* expression in HCT116 cells reflected reinforced and more easily detected signal at sites with some H3K4me1 already present or occurred at truly new sites of H3K4me1 deposition. Consistent with the model proposed by Hu et al. [30], we found that regions that acquired H3K4me1 in *KMT2C<sup>insG</sup>* cells showed some degree of H3K4me1 deposition in parental HCT116 cells (Additional File 1, Fig. S5). In summary, these data suggest that restoration of *KMT2C* expression affected H3K4me1 levels at enhancers. Furthermore, comparison between the two cell lines showed that genome-wide H3K4me1 was more dependent on *KMT2C* expression in RKO cells whereas the modest increase of H3K4me1 levels detected at the enhancers of upregulated genes in HCT116 *KMT2C<sup>insG</sup>* clones suggests that the HCT116 cell line is less dependent on *KMT2C* activity to maintain H3K4me1 marks.

### Characterization of cell growth

To investigate if restoration of *KMT2C* expression affects cell growth phenotypes, we monitored cell proliferation. For RKO cells, both the restored clones showed decreased cell proliferation rate compared to parental RKO, with clones KI1 and KI2 growing ~ 10% and 30% slower than parental RKO cells, respectively (Fig. 4a). The stronger growth phenotype in clone KI2 was also accompanied by a morphological change, with cells becoming enlarged and flattened compared to the parental RKO cells (Additional File 1, Fig. S6). The altered growth phenotype of RKO *KMT2C<sup>insG</sup>* clones was further confirmed in colony formation assays where both RKO *KMT2C<sup>insG</sup>* clones formed markedly fewer colonies than parental RKO cells (Fig. 4b). In contrast, HCT116 *KMT2C<sup>insG</sup>* clones did not show altered growth rate compared to parental HCT116 cells (Fig. 4c). In addition, only one of the three HCT116 *KMT2C<sup>insG</sup>*



**Fig. 4** Restoration of *KMT2C* expression decreases growth of RKO cells. **a** Growth curve for RKO *KMT2C<sup>insG</sup>* clones (KI1-2). Cell confluency was tracked by daily imaging for 1 week. Bars represent standard deviation for three replicate wells per sample (KI1,  $P < 0.05$  and KI2  $P < 0.01$  for time points 48–168 h (Student's  $t$ -test)). **b** Colony formation assay showing ability to form colonies from single cells for RKO *KMT2C<sup>insG</sup>* clones. Triplicate sample wells were normalized to the average relative number of colonies for parental RKO cells. **c** Growth curve for HCT116 *KMT2C<sup>insG</sup>* clones (KI1-3). Cell confluency was traced by daily imaging over the course of 5 days. Bars represent standard deviation for three replicate wells per sample. **d** Colony formation assay for HCT116 parental cells and *KMT2C<sup>insG</sup>* clones. **e** Ability for anchorage independent-growth measured through the soft agar assay for HCT116 parental cells and *KMT2C<sup>insG</sup>* clones. For **d-e**, all wells were normalized to the average for the respective parental HCT116 cells for each plate, and the average relative number of colonies from two independent experiments is shown for each cell line. Error bars, SD. Two tailed Student's  $t$ -test; \* $P < 0.05$ , \*\*\* $P < 0.001$

clones (KI3) showed a trend of reduced colony forming ability in the colony formation assay (Fig. 4d). When investigating the anchorage independent growth by soft agar assays for HCT116 cells, we observed a decreased ability to grow in soft agar for the HCT116 *KMT2C<sup>insG</sup>* KI3 clone (Fig. 4e). As we could not obtain reproducible growth of RKO cells in soft agar, we could not assess RKO cell growth by this assay. In conclusion, correction of *KMT2C* by gene editing had different effects in the two targeted cell lines. RKO cells were strongly affected by restoration of this histone methyltransferase, showing reduced growth rate, morphological changes, and compromised ability to form colonies. In contrast, with the exception of one clone that showed a tendency for decreased colony formation ability, we did not observe any alterations of the growth of HCT116 cells.

#### Knockdown of *TSPAN1* and *PRSS23* in colorectal cancer cells

Next, we investigated the genes found regulated by restored *KMT2C* activity in RKO and HCT116 cells. Based on expression data and literature and database searches, we focused on the upregulated genes *PRSS23* and *TSPAN1*. For *PRSS23*, overexpression (average 4.9%, median 4.7% of samples) is indicated for several cancer types but no frequent mutation or copy number aberration (COSMIC database [33], v86, accessed Nov. 8 2018). The serine protease is evolutionary conserved among vertebrates and suggested involved in endothelial-mesenchymal transition regulation during cardiac development [34] and in ovarian follicle regulation [35]. The small transmembrane glycoprotein *TSPAN1* was previously shown downregulated in HCT116 *KMT2D* KO cells [36]. Downregulation of *TSPAN1* has been associated with metastasis in prostate cancer [37], and knockdown inhibited the proliferation of CRC cells and their ability to migrate in an in vitro invasion assay [38]. Enrichment of H3K4me1 was observed at the enhancer of *PRSS23* in the *KMT2C<sup>insG</sup>* KI clones for both RKO and HCT116 cells but with stronger enrichment profile in RKO cells. For *TSPAN1*, H3K4me1 enrichment was found at the gene body (Additional File 1, Fig. S7). We performed stable *PRSS23* and *TSPAN1* knockdown in the CRC cell lines DLD1 and HT-29, which do not have truncating *KMT2C* or *KMT2D* mutations (COSMIC [33], v86, 11 Oct. 2018), to study how CRC cells respond to silencing of these genes (Additional File 1, Fig. S8). Knockdown of *PRSS23* or *TSPAN1* did not affect DLD1 or HT-29 cell growth rate and growth in soft agar, or the ability for DLD1 cells to form colonies from single cells. For HT-29, cell colonies attached to poorly to the surface to be assayed. Finally, because of possible involvement with cell migration [34, 38], we performed scratch wound healing assays, where both knockdown of *PRSS23* and *TSPAN1* showed tendency to increase the wound closing capability of DLD1 cells, but

not that of HT-29 cells. In summary, these data do not support a major role of *PRSS23* or *TSPAN1* upregulation in the growth phenotype detected in RKO *KMT2C<sup>insG</sup>* cells but indicate that *PRSS23* or *TSPAN1* downregulation can affect CRC cell migration.

#### Discussion

The H3K4 monomethylase *KMT2C* is a frequent target of mutations in CRC and other cancer types [1], and studies have demonstrated different effects of different mutations on *KMT2C* function [22, 39], suggesting the need for further investigation of *KMT2C* in tumorigenesis. Here, we restored wild type function of *KMT2C* in two CRC cell lines with homozygous *KMT2C* c.8390delA frameshift mutation, which is frequent in MSI cancers. No or low expression of truncated protein has been reported for this mutation previously [25, 26], and accordingly *KMT2C* expression in the parental cells was low but detectable. The general tendency for slower growth or decreased colony forming ability detected for the respective RKO and HCT116 derived *KMT2C<sup>insG</sup>* KI cell lines was consistent with the suggested tumor suppressor role of *KMT2C* (reviewed in [5, 6]). The engineered cell lines, however, displayed changes of different magnitude, as was observed also for changes in gene expression and histone methylation patterns.

There was an overall increase in H3K4me1 following restoration of a single *KMT2C* allele, but the RKO and HCT116 cell lines varied in the amount of H3K4me1 present in the parental *KMT2C* deficient cells. Previously, from study of HCT116 cells (which express *KMT2D* but have *KMT2C* homozygous mutation) where *KMT2D* was knocked out, it has been proposed that *KMT2C* and *KMT2D* cooperate by targeting the same genomic regions for deposition of H3K4me1 in CRC cells [30]. In agreement with this functional redundancy model, it may be that HCT116 cells show weak effects on H3K4me1 enrichment upon *KMT2C* restoration because of having active *KMT2D* and a robust genome-wide profile of H3K4me1 at baseline. This fits well with the observation that in HCT116 the increased H3K4me1 enrichment in *KMT2C<sup>insG</sup>* clones was mainly located at sites already marked by H3K4me1 in parental cells. In RKO, sequence data reported in the CCLE [40] and COSMIC [33] (v86) databases (accessed 9 Oct. 2018) indicate the presence of *KMT2D* frameshift and/or nonsense mutations. Furthermore, our RNA-seq data report lowers expression of *KMT2D* in RKO cells than in HCT116 cells (ave FPKM 1.9 for RKO vs 3.0 for HCT116). Inactivity of *KMT2D* may thus contribute to the lower baseline H3K4me1 level and the greater enhancement of H3K4me1 following restoration of *KMT2C* observed in RKO cells. Accordingly, more pronounced effects on gene expression and cell growth were also observed in RKO cells.

Although KMT2C is known to interact with certain sequence specific transcription factors, such as FOXA1 [41], based on our observations, its monomethylase activity is not highly targeted at specific genes as deposition of H3K4me1 in CRC cells where KMT2C is restored is found globally at enhancer regions, not at a specific sub-set of enhancers. Even though we detected variation in regulated gene sets between cell lines and individual KI clones, the overlap in DEGs between *KMT2C*<sup>insG</sup> clones and the respective parental cell line was greater than expected if differential regulation was random, particularly in RKO cells. Furthermore, analysis of gene functions showed similarities between the differentially expressed gene sets enriched in RKO and HCT116-derived clones, identifying cell signaling pathways with well-known association to cancer as affected by *KMT2C* expression. We suggest that this could mean that *KMT2C* loss leads to general destabilization of the cells epigenomic transcriptional control rather than leading to specific downregulation of only certain target genes. This fits well with a previously proposed model for the contribution of inactivating *KMT2C* and *KMT2D* mutations to cancer development, where several mechanisms that reduce enhancer activation alter tumor suppressing gene expression [42]. Some of the variation in differential expression that was observed between clones could potentially stem from the clonal selection procedures used to isolate the targeted clones [43], whereas there is low risk for simultaneous off-target integrations of the rAAV vector with correct rAAV-mediated gene targeting [44]. Interestingly, given that *KMT2C* has been found to regulate estrogen-dependent HOX family gene transcription in chorocarcinoma placenta cells [45–47], and identified an estrogen receptor  $\alpha$  (ER $\alpha$ ) coactivator in breast cancer cells [48], we observed differential regulation of estrogen response genes. Among these were PRSS23, which was upregulated in all *KMT2C*<sup>insG</sup> clones and for which H3K4me1 was enriched at enhancers. Levels of PRSS23 correlate with ER $\alpha$  levels in breast cancer and are upregulated by ER $\alpha$  promoter binding [49]. Our data thus suggest that *KMT2C* may be important for regulation of estrogen signaling in CRC tumorigenesis, but unlike previous reports in breast cancer [49], we saw minor effects of PRSS23 knockdown on CRC cell proliferation, suggesting the possibility that other *KMT2C* regulated genes are more important for promotion of CRC tumorigenesis following loss of *KMT2C*.

Finally, in this paper we have studied a frameshift mutation in exon 38, which precedes the SET domain, FYRC and FYRN domains as well as one PHD-like domain. We focused on investigating effects on enhancer H3K4me1 changes and differential gene expression. From studies of catalytic dead vs knocked out *KMT2C* in mouse embryonic stem cells, it has been suggested

that not all functions of *KMT2C* and *KMT2D* are dependent on the catalytic function of the SET domain [50]. Thus, we cannot exclude the possibility that *KMT2C* c.8390delA mutations contribute to cancer development through other, yet to be defined, mechanisms.

## Conclusions

This is the first study where an endogenous *KMT2C* frameshift mutation is functionally restored in CRC cells. Heterozygous *KMT2C* knock in led to a general tendency of reduced colony formation ability. Gene expression profiling in *KMT2C*<sup>insG</sup> cells did not point to altered expression of specific gene sets but rather to a general transcriptional dysregulation that affected processes or pathways that are known players in cancer development, lending support to this as a plausible mechanism for selection of *KMT2C* mutations during cancer development. The striking difference in H3K4me1 genome-wide profile between HCT116 and RKO cell lines demonstrates how individual mutations can yield different molecular phenotypes. Molecularly, this is not particularly surprising due to the redundancy of molecular pathways in mammalian cells whereby most genes are regulated by several transcription factors, coactivators, and repressors. Thus, when one transcriptional coactivator such as *KMT2C* is lost, it does not lead to a complete failure or suppression of gene expression or even of H3K4me1 deposition. Future studies should address if in clinical settings *KMT2C*-deficient colorectal cancers have distinct H3K4me1 phenotypes. This may pave the way for personalized treatments whereby patients displaying *KMT2C* loss and aberrant H3K4me1 epigenomic profiles may respond well to treatments that restore H3K4me1 while *KMT2C* wild type CRC patients with normal H3K4me1 profiles would likely be unaffected.

## Materials and methods

### *KMT2C* targeting construct

Primers used for PCR are listed in Additional File 2, Table S8. The *KMT2C*<sup>insG</sup>-rAAV gene targeting vector was designed to contain a guanine instead of an adenosine in the fourth position of the A(9) repeat in exon 38. The 5' and 3' homology arm (HA) sequences were amplified from HCT116 gDNA with Phusion High-Fidelity DNA polymerase (Finnzyme, Espoo, Finland) with *attB* tagged primers 1–4 and gel purified. The HA sequences and the selection cassette were cloned into the pAAV-Dest destination vector using the Gateway system (Invitrogen, Carlsbad, CA, USA) as described [51]. Briefly, 100 ng each of the amplified 5' (HA1) and 3' (HA2) HAs were recombined with 150 ng of pDONR<sup>TM</sup>P1-P2 or pDONR<sup>TM</sup>P3-P4, respectively, using BP Clonase II,



generating pEntry-HA1 and pEntry-HA2. Positive clones were identified by PCR with primers 5–6. The G nucleotide was inserted within the A(9) repeat of the 3' HA sequence inserted in pDONR<sup>TM</sup>P3-P4 using the QuickChange II mutagenesis kit (Stratagene, San Diego, CA, USA) with primers 7–8, and the insertion was confirmed by Sanger sequencing. Next, 10 fmol each of pEntry-HA1, the selection-cassette encoding entry vector pBUOY.SA.IRES.Neo.pA, and pEntry-HA2 were recombined with 15 fmol pAAV-Dest vector using LR Clonase II. The presence and orientation of fragments was confirmed by PCR and Sanger sequencing using LR screening primers 9–12. The *KMT2CinsG*-rAAV virus particles were produced by transfection of 70% confluent AAV-293 cells (Stratagene, cultured in DMEM, 10% FBS and 1% penicillin/streptomycin (PEST) (all from Gibco/Life Technologies, Carlsbad, CA, USA)) with Lipofectamine (Invitrogen) and 5 µg each of pAAV-RC, pHELPER (Stratagene), and the targeting construct, with harvesting of cell lysate after 48 h as described [52].

#### Cell lines and *KMT2C* targeting

The human colorectal cancer cell lines RKO (CRL-2577, ATCC, Manassas, VA, USA) and HCT116 (CCL-247, ATCC) were cultured in McCoy's 5A (modified) medium (Gibco) with 10% FBS and 1% PEST. Cells were transfected with *KMT2CinsG*-rAAV as described [52] and selected for 2 weeks at limiting dilution in the presence of 0.8 mg/ml (RKO) or 0.4 mg/ml (HCT116) Geneticin (Gibco). Single-cell clones with site-specific integration of the targeting vector were identified by PCR (primers 13–18), and the insertion of a G in the A9 repeat, as well as the integrity of the sequence surrounding the insertion site, was confirmed by Sanger sequencing. The IRES *neo* selection cassette was removed by Ad-Cre virus (Vector Biolabs, Malvern, PA, USA) infection as described [52], and single-cell clones identified by PCR (primers 19–20) to lack selection cassette were verified by their inability to grow in the presence of Geneticin. The *KMT2CinsG* clones were validated by sequencing of the exon 38 A(9) repeat, and the inserted G was verified to be expressed by Sanger sequencing of RT-PCR products. Parental and isogenic cell lines were authenticated by STR profiling at ATCC (2017) and verified free from mycoplasma using the MycoAlert mycoplasma detection kit (Lonza, Basel, Switzerland) prior to RNA and ChIP sequencing.

#### Generation of stable *PRSS23* and *TSPAN1* knockdown cell lines by lentiviral transduction of shRNAs

The cell lines DLD1 and HT-29 (HTB-38, ATCC) were cultured as described for HCT116 and RKO above. Lentiviral transductions were performed with SMARTvector lentiviral particles (Horizon/Dharmacon) (Additional

File 2, Table S9). The day before transduction, 50,000 cells were plated in each well of a 24-well plate. The next day, the plating medium was removed, and 250 µl of virus diluted in normal growth medium with 7.5 mg/ml Sequa-Brene (Sigma Aldrich) was added to each well. After 24 h incubation at 37 °C, virus-containing media were replaced with fresh complete growth media. After 48 h incubation, the pool of cells containing the shRNA of interest was selected by addition of 1 µg/ml puromycin (Gibco) to the growth medium. After expansion for 2–3 passages under selection with puromycin, knockdown efficiencies were determined by qPCR.

#### Cell morphology, growth, and migration

Growth curves were generated by culturing of cells in plates imaged in real-time inside an IncuCyte HD instrument, recording cell confluency every 12 h. For colony formation assays, 500 cells (HCT116, DLD1) or 1000 cells (RKO) plated in triplicate wells of 6-well plates were stained with 5% methylene blue in methanol 10–14 days after seeding. Soft agar assays were performed in 6-well plates with a 1-ml top layer of 0.3% (RKO, DLD1, HT-29) or 0.35–0.4% (HCT116) low gelling temperature agarose (A9414, Sigma Aldrich, St. Louis, MO, USA) with 1000–1250 resuspended cells, which was overlaid onto 1 ml of a bottom layer of solid 1% (RKO) or 0.8% (HCT116, DLD1, HT-29) low gelling temperature agarose. After solidification, the top layer was covered with 1 ml of fresh culture medium. Colonies were stained with 0.05% crystal violet after 2–3 weeks. For study of cell migration, confluent cell monolayers in 6 well plates were scratched with a 200-µl pipet tip, washed with fresh medium at least three times, and imaged at 0 h and after 72 h culturing. The TScratch software [53] was used to calculate the open wound area at defined spatial areas of the scratch.

#### RNA sequencing

The integrity and concentration of total RNA, isolated using the RNeasy mini kit with on-column DNase digestion (Qiagen, Hilden, Germany), was determined on a Bioanalyzer 2100 instrument (Agilent, Santa Clara, CA, USA) using the RNA 6000 nano chip. Samples were sequenced on the Ion Proton system (Ion Torrent/Life Technologies) at the Uppsala Genome Center at SciLife Lab NGI Uppsala. The sequencing reads were aligned to the human genome hg19 assembly [54] using the Tophat2 software (version 2.0.4) [55]. Quantification of gene expression levels and identification of differential expression were performed using the Cufflinks software (version 2.1.1) [56]. Among the most deregulated genes, 6 were chosen for RT-qPCR validation for each cell line, respectively, although one primer pair failed for the RKO cell line. Overlap with the Hallmarks gene set [57]

using the Molecular Signatures Database (MSigDB) [58, 59] was used for functional classification and enrichment analysis of regulated genes.

#### RT-qPCR

Primers for qPCR are listed in Additional File 2, Table S10. All primers were evaluated by construction of five step 1:5 dilution standard curves. By stability evaluation using the Cotton EST database RefFinder tool [60] across all independent isogenic clones and the respective parental RKO and HCT116 cell lines, TBP and HPRT1 were selected as endogenous reference genes. The cDNA was synthesized from total RNA using the Maxima H minus First Strand cDNA synthesis kit (Thermo Scientific) using the supplied random primers. Technical triplicate 20  $\mu$ l qPCR reaction with 1 $\times$  Maxima SYBR Green/ROX qPCR Master Mix (Thermo Scientific) and 0.3  $\mu$ M of each primer were run on the StepOne Real-Time PCR system (Applied Biosystems, Foster City, CA, USA). Each assay included no-template controls, and each RNA sample was assessed for gDNA contamination by cDNA synthesis reactions where reverse transcriptase was omitted. For assessment of lentiviral knockdown efficiencies for *TSPAN1* and *PRSS23*, all cDNAs were prepared directly from cell lysates using the Cell-to-Ct kit (Life Technologies), and specific TaqMan assays (ThermoFisher Scientific; Additional File 2, Table S7) were used to measure gene expression levels.

#### ChIP sequencing

HCT116 cells and RKO cells were fixed in 1% paraformaldehyde for 8 min, and ChIP was performed using the iDeal ChIP-seq kit (Diagenode, Liege, Belgium) according to the manufacturer's instructions. Chromatin extracts from one million cells were sonicated to yield 100 to 500 bp fragments using a Bioruptor Pico sonication device (Diagenode) on high power for two rounds of 6 cycles with 30 s on 30 s off at 4  $^{\circ}$ C. Fragmentation was verified with agarose gel electrophoresis. ChIP-seq was performed using 1  $\mu$ g of H3K4me1 antibody (ab8895, Abcam) per ChIP. After reverse cross-linking overnight at 65  $^{\circ}$ C, DNA was extracted using the Qiagen MinElute PCR purification kit in 15  $\mu$ l of elution buffer. Libraries were prepared for sequencing using the NEBNext<sup>®</sup> ChIP-Seq Library Prep Reagent Set for Illumina<sup>®</sup> and sequenced on a HiSeq 2500 System (Illumina).

#### Analysis of ChIP-seq

ChIP-seq data in HCT116 and RKO samples were aligned using Bowtie 2. BAM files were uploaded into Seqmonk for further analysis [61]. Peak calling was performed with MACS peak builder at different *p* values as indicated. Ubiquitous and colorectal specific enhancers were obtained from the FANTOM 5 project [62]. FANTOM5 enhancers were identified by bidirectional transcription using the

capped analysis of gene expression technique [63]. H3K4me1 signal was quantified by calculating the log<sub>2</sub> of the reads per million for each sample relative to input (ChIP/Input). Genome browser shots were produced using Seqmonk by detecting the enrichment over input across selected regions with smoothing for adjacent 5 probes. To examine the enhancers of differentially expressed genes, we identified enhancers from the FANTOM5 project that are predicted to regulate each of the differentially expressed genes. Differences between means were examined by the Students *t*-test or ANOVA as indicated.

#### Western blot

Histone extracts were prepared from  $1 \times 10^7$  cells using the Histone Extraction kit (Abcam, ab113476). Equal amounts of proteins were separated on NuPAGE 4–12% Bis-Tris protein gels, transferred onto nitrocellulose membranes using the iBlot system (Invitrogen), and probed with anti-H3K4me1 (ab8895, Abcam, 1:500) and anti-H3 (ab1719, Abcam, 1:3000). Immunoreactive proteins were visualized using SuperSignal West Femto Chemiluminescent Substrate (Thermo Scientific) and Amersham Imager 680 (GE Healthcare). The ratio of H3K4me1 to total H3 was quantified by densitometric analysis using ImageJ [64].

#### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13148-020-00863-z>.

**Additional file 1: Figure S1.** Validation of *KMT2C* gene targeting in knock-in clones. **Figure S2.** RNA sequencing data for RKO and HCT116 *KMT2C<sup>insG</sup>* clones were validated by RT-qPCR. **Figure S3.** The genes ANK1, PRSS23, SAMD9 and TSPAN1 were upregulated in *KMT2C<sup>insG</sup>* HCT116 and RKO cells. **Figure S4.** Detection of H3K4me1 in HCT116 and RKO cells. **Figure S5.** Detection of H3K4me1 enriched regions is enhanced at sites of existing H3K4me1 in HCT116 *KMT2C<sup>insG</sup>* KI clones. **Figure S6.** Restoration of *KMT2C* expression affects the morphology of RKO cells. **Figure S7.** Level of H3K4me1 at genomic regions of *PRSS23* and *TSPAN1* genes. **Figure S8.** Knockdown of *TSPAN1* and *PRSS23* in colorectal cancer cells has minor effects on cell growth.

**Additional file 2: Table S1.** Summary of genes differentially regulated in RKO and HCT116 following restoration of *KMT2C* expression. **Table S2.** Differentially regulated genes in RKO and HCT116 *KMT2C<sup>insG</sup>* clones observed to overlap or expected to overlap by chance if regulation is random. **Table S3.** Upregulated genes in RKO and HCT116 *KMT2C<sup>insG</sup>* clones observed to overlap or expected to overlap by chance if regulation is random. **Table S8.** PCR primer sequences. **Table S9.** The shRNA lentiviruses and TaqMan probes used for stable knockdown cell line generation. **Table S10.** Primers for RT-qPCR with SYBR Green detection.

**Additional file 3: Table S4.** Genes differentially expressed more than 1.5 log<sub>2</sub> fold in RKO cells following restoration of *KMT2C* expression. **Table S5.** Genes differentially expressed more than 1.5 log<sub>2</sub> fold in HCT116 cells following restoration of *KMT2C* expression. **Table S6.** Overlap analysis with the MSigDB Hallmarks gene set for genes differentially regulated >1.5 log<sub>2</sub> fold by restoration of *KMT2C* expression in RKO and HCT116 cells. **Table S7.** Overlap analysis with the MSigDB Hallmarks gene set for genes upregulated >1.5 log<sub>2</sub> fold by restoration of *KMT2C* expression in RKO and HCT116 cells.

**Additional file 4.** Uncropped gels for Figure S1

### Abbreviations

ChIP: Chromatin immunoprecipitation; DEG: Differentially regulated genes; H3K4me1: Histone 3 lysine 4 monomethylation; KMT2: Histone-lysine N-methyltransferase 2; MLL: Mixed lineage leukemia; MSI: Microsatellite instability; NMD: Nonsense-mediated mRNA decay; rAAV: Recombinant adeno-associated virus

### Acknowledgements

We thank Dr. Joakim Ekström for calculating the statistical estimate of expected overlapping DEGs and Maria Karoutsou for technical assistance.

### Authors' contributions

TS and TP conceived the study. TP and KH made the targeting construct. CL and NP made the KI clones, and CL performed and analyzed experiments with support from TD, MAA, SK, and TS, with exception of ChIP studies which were performed and analyzed by LC and LS with support from KE. shRNA cell lines were made by SK. The RNA seq data was processed by LH, and further analyzed by CL. CL, LC, LS, KE, and TS wrote the manuscript; all authors participated in revision and approval of its publication.

### Funding

Work in the TS laboratory was funded by grants from the Swedish Foundation for Strategic Research (RBA08-0114 and F06-0050) and the Swedish Cancer Society (2012/834). CL was funded by postdoctoral grant CAN 2012/1235 from the Swedish Cancer Society. Work in the KE laboratory was supported by the Swedish Research Council and the Swedish Cancer Society. The funding bodies had no role in the design of the study, in collection, analysis, and interpretation of data, or in writing the manuscript. Open access funding provided by Uppsala University.

### Availability of data and materials

The RNA sequencing and ChIP-seq datasets generated and analyzed during this study are available in the NCBI GEO data repository [65] with accession numbers GSE131507 [66] and GSE131755 [67], respectively. All additional data generated and/or analyzed during this study are included in this published article and its supplementary information files.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests with the contents of this article.

### Author details

<sup>1</sup>Department of Immunology, Genetics and Pathology, Uppsala University, Rudbeck Laboratory, SE-751 85 Uppsala, Sweden. <sup>2</sup>Department of Biosciences and Nutrition, NEO, Karolinska Institute, SE-141 83 Huddinge, Sweden. <sup>3</sup>Present address: School of Biological Sciences, University of the Punjab, Lahore, Pakistan. <sup>4</sup>Present address: The Novo Nordisk Foundation for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800 Kgs. Lyngby, Denmark.

Received: 28 May 2019 Accepted: 14 May 2020

Published online: 29 May 2020

### References

- Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. *Nature*. 2013; 502(7471):333–9.
- Shen H, Laird PW. Interplay between the cancer genome and epigenome. *Cell*. 2013;153(1):38–55.
- Shinsky SA, Monteith KE, Viggiano S, Cosgrove MS. Biochemical reconstitution and phylogenetic comparison of human SET1 family core complexes involved in histone methylation. *J Biol Chem*. 2015;290(10): 6361–75.
- Rao RC, Dou Y. Hijacked in cancer: the KMT2 (MLL) family of methyltransferases. *Nat Rev Cancer*. 2015;15(6):334–46.
- Ford DJ, Dingwall AK. The cancer COMPASS: navigating the functions of MLL complexes in cancer. *Cancer Gene Ther*. 2015;208(5):178–91.
- Sze CC, Shilatifard A. MLL3/MLL4/COMPASS family on epigenetic regulation of enhancer function and cancer. *Cold Spring Harb Perspect Med*. 2016;1: 6(11).
- Figueiredo DL, do Amaral Rabello D, Zanette DL, Saggiaro FP, Mamede RC, Zago MA, et al. Assessment of MLL methyltransferase gene expression in larynx carcinoma. *Oncol Rep*. 2015;33(4):2017–22.
- Dawkins JB, Wang J, Maniati E, Heward JA, Koniali L, Kocher HM, et al. Reduced expression of histone methyltransferases KMT2C and KMT2D correlates with improved outcome in pancreatic ductal adenocarcinoma. *Cancer Res*. 2016;76(16):4861–71.
- Li B, Liu HY, Guo SH, Sun P, Gong FM, Jia BQ. Association of MLL3 expression with prognosis in gastric cancer. *Genet Mol Res*. 2014;13(3): 7513–8.
- Ibragimova I, Dulaimi E, Slifker MJ, Chen DY, Uzzo RG, Cairns P. A global profile of gene promoter methylation in treatment-naïve urothelial cancer. *Epigenetics*. 2014;9(5):760–73.
- Dohner K, Brown J, Hehmann U, Hetzel C, Stewart J, Lowther G, et al. Molecular cytogenetic characterization of a critical region in bands 7q35-q36 commonly deleted in malignant myeloid disorders. *Blood*. 1998;92(11): 4031–5.
- Ruault M, Brun ME, Ventura M, Roizes G, De Sario A. MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukaemia. *Gene*. 2002;284(1–2):73–81.
- Ashktorab H, Schaffer AA, Darempouran M, Smoot DT, Lee E, Brim H. Distinct genetic alterations in colorectal cancer. *PLoS One*. 2010;5(1):e8879.
- Sjjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science*. 2006;314(5797):268–74.
- Li B, Liu HY, Guo SH, Sun P, Gong FM, Jia BQ. A missense mutation (S3660L) in MLL3 gene influences risk of gastric cancer. *J BUON*. 2014 Apr-Jun;19(2):394–7.
- Li WD, Li QR, Xu SN, Wei FJ, Ye ZJ, Cheng JK, et al. Exome sequencing identifies an MLL3 gene germ line mutation in a pedigree of colorectal cancer and acute myeloid leukemia. *Blood*. 2013;121(8):1478–9.
- Sasaki MM, Skol AD, Bao R, Rhodes LV, Chambers R, Vokes EE, et al. Integrated genomic analysis suggests MLL3 is a novel candidate susceptibility gene for familial nasopharyngeal carcinoma. *Cancer Epidemiol Biomark Prev*. 2015;24(8):1222–8.
- Villacis RA, Miranda PM, Gomy I, Santos EM, Carraro DM, Achatz MI, et al. Contribution of rare germline copy number variations and common susceptibility loci in Lynch syndrome patients negative for mutations in the mismatch repair genes. *Int J Cancer*. 2016;138(8):1928–35.
- Mann KM, Ward JM, Yew CC, Kovoichich A, Dawson DW, Black MA, et al. Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. *Proc Natl Acad Sci U S A*. 2012;109(16):5934–41.
- March HN, Rust AG, Wright NA, ten Hoeve J, de Ridder J, Eldridge M, et al. Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nat Genet*. 2011;43(12):1202–9.
- Lee J, Kim DH, Lee S, Yang QH, Lee DK, Lee SK, et al. A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. *Proc Natl Acad Sci U S A*. 2009;106(21):8513–8.
- Wang L, Zhao Z, Ozark PA, Fantini D, Marshall SA, Rendleman EJ, et al. Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. *Nat Med*. 2018;24(6):758–69.
- Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for clonal carcinogenesis. *Nature*. 1993;363(6429):558–61.
- Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*. 2010;138(6):2073–87 e3.
- Shin N, You KT, Lee H, Kim WK, Song M, Choi HJ, et al. Identification of frequently mutated genes with relevance to nonsense mediated mRNA decay in the high microsatellite instability cancers. *Int J Cancer*. 2011; 128(12):2872–80.
- Watanabe Y, Castoro RJ, Kim HS, North B, Oikawa R, Hiraishi T, et al. Frequent alteration of MLL3 frameshift mutations in microsatellite deficient colorectal cancer. *PLoS One*. 2011;6(8):e23320.
- Je EM, Lee SH, Yoo NJ. Mutational and expression analysis of MLL genes in gastric and colorectal cancers with microsatellite instability. *Neoplasma*. 2013;60(2):188–95.

28. Pandzic T, Rendo V, Lim J, Larsson C, Larsson J, Stoimenov I, et al. Somatic PRDM2 c.4467delA mutations in colorectal cancers control histone methylation and tumor growth. *Oncotarget*. 2017;8(58):98646–59.
29. Ernst J, Kheradpour P, Mikkelson TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473(7345):43–9.
30. Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol Cell Biol*. 2013;33(23):4745–54.
31. Cheng J, Blum R, Bowman C, Hu D, Shilatifard A, Shen S, et al. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Mol Cell*. 2014;53(6):979–92.
32. Herz HM, Mohan M, Garruss AS, Liang K, Takahashi YH, Mickey K, et al. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian MLL3/MLL4. *Genes Dev*. 2012;26(23):2604–20.
33. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017;45(D1):D777–D83.
34. Chen IH, Wang HH, Hsieh YS, Huang WC, Yeh HI, Chuang YJ. PRSS23 is essential for the Snail-dependent endothelial-to-mesenchymal transition during valvulogenesis in zebrafish. *Cardiovasc Res*. 2013;97(3):443–53.
35. Wahlberg P, Nylander A, Ahlskog N, Liu K, Ny T. Expression and localization of the serine proteases high-temperature requirement factor A1, serine protease 23, and serine protease 35 in the mouse ovary. *Endocrinology*. 2008;149(10):5070–7.
36. Guo C, Chang CC, Wortham M, Chen LH, Kernagis DN, Qin X, et al. Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. *Proc Natl Acad Sci U S A*. 2012;109(43):17603–8.
37. Xu F, Gao Y, Wang Y, Pan J, Sha J, Shao X, et al. Decreased TSPAN1 promotes prostate cancer progression and is a marker for early biochemical recurrence after radical prostatectomy. *Oncotarget*. 2016;7(39):63294–305.
38. Chen L, Yuan D, Zhao R, Li H, Zhu J. Suppression of TSPAN1 by RNA interference inhibits proliferation and invasion of colon cancer cells in vitro. *Tumori*. 2010;96(5):744–50.
39. Weirich S, Kudithipudi S, Kycia I, Jeltsch A. Somatic cancer mutations in the MLL3-SET domain alter the catalytic properties of the enzyme. *Clin Epigenetics*. 2015;7:36.
40. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483(7391):603–7.
41. Jozwik KM, Chernukhin I, Serandour AA, Nagarajan S, Carroll JS. FOXA1 directs H3K4 monomethylation at enhancers via recruitment of the methyltransferase MLL3. *Cell Rep*. 2016;17(10):2715–23.
42. Herz HM, Hu D, Shilatifard A. Enhancer malfunction in cancer. *Mol Cell*. 2014 Mar 20;53(6):859–66.
43. Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, et al. Genetic and transcriptional evolution alters cancer cell line drug response. *Nature*. 2018;560(7718):325–30.
44. Khan IF, Hirata RK, Russell DW. AAV-mediated gene targeting methods for human cells. *Nat Protoc*. 2011;6(4):482–501.
45. Ansari KI, Hussain I, Kasiri S, Mandal SS. HOXC10 is overexpressed in breast cancer and transcriptionally regulated by estrogen via involvement of histone methylases MLL3 and MLL4. *J Mol Endocrinol*. 2012;48(1):61–75.
46. Ansari KI, Hussain I, Shrestha B, Kasiri S, Mandal SS. HOXC6 is transcriptionally regulated via coordination of MLL histone methylase and estrogen receptor in an estrogen environment. *J Mol Biol*. 2011; 411(2):334–49.
47. Ansari KI, Shrestha B, Hussain I, Kasiri S, Mandal SS. Histone methylases MLL1 and MLL3 coordinate with estrogen receptors in estrogen-mediated HOXB9 expression. *Biochemistry*. 2011;50(17):3517–27.
48. Gala K, Li Q, Sinha A, Razavi P, Dorso M, Sanchez-Vega F, et al. KMT2C mediates the estrogen dependence of breast cancer through regulation of ER $\alpha$  enhancer function. *Oncogene*. 2018;37(34):4692–710.
49. Chan HS, Chang SJ, Wang TY, Ko HJ, Lin YC, Lin KT, et al. Serine protease PRSS23 is upregulated by estrogen receptor  $\alpha$  and associated with proliferation of breast cancer cells. *PLoS One*. 2012;7(1):e30397.
50. Dorighi KM, Swigut T, Henriques T, Bhanu NV, Scruggs BS, Nady N, et al. MLL3 and MLL4 facilitate enhancer RNA synthesis and transcription from promoters independently of H3K4 monomethylation. *Mol Cell*. 2017;66(4):568–76 e4.
51. Stoimenov I, Ali MA, Pandzic T, Sjoblom T. Computational and molecular tools for scalable rAAV-mediated genome editing. *Nucleic Acids Res*. 2015; 43(5):e30.
52. Rago C, Vogelstein B, Bunz F. Genetic knockouts and knockins in human somatic cells. *Nat Protoc*. 2007;2(11):2734–46.
53. Geback T, Schulz MM, Koumoutsakos P, Detmar M. TScratch: a novel and simple software tool for automated analysis of monolayer wound healing assays. *Biotechniques*. 2009;46(4):265–74.
54. UCSC Genome Browser. <https://genome.ucsc.edu/>.
55. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14(4):R36.
56. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28(5):511–5.
57. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–25.
58. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
59. Molecular Signatures Database. <http://software.broadinstitute.org/gsea/msigdb/index.jsp>. Accessed 15 Feb 2016.
60. Xie F, Xiao P, Chen D, Xu L, Zhang B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol Biol*. 2012.
61. SeqMonk, Babraham Bioinformatics. <http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>.
62. FANTOM 5 Predefined enhancer data. <http://enhancer.binf.ku.dk/presets/>.
63. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507(7493):455–61.
64. Rasband WS. ImageJ, U.S. National Institutes of Health, Bethesda, Maryland, USA. <https://imagej.nih.gov/ij/>. (1997–2019).
65. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002; 30(1):207–10.
66. NCBI Gene Expression Omnibus data repository. <https://www.ncbi.nlm.nih.gov/geo/>. Accession number GSE131507.
67. NCBI Gene Expression Omnibus data repository. <https://www.ncbi.nlm.nih.gov/geo/>. Accession number GSE131755.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

