Regulation of ZNF217 oncogene by a co-amplified 20q13 enhancer cluster in gastric cancer

Colleen Russett
Department of Human Genetics
Faculty of Medicine and Health Sciences
McGill University, Montreal
August 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Human Genetics.

Abstract

The chromosome 20q13 region is recurrently amplified in many different cancer types, including breast and stomach cancer. This genomic region harbors the oncogenic zinc-finger transcription factor ZNF217, whose overexpression is associated with the immortalization of cancer cells, metastasis phenotypes, and poor patient prognosis. However, whether additional mechanisms, other than increased copy number, are responsible for ZNF217 hyperactivation remain to be elucidated. Several emerging paradigms suggest that copy number variants (CNVs) target regulatory non-coding regions, either on their own, or together with oncogenes. By analyzing publicly available array data from The Cancer Genome Atlas (TCGA), we discovered a significantly co-amplified noncoding region adjacent to ZNF217 across 1075 breast and 438 gastric cancer patients that harbors multiple active enhancer elements. To assess the importance of these co-selected enhancers in ZNF217 hyperactivation, we use HiChIP data from several breast and gastric cell lines to prioritize enhancers interacting with the ZNF217 promoter and employ CRISPRi to determine the contribution of these enhancers to ZNF217 hyperactivation. ZNF217 is currently undruggable, therefore identifying the mechanism(s) behind its hyperactivation, such as a dependency on specific regulatory elements, may lead to alternative therapeutic interventions for 20q13 amplified cancers.

Résumé

La région chromosomique 20q13 est amplifiée de façon récurrente dans de nombreux types de cancers, y compris le cancer du sein et de l'estomac. Cette région génomique héberge le facteur de transcription oncogénique, ZNF217, dont la surexpression est associée à l'immortalisation des cellules cancéreuses, la métastase et un pronostic pauvre. Cependant, il reste à élucider s'il y a des mécanismes supplémentaires—autre que l'augmentation du nombre de copies—responsables de l'hyperactivation de ZNF217. Plusieurs paradigmes émergents suggèrent que les variantes du nombre de copie (VNC) ciblent les régions régulatrices non-codantes, soit seules, soit en conjonction avec les oncogènes. En analysant les données de puces d'ADN de l'Atlas du Génome du Cancer (TCGA), nous avons découvert une région non-codante adjacente à ZNF217 qui est significativement coamplifiée chez 1075 patients atteints de cancer du sein et 438 du cancer gastrique et qui abrite plusieurs éléments amplificateurs actifs. Pour évaluer l'importance de ces éléments amplificateurs cosélectionnés dans l'hyperactivation de ZNF217, nous utilisons les données « HiChIP » de plusieurs lignées cellulaires mammaires et gastriques afin de prioriser les éléments amplificateurs qui interagissent avec le promoteur de ZNF217 et utilisons le CRISPRi pour déterminer la contribution de ces éléments à l'hyperactivation de ZNF217. Actuellement, il n'y a aucune drogue active contre le ZNF217. Par conséquent, identifier les mécanisme(s) à l'origine de son hyperactivation, telle qu'une dépendance à des éléments régulateurs spécifiques, pourrait mener à des interventions thérapeutiques alternatives pour les cancers amplifiés en 20q13.

Table of Contents

List of Abbrieviations

GEA	Gastro-esophageal adenocarcinoma
HDGC	Hereditary diffuse gastric cancer
PM	Peritoneal metastasis
TCGA	The Cancer Genome Atlas
TF	Transcription factor
GWAS	Genome-wide association study
SV	Structural variant
CNV	Copy number variation
SCNA	Somatic copy number alteration
ecDNA	Extra-chromosomal DNA
GISTIC	Genomic Identification of Significant Targets in Cancer
m^6A	N6-methyladenosine
EMT	Epithelial-mesenchymal transition
BET	Bromo- and extra-terminal domain
(CRISPR)i	(Clustered regularly interspersed short palindromic repeats) interference
HDAC	Histone deacetylase
NHEJ	Non-homologous end joining
sgRNA	Single guide RNA
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
RT-PCR	Reverse transcription PCR
IF	Immunofluorescence

ATAC-seq	Assay for transposase-accessible chromatin using sequencing
lncRNA	Long non-coding RNA
SNP	Single nucleotide polymorphism
PET	Paired-end tag
ChIP	Chromatin immunoprecipitation
prom	Promoter
e(1-7)	Enhancer 1-7

List of Figures

List of Tables

Acknowledgements

First, I'd like to thank my supervisor, Dr. Swneke Bailey, for giving me the opportunity to do science in his lab and for guidance throughout my two years. Thank you to my lab mates—in particular Juliana Cavalcante de Moura for being my first experience training someone and also my lab buddy. Thank you to close collaborator Dr. Xiaoyang Zhang for scientific discussion and troubleshooting help. Thank you to committee members Dr. Josee Dostie and Dr. Patricia Tonin for their guidance and respective expertises. I'd also like to thank the Bailey, Garzia, and Ferri labs in general for shared knowledge and resources. Finally, thank you to our funding sources, in particular the CGS-M award, for their support of this research.

Contribution of Authors

Chapter 1: Literature Review & Hypothesis

Chapter 1 was written by the MSc candidate, Colleen Russett. Editorial review for this chapter was done by Dr. Swneke Bailey.

Chapter 2: Materials and Methods

Chapter 2 was written by the MSc candidate, Colleen Russett. Editorial review for this chapter was done by Dr. Swneke Bailey.

Chapter 3: Results

Skewing analysis was done with the help and guidance of Dr. Swneke Bailey.

The GWAS SNP analysis was performed by Anna Li (undergraduate) and Dr. Swneke Bailey.

Juliana Cavalcante de Moura (non-thesis Master's) assisted with qPCR experiments and analysis.

Figures 8ab were prepared with assistance from Dr. Swneke Bailey.

All other experimentation and analyses were performed by the candidate, Colleen Russett.

Chapter 4: Discussion

Chapter 4 was written by the MSc candidate, Colleen Russett. Editorial review for this chapter was done by Dr. Swneke Bailey.

Chapter 5: Conclusions and Future directions

Chapter 4 was written by the MSc candidate, Colleen Russett. Editorial review for this chapter was done by Dr. Swneke Bailey.

Chapter 6: Supplementary Figures and Tables

Chapter 6 was prepared by the MSc candidate. HOMER analysis in Fig. 14 was performed by Dr. Swneke Bailey; the figure itself was prepared by the MSc candidate.

Chapter 1: Literature Review and Hypothesis

1.1 Literature review

1.1.1 Introduction to gastro-esophageal adenocarcinomas

Gastro-esophageal adenocarcinomas (GEAs) represent a large portion of the global disease burden[1]. Roughly 1.1 million new diagnoses and 770,000 gastric cancer deaths were reported in 2020, making it the fourth-leading cause of cancer-related mortality[2]. Development of GEAs is thought to primarily be driven by environmental agents, of which the most broadly relevant are: *Heliobacter pylori* infection[3], and Epstein-Barr virus infection[4,5], and some evidence for a causal role for poor nutrition[6], smoking[7], and heavy alcohol consumption[8]. Accordingly, gastric cancer typically affects older populations, with some 80% percent of all cases occurring in people over the age of 45[9]. A subset of gastric cancers (~10%) are early-onset, appearing under the age of 45. Hereditary diffuse gastric cancer (HDGC) represents a known, albeit very small (~3%) portion of gastric cancer incidences.

While GEA incidence and mortality is clearly declining in many parts of the world, due largely to structural changes in hygiene and nutrition standards and eradication of GEA-associated infectious diseases, GEA remains a significant problem to contend with once acquired. Mortality remains high, particularly as the vast majority of cases present in the late stages of disease. Furthermore, metastasis to the peritoneum is a common and deadly progression of GEA with a median survival of less than four months[10]. Peritoneal metastasis (PM) samples are difficult to obtain as patients largely experience unresectable tumor progressions and receive palliative care, complicating research into metastasis-promoting genomic alterations, as well as into non-surgical

therapeutic alternatives[1]. Non-metastatic primary tumors are treated with surgical resection, sometimes in combination with adjuvant and neo-adjuvant therapy.

Gastric cancer can be anatomically split into two groups: true gastric adenocarcinomas and adenocarcinomas of the gastro-esophageal junction. The most common histological classification schemas are the Lauren and World Health Organization (WHO) classifications. The former identifies two major groups: intestinal (high cohesion between tumor cells, visible glands) and diffuse (poorly cohesive, little/no gland formation, poor infiltration of the gastric wall)[11]. Conversely, the latter finds five categories of tumor: tubular, papillary, mucinous, poorly cohesive, and rare variants, of which "tubular" and "papillary" roughly fall under the Lauren "intestinal" type while "poorly cohesive" corresponds to the Lauren "diffuse" type[12].

Increasingly, however, cancer classification is moving towards molecular groupings of tumor types as technological ability grows and the promised future of precision medicine grows nearer. In 2014, The Cancer Genome Atlas (TCGA)[4] published an extensive evaluation of the molecular characteristics of primary gastric tumors, dividing them into four subtypes. These are: tumors positive for Epstein-Barr virus, microsatellite unstable tumors, genomically stable tumors, and chromosomally unstable tumors. It is clear that the analysis of genetic alterations, epigenetic modifications, and gene expression data in concert will be paramount to identification of novel treatment targets in gastric cancer as well as to patient stratification therein. Going forward, many alterations highlighted in or uncovered by the datasets in such large-scale genomic and epigenomic studies will necessitate closer investigation before they can be deemed clinically actionable.

1.1.2 Enhancers and super-enhancers

Enhancers are a class of regulatory elements which increase the expression of a gene, frequently in a tissue-specific or developmental stage-specific manner. Putative enhancers are

identifiable by several characteristics including location within nucleosome-free chromatin and the presence of active chromatin marks such as histone 3 lysine 27 acetylation (H3K27ac) and histone 3 lysine 4 methylation (H3K4me) as well as of transcription factor-binding sites. They may be proximal or distal to the gene(s) in question.

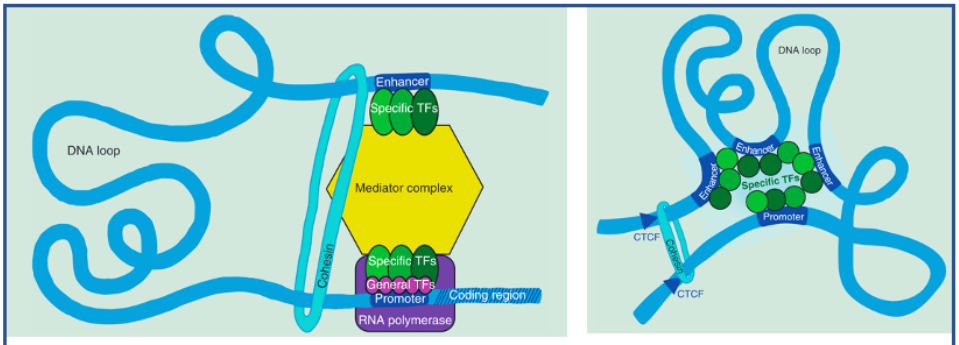


Fig. 1 Mode of enhancer action.

a) Proposed model of enhancer action based on the cohesin loop extrusion; wherein intervening chromatin is looped out to bring distal enhancer region in relative proximity to its target gene. Alternate models of enhancer action, including non-contact models, have been proposed; however, the players involved—such as TFs and the Mediator complex—should remain the same. b) Multiple enhancers and a target gene acting as part of a transcriptional hub.

Used with permission granted from the Copyright Clearance Center. Oct 22, 2022[14].

In order to exert their enhancing function across long genomic distances, distal enhancers are typically brought in proximity of target gene promoters through what is theorized to be a cohesin-based loop extrusion mechanism and/or via self-aggregation in "transcription hubs"[13] (Fig. 1[14]). A proposed underlying mechanism behind self-aggregation is phase separation. Evidence

suggests that transcription hubs constitute dynamically-formed, membrane-less organelles governed by weak and multivalent protein (and RNA) interactions, known as a biomolecular condensates[15] (Fig. 2[16]). Within condensates, loci of similar transcriptional and chromatin states associate themselves to the exclusion of more dissimilar regions.

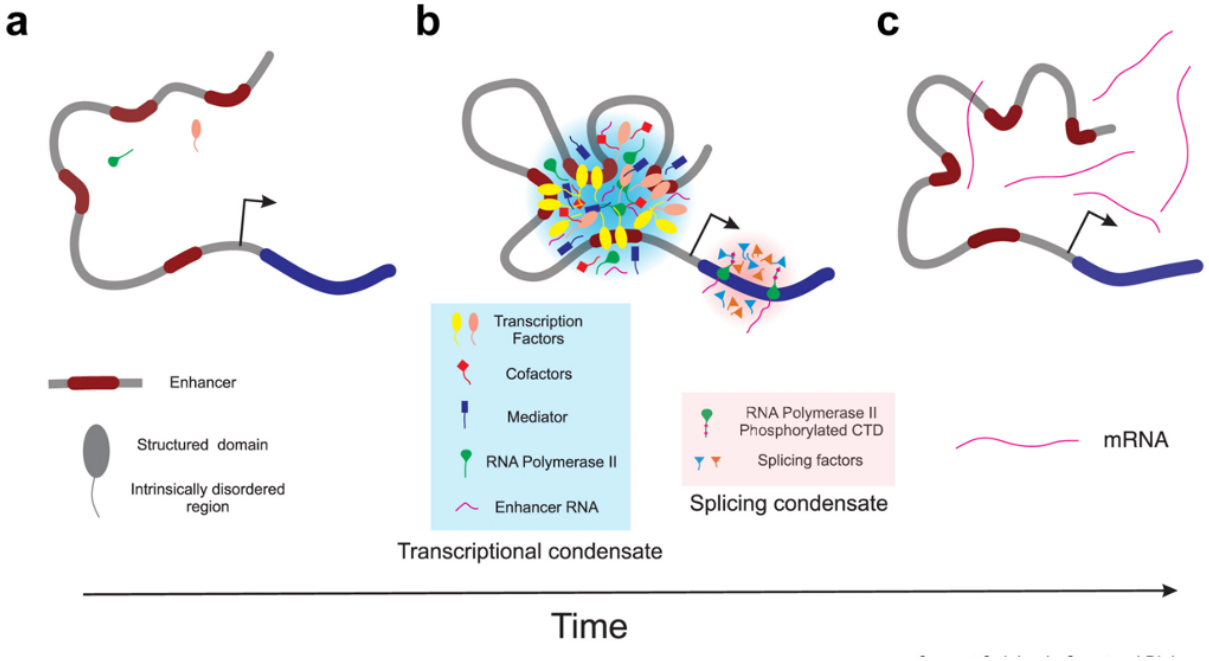


Fig. 2 The role of phase separation in transcription.

Transcriptional hubs are theorized to be at least partially self-organizing, forming dynamic biomolecular condensates through interactions between the intrinsically disordered regions of proteins as well as some other forms of weak, multivalent interactions. Schema depicts the transient formation of a transcriptional condensate from timepoint **a** to timepoint **c** involving multiple enhancers and one target gene, resulting in the production of mRNAs. Used with permission granted from the Copyright Clearance Center. Oct 17, 2022[16].

While it is known that, generally, enhancer activity is initiated by pioneer transcription factors (TFs), which then recruit further TFs, coactivators, histone readers, and finally Mediator/RNAPII to stimulate transcriptional activity, we rather lack a systematic understanding

of enhancer action, especially in mammals[17]. Practically speaking, neither enhancer-promoter contacts measured by chromosome conformation capture, nor other features associated with enhancers, are necessarily definitive in establishing a functional link between a given enhancer and gene. It is therefore difficult to determine that an element is actually an enhancer, which target gene(s) it has relationships with, in what cellular contexts, and to what extent the enhancer affects target gene expression. More decisive detection of enhancers typically requires the use of a reporter assay or of perturbation (most often, deletion, mutation, or heterochromatinization of the enhancer, leading to a corresponding decrease in gene expression of the enhancer's target gene(s)).

Interestingly, enhancers do not necessarily act alone. Indeed, multiple enhancers may cooperatively regulate one or more target genes, either within the same transcriptional hub or as part of a linear cluster of enhancers[18]. Linear enhancer clusters are sometimes termed super-enhancers, when they coincide with other characteristics including the binding of high levels of chromatin regulators and transcription factors—namely BRD4 and Mediator—and the driving of high levels of transcription vis-à-vis their target gene(s)[19]. Super-enhancers have also been noted to fall near lineage-specifying genes, suggesting a role in key developmental processes[20]. The term "super-enhancer" is somewhat controversial, however, as it is not clear that it designates an entity with qualitatively different properties to individual enhancers or even from groups of enhancers that are not bioinformatically designated as "super-enhancers" *per se* (i.e. groups of enhancers that do not meet the thresholds for distance between consecutive enhancer elements)[21,22]. For our purposes, we will make use of literature on both enhancer clusters, in general, as well as super-enhancers to illustrate some of the diverse ways enhancers might behave in groups.

To the former point, some enhancer clusters—such as the cluster of five enhancers which drives mouse *α-globin*—have been found to be distinctly lacking in any emergent properties, as

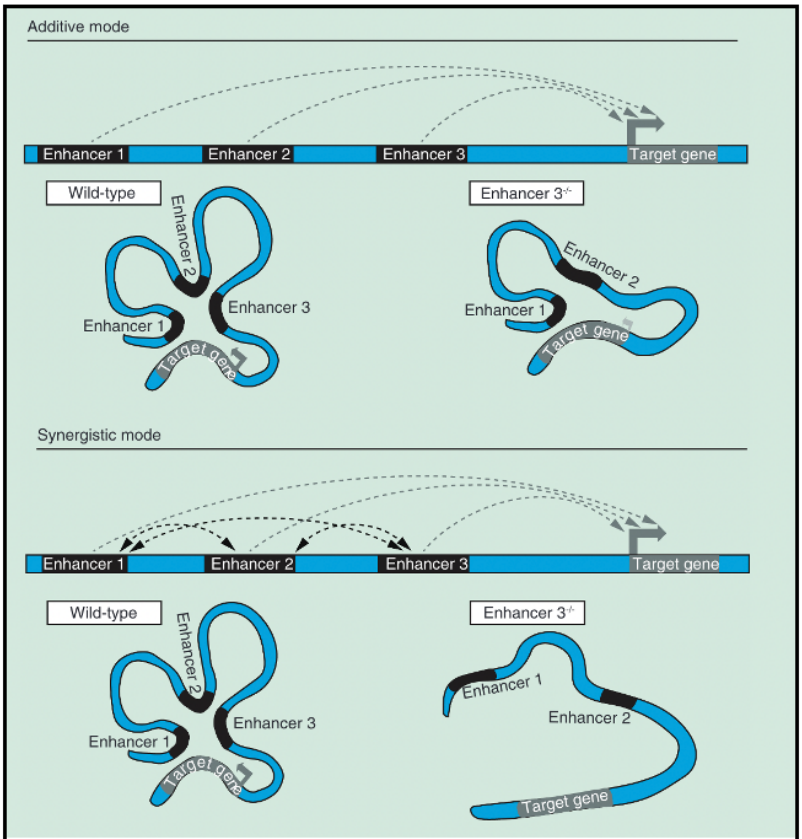


Fig. 3 Additivity and synergy in enhancer clusters. Schema depicting two possible ways for enhancers in an enhancer cluster to relate to each other. In a synergistic cluster, the activity of individual enhancers influence and may be dependent on each other.

Used with permission granted from the Copyright Clearance Center. Oct 22, 2022[14].

constituent enhancers behave in an additive manner with respect to target gene expression[23] (Fig. 3). This is, however, not the only mechanism governing the behavior of enhancer clusters that has been identified. Other enhancer clusters demonstrate non-additive effects as opposed to or in addition to additive effects. Synergistic enhancers reflect a change in target gene expression larger than the sum of individual enhancer activities as derived from transgene experiments. In these cases, the activity of one enhancer is tied to the activities of other enhancers in the cluster. Partial redundancy, a feature now known to be exceedingly common (see: "shadow enhancers"[24]), is also present in enhancer clusters. While each separate enhancer may have only a small or negligible impact on gene expression, such a mechanism may allow constituent enhancers to fine-tune their control of a target gene and/or act as a buffer against genetic or environmental disturbances[18,25].

Some constituent enhancers appear to be disproportionately responsible for driving target gene expression and otherwise contributing to 3D chromatin interactions across the cluster, forming what can be considered a hierarchy of enhancers[26]. Huang and colleagues have termed constituent enhancers at the top of the hierarchy of hierarchical super-enhancer, "hub enhancers." At the mouse *Wap* locus—an example of a hierarchical enhancer cluster showing both additive and synergistic effects—disabling TF-binding sites at the most proximal constituent enhancer disables function of the entire super-enhancer[27]. Taken together, the roles that enhancers play in regulation on their own and in the context of other enhancers are varied and, as of yet, difficult to predict on a wide scale. There remains a need to dissect enhancer clusters on an individual basis in cell types and developmental stages of interest in order to capture the precise function and mechanisms at play.

1.1.2.1 Enhancers and super-enhancers in disease

Enhancers also represent interesting and underexplored conduits for disease. It has been known for some time that most genetic variants probed by genome-wide association (GWAS) studies (GWAS) map to non-coding regions[28]. It is not a surprise, then, that disease-associated variation has been found to be enriched in enhancers[28], super-enhancers[20], as well as specifically within the hub enhancers of hierarchical super-enhancers[26]. Although loss-of-function mutation of enhancers are possible—i.e. variants which disrupt binding of key TFs or interfere with the enhancer's ability to localize to or act on its target gene(s)—deleterious gain-of-function mutations may be more common[29]. Depending on the exact nature of the variant, enhancers activity may be bolstered or, in the case of larger-scale variants, the location of enhancers themselves may be altered with respect to the location of new, potential target genes.

1.1.3 On somatic copy number alterations in cancer

Structural variants (SVs) such as amplifications, deletions, and chromosomal rearrangements are frequent events in tumorigenesis and form an integral part of this complex, multi-step process. A subset of structural variants, known as copy number variants (CNVs), refer to those variants that comprise duplication or deletion of parts of the genome. The size of copy number losses or gains can be as small as trinucleotide repeats or large enough to implicate an entire gene or multiple genes, often correlating with an altered level of gene expression[30]. CNVs which arise post-zygotically and result in copy number changes seen in the DNA of somatic cells—most saliently in cancer cells—but not in paired germline DNA are thus often referred to as somatic copy number alterations (SCNAs).

In particular, SCNAs involving copy number gain may be inserted on the chromosomes itself or—as is more frequently the case with extremely high copy number gains—be maintained on circularized bits of extra-chromosomal DNA (ecDNAs). These SCNAs are often associated with oncogene overexpression[31] and dependency[32] in cancer, reflecting the selective advantages, as well as vulnerabilities, that they can impart to cancer cells. Correspondingly, increasing evidence suggests that SCNA genes are useful as criteria for molecular subtyping of cancers, as biomarkers, and as potential therapeutic targets[33].

Although much research has focused on SVs targeting oncogenic protein-coding genes, SVs also frequently appear to target non-coding regions. Several emerging paradigms suggest that variants recurrently targeting regulatory non-coding regions, either on their own[34] or in tandem with oncogenes[35], also have functional consequences. As drivers of gene expression, chromosomal rearrangements and other SVs reposition enhancer elements near proto-oncogenes aberrantly activate or hyper-activate them in a paradigm that has been termed “enhancer hijacking”[36]. Focal

amplifications of regular or super-enhancers (SEs) – defined as enhancer clusters with unusually high levels of Mediator binding as well as other indicators of enhancer activity – produce a similar effect, often by altering local chromatin interactions[34].

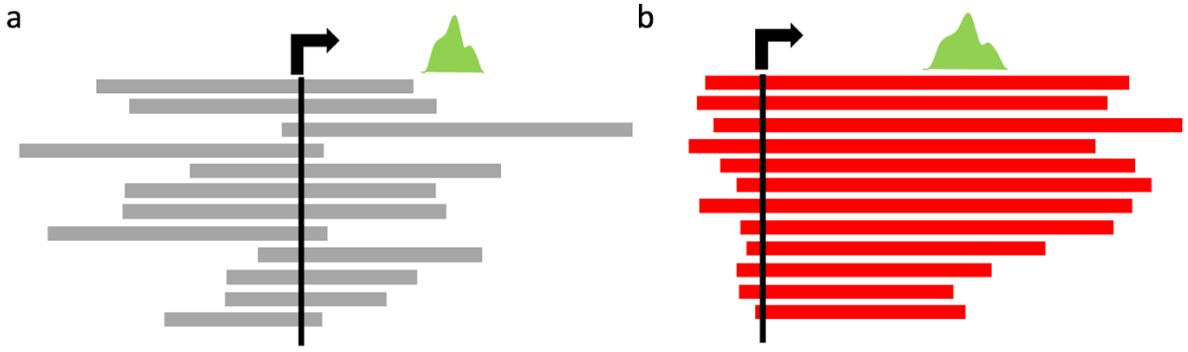


Fig. 4 Amplicons select for enhancer elements.

Schema of oncogene-containing cancer amplicons showing (a) no skew and (b) skew towards the inclusion of endogenous enhancer elements. Oncogene is represented by the black arrow and enhancer element by the green peak. Grey and red rectangles represent amplicons derived from patient samples in amplified cancers.

Finally, Morton et al. (2019) demonstrated that amplified regions in multiple cancers appear to select for the inclusion of endogenous enhancer elements, alone or along with amplified oncogenes (Fig. 4). Interestingly, enhancer amplifications maintained on ecDNAs are able to provide novel topological contexts for enhancer-promoter contacts (Fig. 5). These novel contacts or perhaps increased chromatin accessibility of the ecDNAs may partially explain why copy-number-normalized oncogene expression on ecDNAs is higher than on non-circular amplicons[25]. Additionally, (super)enhancer elements on ecDNAs function as mobile enhancers for chromosomal DNA[37] as well as for other ecDNAs[38], driving widespread transcriptional activation, including of oncogenes. These examples point to a strong role for enhancers—and the variants that target them—as cancer-driving mechanisms.

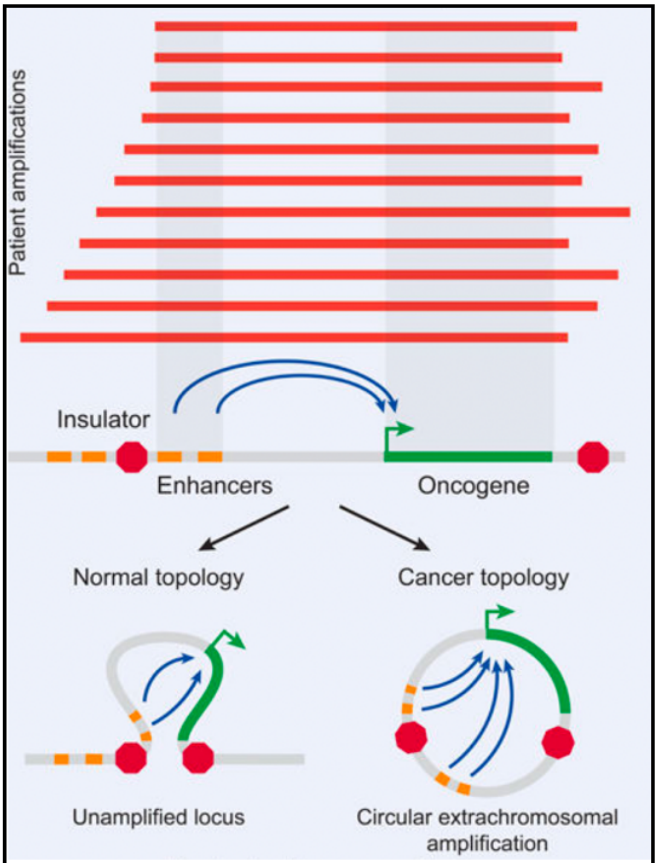


Fig. 5 Amplified loci have altered topology. Schema representing the differential topology of ecDNA-derived amplicons compared to unamplified loci with respect to their enhancer-promoter contacts.

Used with permission granted from Copyright Clearance Center. Oct 17, 2022[35].

Our locus of interest—the 20q13 chromosomal region—has long been identified as a region subject to recurrent copy number amplifications across many cancer types. Early reports focused on the 5.4%-40% of breast cancer primary tumors and cell lines that were 20q13-amplified and which were associated with aggressive phenotype and poor outcome[39–41]. Several candidate oncogenes were proposed for the region before Collins et al. (1998) described a ZNF217-containing, 260 kb region of maximal amplification within 20q13.2[42]. Multiple papers have extended these findings to other cancers, including ovarian clear cell[43], prostate[44], colon[45], glioblastoma[46], and, cancer-of-interest, gastric cancer[4,47]. Of particular note, The Cancer Genome Atlas published a comprehensive molecular characterization of gastric cancer in 2014, which included analysis of somatic copy number alterations (SCNAs) by GISTIC, confirming focal amplification of ZNF217 in the chromosomal instability (CIN) subtype of gastric cancer[4].

1.1.4 On ZNF217 regulation and downstream effects

1.1.4.1 ZNF217 regulatory functions

Zinc finger protein 217, or ZNF217, is a Krüppel-like zinc-finger transcription factor located within the 20q13.2 chromosomal region. ZNF217 is expressed in normal tissue and confers oncogenic properties when overexpressed. In the nucleus, ZNF217 effectuates gene expression changes by directly binding to DNA with two of its eight zinc-fingers[48] and indirectly, by recruiting other proteins with repressor functions. Among ZNF217's frequent interactors are the histone demethylases JARID1B and LSD1, the histone methyltransferases EZH2 and G9A, and the co-repressor complexes CTBP1/2 and CoREST, which recruit further histone-modification enzymes such as histone deacetylases (Fig. 6)[49].

Although a majority of the literature has focused on ZNF217's repressive functions, a subset of genes is upregulated upon ZNF217 knockdown, indicating that ZNF217 may also act as a transcriptional activator[50,51]. Furthermore, there are a few examples of ZNF217 positively regulating gene expression through direct binding to gene promoters[52]. Notably, the interaction with LSD1 provides a potential avenue through which ZNF217 could enact gene activation. LSD1 is a demethylase of histone 3 lysine 4 (H3K4) methylation, a mark associated with gene activation; however, LSD1 is also capable of demethylating repressive histone marks, such as histone 3 lysine 9 (H3K9) mono- and di-methylation, which could lead to reactivation of these genes[53,54]. The exact mechanism by which LSD1 and potential co-activators specifically activate genes rather than repress them in a ZNF217-relevant context, remains to be elucidated.

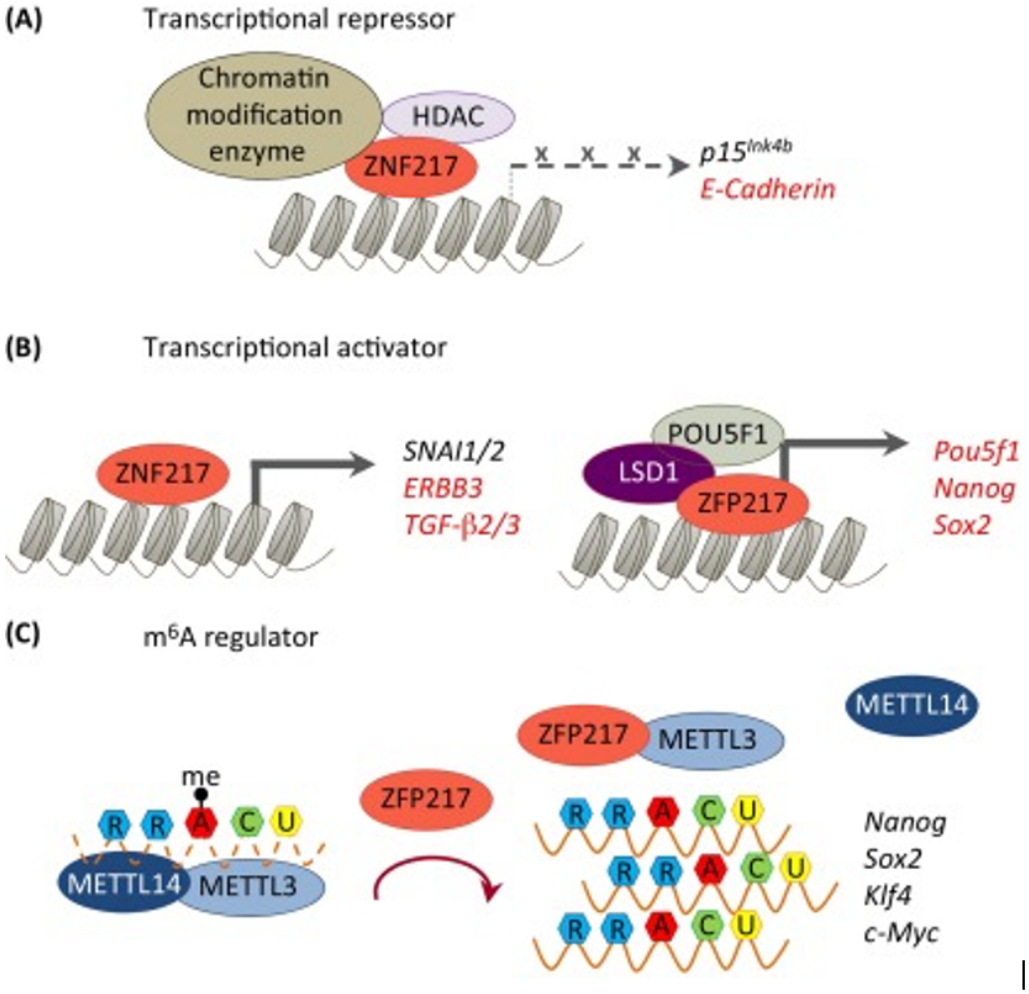


Fig. 6 ZNF217 mechanisms of action.

Schemas showing proposed mechanisms by which ZNF217 acts as (a) a transcriptional repressor, (b) a transcriptional activator, and (c) a post-transcriptional regulator of m^{6}A.

Used with permission granted from the Copyright Clearance Center. Oct 17, 2022[49].

Interestingly, ZNF217 is also known to have a role in post-transcriptional regulation through its interaction with mediators of m^{6}A deposition, such as methyltransferase-like 3 (METTL3) and 14 (METTL14). The deposition of m^{6}A on mRNAs modulates translation efficiency, nuclear

retention, splicing, and stability of mRNA molecules[55]. In particular, it has been linked to the regulation of core pluripotency factors. ZNF217 appears to positively regulate the stability of pluripotency-related mRNA transcripts by sequestering METTL3 in an inactive complex, reducing the amount of transcripts modified by m^6A, thus preventing them from being targeted for degradation[55].

1.1.4.2 ZNF217 amplification and overexpression in cancer

Since the discovery of its recurrent amplification, studies have linked ZNF217 to a number of *in vitro* cancer phenotypes, including cell immortalization[56] and stemness, anti-apoptosis[57–59], epithelial-to-mesenchymal transition (EMT)[60,61], and several metastasis-related phenotypes, among others. To begin, cancer cell immortalization is a common feature of tumor progression, which allows cancer cells to proliferate beyond the limits typically set by a cell's telomeres. In human mammary epithelial cells (HMECs) with silencing of the tumor suppressor, $p16^{lnk4a}$, ectopic ZNF217 expression led to immortalization[56]. The immortalized state was accompanied by increased presence of TRF2 at telomeres, indicating that ZNF217-overexpressing cells may avoid senescence at least partially through the stabilization of telomere length[62].

Relatedly, ZNF217 overexpression has been repeatedly linked to evasion of apoptosis. When TRF1 and TRF2 are manipulated so as to trigger apoptosis through telomeric dysfunction, ZNF217-overexpressing cells showed attenuated cell death[57]. The enhanced lifespans of ZNF217-overexpressing cells may be a consequence ZNF217-mediated upregulation of eEF1A2, whose inactivation typically promotes apoptosis[58]. In addition to attenuating apoptotic signals originating from DNA damage, ZNF217 was also able to interfere with apoptotic signals imparted by chemotherapeutic agent, paraclitaxel, as part of the Bcl2-controlled mitochondrial apoptosis pathway in MDA-MB-231 breast cancer cells[59].

One feature of ZNF217-overexpressing cancers that is particularly of interest with respect to ZNF217's theorized role in metastasis is epithelial-mesenchymal transition, or EMT. EMT is a process by which epithelial cells lose epithelial characteristics such as E-cadherin-mediated cell-cell adhesion and cell polarity, in favour of gaining mesenchymal properties. These mesenchymal properties—such as invasion and migration—aid in the initiation of tumor invasion and metastasis. In human mammary epithelial cells, ZNF217 overexpression has been shown to promote several features of EMT, including downregulation of E-cadherin and other epithelial markers, upregulation of mesenchymal markers such as N-cadherin and vimentin, increased stem cell-like properties such as enhanced ability to form mammospheres, as well as demonstrating a more fibroblastic morphology[60]. Acquisition of these features is at least partially driven by induction of the TGF-ß-activated Smad pathway, which ZNF217 may be able to directly regulate through binding to TGFB2 and TGFB3 promoters. In MDA-MB-231 breast carcinoma cells and SCp2 mouse mammary epithelial cell lines, ZNF217 overexpression was associated with increased invasiveness in Matrigel/Boyden chamber invasion assays as well as increased motility in scratch wound assays[60,61].

As part of the transition, cells undergoing EMT must evade a subset of apoptosis which occurs in epithelial cells in response to detachment from the correct extracellular matrix (ECM), known as anoikis, and acquire the ability to grow in an anchorage-independent manner. In a soft agar assay performed in the fibroblast cell line, NIH3T3, overexpression of ZNF217 promoted the formation of larger and a greater number of colonies than controls[61]. Notably, the exon 4-skipping isoform of ZNF217 (ZNF217-ΔE4) has been shown to have an even greater effect on anchorage-independent growth than wild-type[63].

The ZNF217-mediated acquisition or maintenance of stem-cell-like features has been observed in cancer stem cells (CSCs), which are defined as a subpopulation of cancer cells which drive tumor initiation[64,65]. In glioma stem cells (GSCs)—which are thought to drive tumorigenesis in glioblastoma—ZNF217 is upregulated and contributes to maintenance of stem cell properties during culture under hypoxic conditions, such as self-renewal and differentiation capacity[46]. This upregulation appears to be mediated by hypoxia-inducible factors (HIFs), HIF1**α** and HIF2**α,** whose knockdown represses ZNF217. In breast cancer cells, hypoxic conditions/HIFs were found to trigger ZNF217-dependent inhibition of m6A deposition on core pluripotency factor mRNAs such as Nanog and KLF4 by METTL3 and METTL14, leading to increased presentation of breast cancer stem cell (BCSC) phenotypes[66].

Finally, there is some evidence that ZNF217 has utility as a prognostic and predictive indicator. In patients with estrogen receptor positive (ER+), human epithelial growth factor receptor 2 (HER2) negative, lymph node negative breast cancer, high ZNF217 expression predicted worse prognosis in terms of overall survival, relapse-free survival, and metastasis-free survival[61]. In addition, ZNF217 expression is an independent prognostic factor for relapse-free survival in gastric cancer[67], as well as a predictor of poor response to endocrine therapy, also in ER+ breast cancer[68].

Taken together, there is substantial evidence to suggest that ZNF217 acts as an oncogene in multiple cancers and furthermore that ZNF217 inhibition may be worthwhile pursuing in the context of drug therapies; however, a comprehensive understanding of ZNF217's mode(s) of action in cancer cells—starting from its hyperactivation to its role in oncogenic signaling pathways, and to its multiple functions as a repressor, activator, and post-translational modifier—has yet to be fully elaborated.

1.1.5 On therapeutic interventions for ZNF217-amplified cancers

While ZNF217 has been proposed as a potential therapeutic target[61], currently there exist no drugs which directly target ZNF217. ZNF217-amplified cancers may, however, benefit from therapeutic interventions that can intervene at other junctions of ZNF217's oncogenic gene expression program. One example of such a therapy might be bromo- and extra-terminal (BET) inhibitors or degraders, which function by, respectively, inhibiting or degrading BET family members such as BRD4. BET family proteins preferentially bind to acetylated—activating—histone lysine residues and recruit transcriptional activators (Fig. 7[69])[70]. BET inhibitors and degraders introduce the intriguing possibility of intervening upstream of oncogene activity, by way of using oncogene-driving enhancers as direct therapeutic targets.

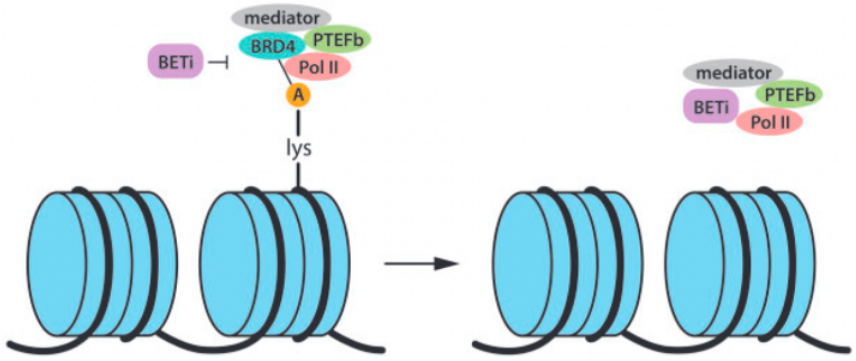


Fig. 7 BET inhibitor mode of action.

Schema showing a BET inhibitor (BETi) molecule interfering with binding of the epigenetic reader, BRD4, to the acetylated lysine residues of histones. BRD4 is thus unable to recruit Mediator, RNA polymerase II, and others which play a role in active transcription. Used with permission granted from the Copyright Clearance Center. Oct 17, 2022[69].

Although BET family proteins localize to many thousands of enhancers in the genome, targeting them may still have gene-specific effects. For example, a recent paper made use of the

proteolysis targeting chimera (PROTAC) molecule ARV-771 to degrade BRD4 via recruitment of the E3 ligase cereblon and observed that the oncogene SOX2 was hypersensitive to BRD4 degradation *in vitro*, leading to a decrease in its expression[71]. Interestingly, the BRD4 inhibitor, JQ1, has been shown to selectively affect BRD4 and BRD4 interactor occupancy at super-enhancers resulting in preferential loss of transcription at super-enhancer-associated genes in multiple myeloma cell lines[70]. How specific cancer drugs achieve or can be leveraged to achieve this kind of intracellular target specificity remains to be clarified; however, one theory suggests that it may occur through biomolecular condensate formation[72]. There are several reports of DNA features common in enhancer clusters driving condensate formation, including those that accrue Mediator and RNAPII[73], BRD4[74], and TFs in large numbers[75]. It has also been suggested that ecDNAs form condensate-like congregations at "ecDNA hubs," where ecDNAs co-localize and facilitate inter-ecDNA promoter-enhancer activation. In one study, JQ1 was able to dispel these congregations, thus preferentially suppressing ecDNA-amplified-oncogene expression[38]. Recently, contradictory reports have emerged which found no evidence of ecDNAs spatially clustering together or indeed overlapping with transcriptional condensates (preprint)[76]. Nonetheless, given the known association between oncogenes and super-enhancers as well as oncogenes enrichment on ecDNAs, it's possible that certain cancers will be selectively sensitized to the effects of BET inhibitors/degraders at super-enhancer-driven and/or highly amplified oncogenes.

Thus far, several clinical trials testing the efficacy of BET inhibitors in human cancer patients have encountered dose-limiting toxicities; however, a priority in the next generation of clinical trials will be to increase the selectivity of BET inhibitor therapy, both on the level of (1) mechanism (i.e., by testing compounds that are specific to the individual bromodomains of BET proteins) and (2) using predictive biomarkers to identify potential responders[77]. Determining the

mechanisms leading to the hyperactivation of oncogenes, such as ZNF217, may thus also reveal predictive biomarkers for sensitivity to BET inhibitor therapy.

1.1.6 Methodology: CRISPRi and its use in enhancer knockdown

The introduction of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein 9 (Cas9) systems as a tool for molecular biology has had considerable impact on research. Shortly after conventional CRISPR/Cas9 systems first appeared in publication, CRISPR-based alternatives began appearing as well, including CRISPR interference (CRISPRi)[78,79]. CRISPRi involves a catalytically inactive Cas9, called dead Cas9 (dCas9), which can still interact with a single guide RNA (sgRNA) in order to target a specific genomic region, but can no longer make double-stranded breaks. In its original conception, dCas9 instead sits at or near a gene promoter in order to block transcriptional initiation or elongation. In theory, CRISPRi results in gene "knockdown" as opposed to a "knockout," not unlike in RNA interference (RNAi), albeit by targeting the production of mRNA as opposed to the mRNA itself.

Subsequent innovations in CRISPRi technology have seen the fusion of dCas9 to repressor domains in order to improve the efficiency of knockdowns. In our work, we have opted to use the variation developed by Yeo et al. (2018), which uses a combined Kruppel-associated box (KRAB) domain and methyl CpG binding protein 2 (MECP2) fusion to improve upon previous designs containing only one repressor domain[80]. The KRAB domain acts by recruiting KAP1 protein, which acts as a scaffold for histone deacetylases (HDACs), the H3K9-specific methylase, SETDB1, and finally, heterochromatin protein 1 (HP1), leading to facultative heterochromatin formation[81]. MECP2, on the other hand, binds to methylated CpGs in the vicinity and complexes with a co-repressor containing mSin3A as well as HDACs[82]. 60% of guides tested by Yeo *et al.*

(2018) against endogenous genes showed improved repression by dCas9-KRAB-MECP2 compared to dCas9-KRAB[80]. The effect was independent of the DNA strand targeted.

Although CRISPRi in its various forms does not appear to be as efficient as conventional CRISPR/Cas9, it also presents some distinct advantages[83]. Small deletions such as those catalyzed by non-homologous end joining (NHEJ) may not be sufficient to interfere with enhancer function, especially as it may not always be possible to disrupt critical TF binding sites[84]. In particular, conventional CRISPR has been found to pose issues in cell lines with copy number gains as the creation of double-stranded breaks is correlated with an anti-proliferative effect likely linked to the DNA damage response[85,86]. Furthermore, amplifications in tandem on the chromosome may lead to loss of entire loci, including of regions containing oncogenic driver genes, on which cells may be dependent for proliferation and/or survival. Thus, when targeting amplified regions with CRISPR, such as in the case of some cancers, it is advisable to use a method of gene repression that does not rely on a nuclease activity.

1.2 Rationale, Hypothesis and Aims

1.2.1 Rationale

Structural variants (SVs) such as amplifications, deletions, and chromosomal rearrangements are frequent events in tumorigenesis and can function as actionable biomarkers by guiding therapeutic decision-making for at-risk patients. The 20q13.2 chromosomal region has been found to be recurrently amplified in a number of different cancers, including gastroesophageal cancers and breast cancer. The leading candidate oncogene in the region is ZNF217, a zinc-finger transcription factor which acts as part of various complexes as a transcriptional repressor or activator. Overexpression of ZNF217 has been most extensively characterized in breast cancer and is associated with immortalization of cancer cells, metastasis phenotypes *in vitro* and *in vivo*, as well as poor prognosis in human patients[61,87]. Despite these reports, a mechanism responsible for ZNF217 hyperactivation in these cancers has yet to be established. This is particularly relevant as identifying mechanisms of hyperactivation for prognostic and/or predictive biomarkers like ZNF217 lead to alternative therapeutic interventions. Of note, we have identified a recurrently co-amplified non-coding region adjacent to ZNF217, using publicly available array data from TCGA, that appears to be under positive selection together with the ZNF217 gene. This region—common to gastric, breast, and prostate cancer, and possibly others—harbors multiple active enhancer elements.

1.2.1 Hypothesis

In keeping with this background, we hypothesize that focal amplification of an SE in the 20q13.2 drives aberrant ZNF217 hyperactivation through the alteration of chromatin contacts in the region, leading to tumor progression in GEA.

1.2.2 Aims

In this thesis, I aim to:

(1) narrow down a subset of enhancers within the co-amplified region and

(2) employ CRISPRi to determine the contribution of these enhancers to ZNF217 hyperactivation in amplified gastric cancer lines.

Chapter 2: Materials and Methods

2.1 Cell culture

We selected the gastric cell lines MKN1 and MKN74 as they were found to harbour amplification of the 20q13.2 region[88]

. Cell lines were cultured according to ATCC guidelines.

2.2 Puromycin/blasticidin kill curves

MKN1 or MKN74 cells were plated in 24-well plates for a confluency of ~70% the following day. For blasticidin S hydrochloride (Sigma 15205-25MG), wells for treatment at 0, 2, 4, 6, 8, and 10 ug/mL were plated in duplicate. For puromycin, wells for treatment at 0, 1, 2, 3, 4, and 5 ug/mL were plated in duplicate. Doses were chosen based on the minimum dose that killed cells in ~3-5 days (5 ug/mL blasticidin and 2 ug/mL puromycin).

2.3 Creation of stable CRISPRi cell lines

2.3.1 Lentiviral production

Plasmids of interest (dCas9-KRAB-MECP2[80], lentiGuide-puro: Addgene #52963[89], GFP-containing pPRIME test vector) and 2nd generation lentiviral packaging plasmids, VSV-G and psPAX2 were purified using Qiagen Plasmid Plus Midi Kit (Qiagen 12943). 293Ts were plated the day before the transfection such that the cells are less than 70% confluent on the day of transfection. For a 6-well plate, 1800 ng DNA was used per well at a ratio of 3: 2: 1, i.e. 900 ng transfer plasmid, 600 ng PAX2 (50 ng/uL), 300 ng VSV-G (50 ng/uL). DNA was diluted with enough PBS to make a 50 uL total volume. The ratio of DNA to polyetherimide (PEI) was 1:4, i.e. 1800 ng DNA x 4 = 7200 ng PEI, added into a second Eppendorf tube and diluted to 50 uL. DNA tubes and PEI

tubes were mixed and incubated for ~20-25 minutes at room temperature (RT). Contents were added dropwise to 293Ts. Transfection success for pPRIME-GFP vector was checked using the EVOS M7000 microscope. After 48 hours, virus-containing media was collected, placed in 15 mL tube at 4C, and cell media replaced. After 72 hrs, virus was collected again and mixed with virus collected at 48 hours, and used immediately to infect cells.

2.3.2 Viral infection and blasticidin selection

MKN1 and MKN74 cells were plated for next-day 70% confluency in 6-well plates and incubated for 24 hours with different amounts of viral supernatant, from 1 mL of 1:5 supernatant to RPMI up to 1 mL of supernatant only. Cells recovered for one day in normal media before starting blasticidin selection the following day. Concentrations of viral supernatant that produced 10-20% remaining live cells after blast selection were grown out and frozen down. Final experiments were performed with MKN1 and MKN74 cells treated with 1:2 and 1:5 supernatant to RPMI media, respectively.

2.4 sgRNA cloning into LentiGuide-Puro vector

sgRNA cloning was performed as previously described in the Zhang lab's lentiCRISPR cloning protocol. LentiGuide-Puro vector was digested and dephosphorylated with BsmBI for 30 min at 37C. Gel purification of digested plasmid was done using the QIAEX II Gel Extraction Kit (Qiagen 20021). Oligos were phosphorylated and annealed simultaneously in a thermocycler with the following parameters: 37 C for 30 min, 95 C for 5 min, followed by a ramp to 25 C at 0.1 C/s. Oligos were diluted 1:200 in sterile water, then ligated at room temperature for 10 min using Quick Ligase (NEB M2200S). Finally, plasmids were transformed into Stbl3 competent cells (Invitrogen C737303).

2.5 Infection of CRISPRi cell lines with lentiGuide-Puro and puromycin selection

100K MKN1 or 150K MKN74 cells were plated in 6-wells and infected the day after with 1 mL of lentiGuide-Puro viral supernatant containing 5 ug/uL polybrene (Sigma-Aldrich TR-1003-G). Cells were allowed to recover for one day in regular media followed by puromycin selection (2 ug/mL) for 4-7 days.

2.6 qPCR

On d4 or d7 of puro selection, cells were lysed from single well of a 6-well plate with 350 uL lysis buffer and extracted according to Qiagen RNeasy RNA Mini Kit (Qiagen 74104). RNA in RNase/DNase-free water was stored at -80 C until cDNA synthesis could be performed with the Superscript IV First-Strand Synthesis kit (Invitrogen 18090050). The protocol was followed using 1 ug of RNA per reaction and a 1:1 ratio of random hexamers:oligod(t) primers. New qPCR primers were tested by RT-PCR and gel electrophoresis as well as by performing primer efficiencies (acceptable range = 90-110%) and melt curves by qPCR. qPCR experiments were performed in triplicate using Ssoadvanced Universal SYBR Green Supermix (Bio-Rad 1725272), 250 nM primers (final concentration), and cDNA of the appropriate dilution (final experiments were done 1:20) in semi-skirted 96-well plates (Bio-Rad 2239441) sealed with optical sealing tape (Bio-Rad 2239444) on the Bio-Rad MyIQ Single-Color Real-Time PCR Detection System (Bio-Rad 170-9740). Analysis was done in Excel according to the method described in Vandesompele et al. (2002)[90]. Expression was normalized using three reference genes: GADPH, HPRT1, and ACTB.

2.6.1 qPCR program

- Polymerase activation/initial denaturation: 95 C for 30 s
- 40 cycles 95 C for 10 s
  60 C for 30 s
- Melt curve (optional): 65 C 95 C in 0.5 C increments

2.7 Immunofluorescence (IF)

MKN74 cells were plated at ~50K on glass cover slips in a 24-well plate and allowed to grow for 1-2 days. Cells were washed with PBS x2 to remove media, waiting 5 min in between each wash, then aspirating the PBS at the end of the final wash. 300 uL of 4% PFA was added under the chemical hood and incubated with cells for 15 min at RT, followed by PBS wash x2. Cells were blocked with 100 uL of a 5 mL w/v solution of 0.1 g BSA, 12.5 uL Triton-X in PBS, by adding the blocking buffer to coverslips placed on Parafilm on an opaque container with a lid. Blocking proceeded for 30 min, followed by PBS-T wash x3. Anti-Cas9 polyclonal antibody (Diagenode, C15310258) was diluted 1:1000 in blocking buffer and incubated in 4 C overnight. Primary was removed and slips were washed with PBS-T x3, then incubated for 1h30 min with donkey anti-rabbit-Alexa 647 (Abcam 150075) diluted 1:1000 in blocking buffer. Secondary was removed and cover slips were washed with PBS-T x3. Coverslips were then dried, by placing on side on paper towel, and mounted on a glass slide using ProLong Gold Antifade with DAPI (Invitrogen P36941). Slides were left to cure overnight in a dark place, then imaged on EVOS. All PBS washes in the 24-well plate were done using 500 uL PBS and all washes done on coverslips on Parafilm were done using 100 uL PBS.

2.8 Skewing analysis

Analysis was adapted from Morton et al. (2019)[35]. Amplicons overlapping ZNF217 with a log relative copy number ratio greater than or equal to 1 were used in the analysis. Briefly, amplicons were extended if the adjacent relative copy number ratio of neighboring amplicons

were at least 80% of the neighboring amplicon. Permutations were performed to randomly shuffle amplicons 1,000 maintaining the length of each amplicon and keeping the gene, ZNF217, within the amplicon. 1000 base pair windows were used to calculate the overlapping frequency of the shuffled amplicons, the expected distribution.

2.9 HiChIP analysis

HiChIP data for the MCF7, ZR75, T47D, AGS, OE33, and HT55 cell lines was generated through collaboration with Xiaoyang Zhang. HiC-Pro[91] was used to process the HiChIP data and this output was subsequently run through hichipper[92] for quality assessment and loop visualization.

Chapter 3: Results

3.1 Identification of significantly co-amplified, enhancer-containing region adjacent to ZNF217

In 2014, The Cancer Genome Atlas (TCGA) released a comprehensive molecular characterization of gastric adenocarcinoma, including GISTIC analysis of recurrently amplified regions[4]. In this analysis, the 20q13.2 region containing ZNF217 was found to be recurrently amplified, specifically in the chromosomal instability (CIN) subgroup (Fig. 8a). We reanalyzed GISTIC data from the TCGA, paired it with TCGA ATAC-seq tracks as well as our own H3K27ac Chipmentation data, and delineated a non-coding region which is both adjacent to ZNF217 and which overlaps the most frequently amplified region across 1075 breast and 438 gastric cancer patients (Fig. 8b). This non-coding region contains no protein-coding genes, two uncharacterized lncRNAs, and a number of open chromatin regions. We hypothesize a portion of these open chromatin regions to be active enhancer elements linked to ZNF217 expression.

Amplified regions in multiple cancers appear to select for the inclusion of endogenous enhancer elements, alone or along with amplified oncogenes, which can then drive aberrant gene expression[82]. In order to show that our locus is co-selecting for the candidate enhancer-containing region adjacent to ZNF217, we adapted a method from Morton et al. (2019), which aims to compare the actual distribution of amplicons around our region of interest to an expected null distribution (Fig. 8c)[35].Where the actual frequency of overlapping amplicons is greater or less than the expected frequency, indicates a region that is amplified more or less frequently than expected by random chance, suggesting that there is selection for the inclusion of these regions in cancer.

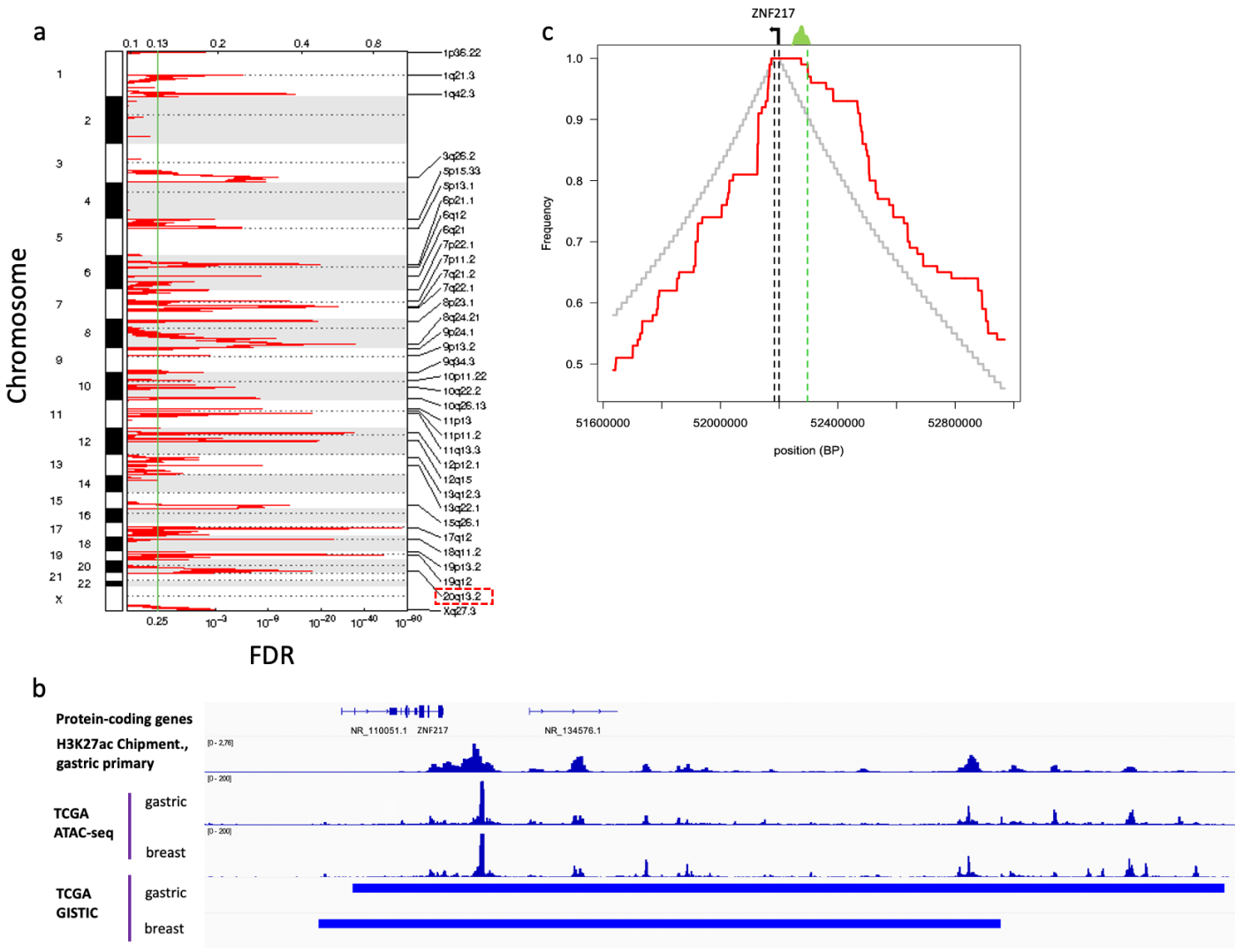


Fig. 8 Identification of significantly co-amplified, enhancer-containing region adjacent to ZNF217.

(a) GISTIC 2.0 analysis of focal amplifications in gastric cancer plotted by chromosomal location and false discovery rate (FDR). Peaks are annotated by cytoband. (b) Schema showing the location of known oncogene, ZNF217, and lncRNA genes, NR_110051.1 and NR_134576.1, within 99% confidence region from GISTIC analysis of TCGA breast and gastric cancer samples. ATAC-seq (TCGA) and H3K27ac Chipmentation peaks indicate potential enhancers elements within the co-amplified non-coding region. (c) Actual distribution of amplicons from 1075 breast and 438 gastric cancer compared to null distribution. Non-coding shelf to the right of ZNF217 region is amplified more frequently than expected by random chance. Priority enhancers were identified from overlap between shelf and enhancer cluster of interest.

We found that there is co-selection for a non-coding "shelf" adjacent to ZNF217, indicating that the contents of this shelf may be functionally important.

3.2 Identification and prioritization of candidate enhancers in 20q13.2 region

After accounting for the cutoff indicated by the non-coding shelf, there were still many candidate enhancer elements to consider. Therefore, we used two additional approaches in order to narrow down which candidate enhancers were most likely to contribute substantially to ZNF217 hyperactivation. First, given that hub enhancers appear to be enriched in disease-associated SNPs, we compiled a list of significant SNPs from two breast cancer GWAS studies and determined whether these SNPs, or SNPs that were in linkage disequilibrium (LD) with them, overlapped with any candidate enhancers (analysis performed by Anna Li and Swneke Bailey)[26,93,94]. From this analysis, we came up with a list of four candidate enhancers which met the criteria (e2-e5). Secondly, we analyzed HiChIP data from multiple breast (MCF7, ZR75, T47D) and gastro-esophageal (AGS, HT55, OE33) cancer cell lines, and identified the enhancers which had the strongest loops (paired-end tags (PETs) > 50) to the ZNF217 promoter across the chosen cell lines (Fig. 9a-b). From this, we identified two additional candidate enhancers (e1 and e6). We also chose a final enhancer which represented one of strongest ATAC-seq and H3K27ac Chipmentation peaks in the region (e7). In total, we compiled a list of seven candidate enhancers to prioritize in our first round of CRISPRi (Fig. 9c).

3.3 Generation of stable dCas9-KRAB-MeCP2-expressing amplified GEA lines

Recently, CRISPRi—a variation of conventional CRISPR which makes use of a catalytically inactive Cas9 fused to two repressive domains, KRAB and MECP2—to knockdown

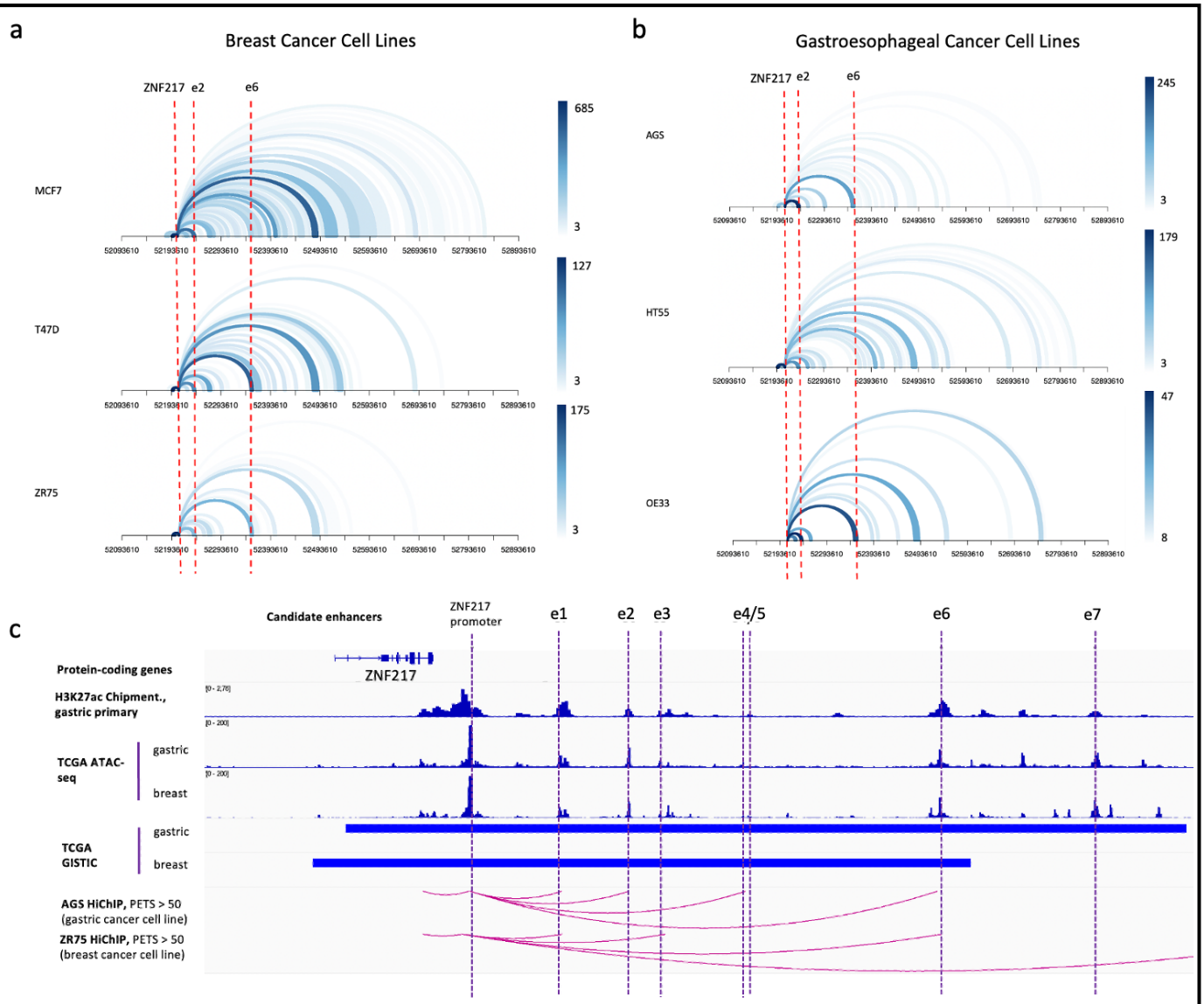


Fig. 9 Identification and prioritization of candidate enhancers in 20q13.2 region.

Heatmap of HiChIP data frrom (a) MCF7, T47D, and ZR75 (breast cancer cell lines) and (b) AGS (gastric), HT55 (colon), and OE33 (esophageal) cancer cell lines, looping from the ZNF217 promoter to elsewhere. Colour indicates number of paired end tags (PETs). e2 and e6 are candidate enhancers chosen for having strongly recurring loop to ZNF217 across cell lines. (c) Schema showing ATAC-seq (TCGA), H3K27ac Chipmentation, HiChIP, and GISTIC (TCGA) data used to identify candidate enhancers within the co-amplified non-coding region. HiChIP data PET cutoff was chosen to be 50, showing only the strongest loops in the region. Seven enhancers were chosen to profile from candidates above.

individual enhancers in an amplified, hierarchical enhancer cluster, avoiding previously reported complications of creating double-stranded breaks in amplified cell lines[85,86] (Fig. 10). CRISPRi-mediated gene and enhancer repression using a dCas9-KRAB-MECP2-containing vector with blasticidin resistance from Yeo et al. (2018) is executed in two steps (Fig. 11b). We first generated stable CRISPRi cell lines in two of the most highly 20q13.2-amplified GEA cell lines, MKN1 and MKN74, which will be henceforth designated MKN1i and MKN74i. We validated dCas9 mRNA and protein expression in the CRISPRi lines by RT-PCR (Fig. 11c) and IF (Fig. 11d), respectively.

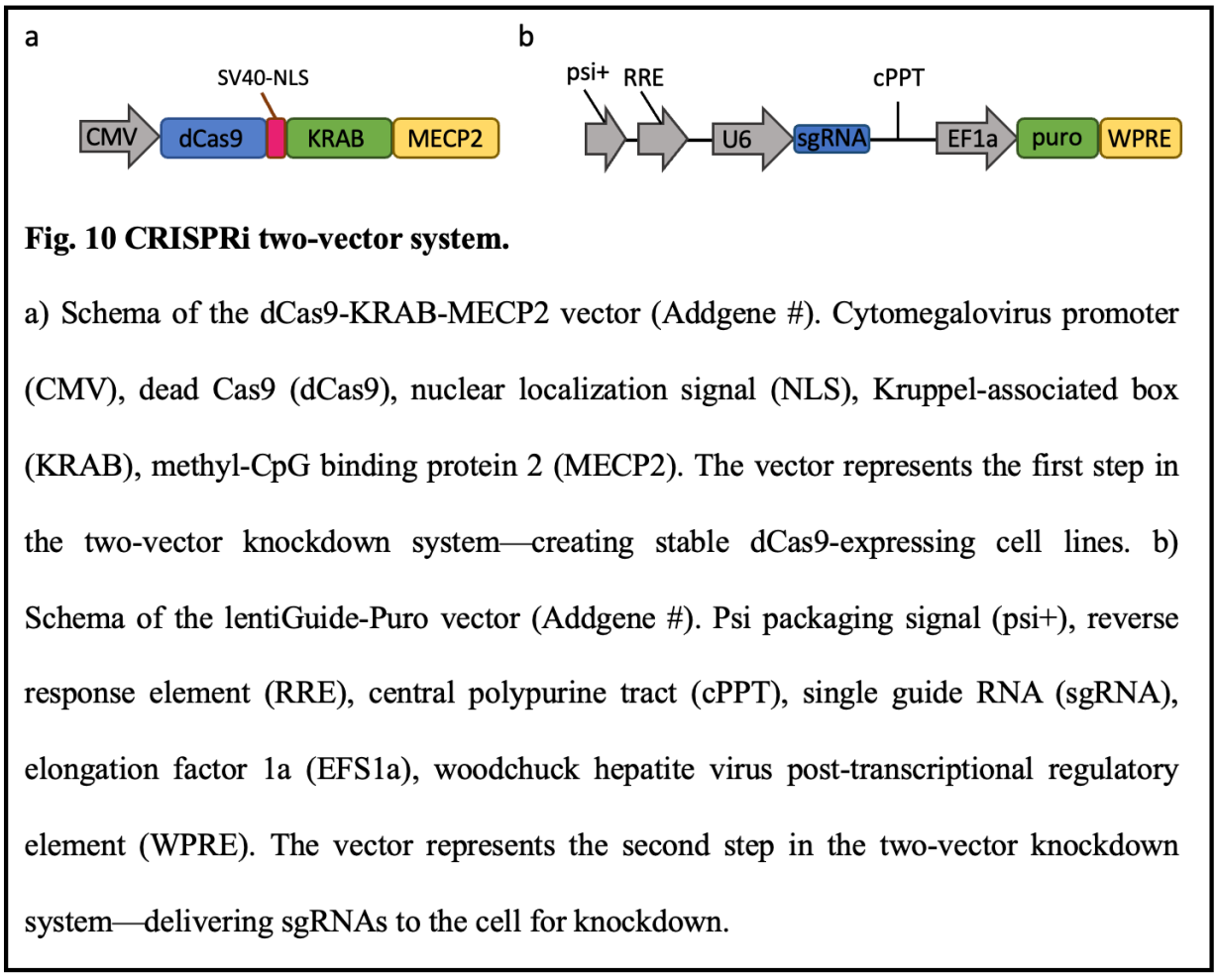


Fig. 10 CRISPRi two-vector system.

a) Schema of the dCas9-KRAB-MECP2 vector (Addgene #). Cytomegalovirus promoter (CMV), dead Cas9 (dCas9), nuclear localization signal (NLS), Kruppel-associated box (KRAB), methyl-CpG binding protein 2 (MECP2). The vector represents the first step in the two-vector knockdown system—creating stable dCas9-expressing cell lines. b) Schema of the lentiGuide-Puro vector (Addgene #). Psi packaging signal (psi+), reverse response element (RRE), central polypurine tract (cPPT), single guide RNA (sgRNA), elongation factor 1a (EFS1a), woodchuck hepatite virus post-transcriptional regulatory element (WPRE). The vector represents the second step in the two-vector knockdown system—delivering sgRNAs to the cell for knockdown.

3.4 Successful knockdown of positive control gene, *NET1* short isoform

The second step of the two-vector CRISPRi system was to design and clone sgRNAs into the second vector, lentiGuide-Puro[89], followed by infection of the stable CRISPRi lines and puromycin selection. In order to confirm that our CRISPRi system was capable of substantially and significantly knocking down genes, we chose a positive control gene with previously validated sgRNAs—the short isoform of *NET1*—to target first. In MKN74i, we observed an 84.7-93.7% knockdown by qPCR across three biological replicates (p-value = 0.00902) (Fig. 11e).

3.5 Knockdown of ZNF217 promoter and candidate enhancers

Next, we wanted to knock down the ZNF217 promoter and our candidate enhancers. We designed guides using either the Broad Institute CRISPick or IDT CRISPR guide design tools and chose 2-3 guides against each region of interest with the highest predicted efficiency. Guides against enhancers were designed to fall within a ~500 bp cut of the highest ATAC-seq peaks (data from TCGA), in the interest of targeting open chromatin. These guides were successfully cloned into the lentiGuide-puro vector and verified by Sanger sequencing (Fig. 13).

We subsequently tested guides against SNP enhancer 5, strong loop enhancers 1 and 6, as well as against the ZNF217 promoter in MKN74i. There was a small amount of KD in candidate SNP enhancer 5 (up to 17.1% KD), but virtually no KD in the strongly looping enhancers compared to the 93.7% KD of the NET1 positive control performed in parallel (Fig. 12ab). Targeting of the ZNF217 promoter was more efficient, but only yielded a maximal KD effect of 40.4% between the three tested guides (Fig. 12cd). It is possible that the fact of the enhancers showing low amounts of KD may indeed be reflective of their actual relationship to ZNF217

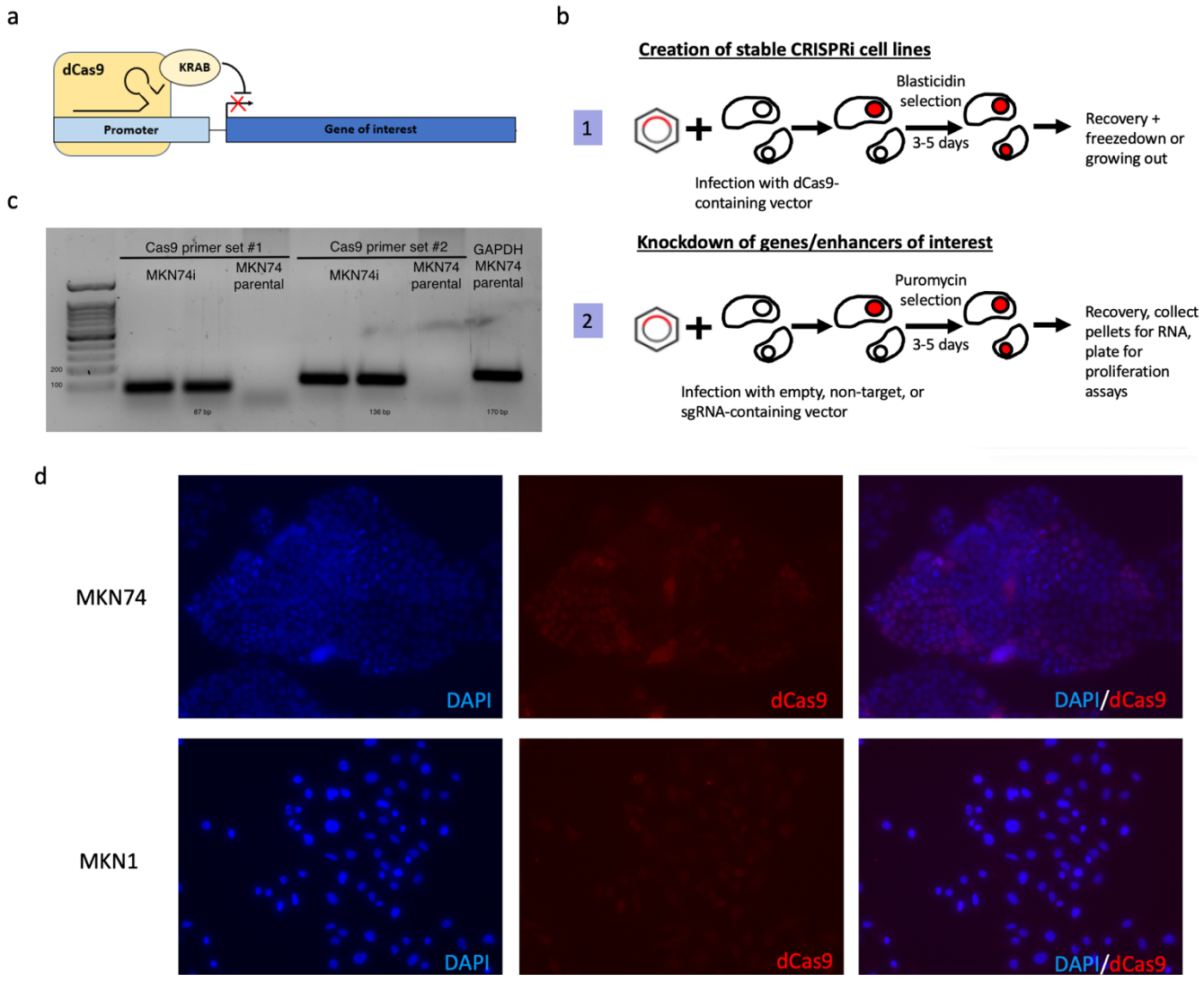


Fig. 11 Generation of stable dCas9-KRAB-MeCP2-expressing amplified GEA lines. (a) Schematic of CRISPRi acting on gene of interest, where catalytically inactive dCas9, fused to KRAB and MECP2 repressor domains, is targeted to the gene interest via a single guide RNA. (b) Schematic of experimental design for the creation of stable CRISPRi cell lines, followed by the knockdown of target regions. (c) RT-PCR of dCas9 transcript in stable CRISPRi cell lines, MKN74i and MKN1i. (d) IF of MKN74i and MKN1i showing expression of DAPI (blue), dCas9 (red), and merge. (e) qPCR of successful knockdown of positive control gene—the short isoform of NET1—compared to a non-target control sample in three biological replicates of MKN74 CRISPRi line (p-value = 0.00902).

expression; however, our difficulty in knocking down a second positive control gene, ZNF217, with guides that have not previously been validated for efficiency in the literature, indicates that the guides that we have designed or can design for our loci are not maximally efficient.

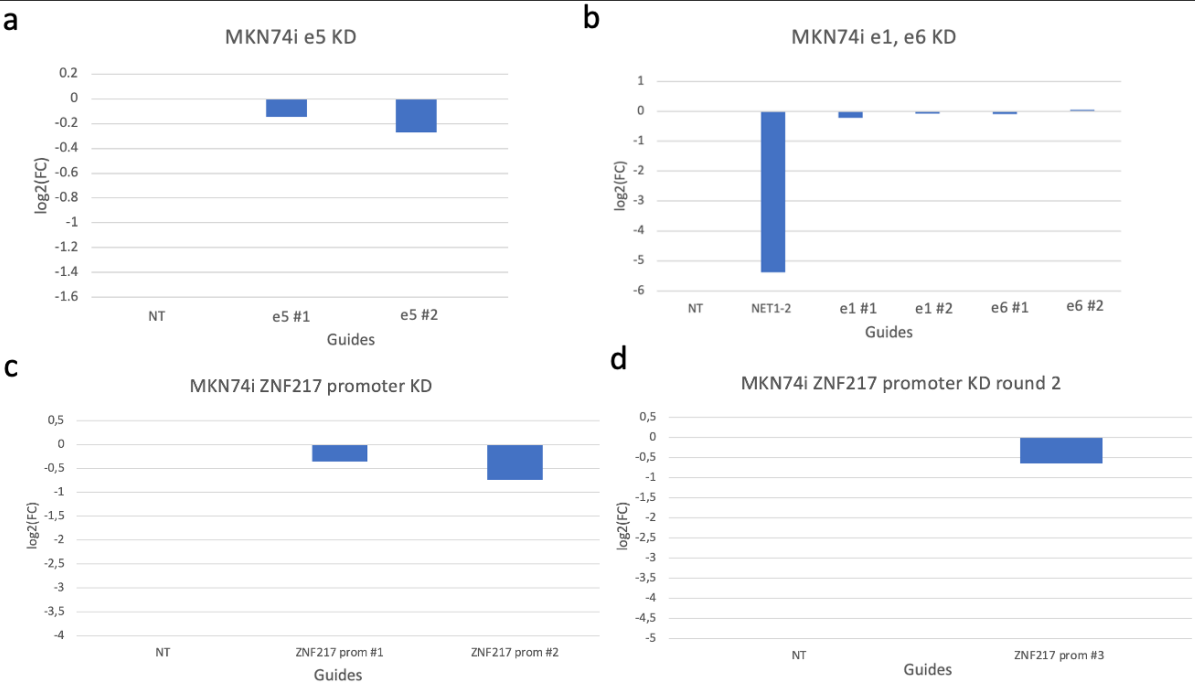


Fig. 12 Knockdown of candidate enhancers and of ZNF217 promoter.

(a) and (b) qPCR of knockdown of SNP enhancer e5, and strong loop enhancers e1 and e6 in MKN74i using two different sgRNAs, compared to non-target control. Samples were collected on day seven of puromycin selection. (c) and (d) qPCR of knockdown of ZNF217 promoter using three different sgRNAs compared to non-target control.

CHAPTER 4: DISCUSSION

In conclusion, we have identified a recurrently amplified non-coding region containing a super-enhancer which is co-selected for in gastric cancer alongside the previously described oncogene, ZNF217. Multiple candidate enhancers from this non-coding region appear to loop to the ZNF217 promoter common to several different gastroesophageal and breast cancer cell lines and tumors, suggesting that ZNF217 expression and overexpression may be functionally dependent on one or more of these. If this is the case, 20q13.2-amplified cancers may contribute to a growing body of examples in which structural variants targeting non-coding regions—specifically highly active regions such as super-enhancers—drive oncogene expression.

We have also successfully implemented CRISPRi in two gastric cancer cell lines, MKN74 and MKN1; however, we had difficulty knocking down our regions of interest, which have not previously been knocked down using CRISPRi. We considered whether the shortcomings we encountered were due to technical limitations of the CRISPRi system or difficulties pertaining to our biological system of choice.

One potential technical barrier that has been proposed to CRISPRi efficiency is the interaction of the dCas9-sgRNA complex with local chromatin structure, as sgRNA functionality appears to correlate with accessibility of chromatin[86]. Although we derived our enhancers of interest from overlapping HiChIP (Xiaoyang Zhang unpublished), ATAC-seq (TCGA) and our Chipmentation data in a various of gastric and breast cancer cell lines, open chromatin regions and therefore enhancer-promoter relationships may differ between cell types. One future experiment is to perform H3K27ac Chipmentation for MKN1 and MKN74 to confirm that active chromatin regions of interest correlate highly with active chromatin regions in related cancer cell lines, and alter our list of candidate enhancers accordingly. Furthermore, research has suggested that if Cas9

is successfully being expressed, CRISPR KD efficiency is mostly contingent on the potency of the guides used[87]; however, guide design tools do not take into account chromatin landscape in specific cell lines, nor can they integrate the locations of key binding sites for transcription factors involved in enhancer activity into their output, among other factors which may be relevant for guide efficiency. As such, we may be overestimating the reliability of guides predicted to be the most efficient. By designing more guides per enhancer (6-8 guides) and thus targeting a greater span of our region of interest, we may increase our chance of finding efficient guides.

In addition, it remains possible that amplification status impacts knockdown efficiency. Yuen et al. (2017) tested conventional CRISPR/Cas9's ability to knock out genes with copy number gains up to 21 copies and found no significant impairment in efficiency[95]. While it not is clear from this study whether these findings extend to amplifications above ~20 copies (highly-amplified loci may number in the 100s of the copies) or to the knockdown efficiency of CRISPRi, there have been recent reports of successfully knockdown of enhancers and genes at amplified loci using CRISPRi[71]—including of MYC highly amplified on ecDNAs[38]—indicating that, in principle, this model of using of CRISPRi to interrogate amplified regions is feasible. We have generated additional CRISPRi lines in non-amplified gastric cell lines expressing ZNF217 which will be used to confirm whether our guides can be successful in reducing ZNF217 expression in normal copy number conditions.

Additionally, some features of enhancer clusters—such as functional redundancy between enhancers which regulate the same target gene[25,29,96]—may make it difficult to resolve a strong functional relationship between any individual enhancer in our cluster and ZNF217 gene expression. While individual knockdown of enhancers may lead to small, but measurable effects on gene expression, redundancy in enhancers creates phenotypic robustness[29,96]. In order to fully

dissect the architecture of our enhancer cluster, it may be necessary to perform knockdowns of combinations of two or more enhancers at a time, allowing us to observe the result when other enhancers are not able to compensate for each other. As we now have stable CRISPRi cell lines made, and robust cloning, infection, and qPCR protocols set up in our lab, it will be easier to test more options such as this in parallel in the future. Furthermore, to differentiate this case from a case where our guides are not effectively targeting our loci, we could perform dCas9 and/or H3K9me3 ChIP-qPCR to ensure targeting to the correct location and subsequent heterochromatinization.

Beyond this, super-enhancers—defined as strong enhancers or groups of enhancers which span large genomic regions and bind unusually high levels of enhancer-associated TFs—and their associated oncogenes may be particularly susceptible to a class of potential cancer therapeutics known as BET inhibitors and degraders[70,71]. Across several amplified breast cancer cell lines, we can show that multiple stretches within the 20q13.2 region meet the criteria[97] for super-enhancers as identified by HOMER[98], a peak-calling software, and in the future, could expand this analysis to include our gastric cancer cell lines (Fig. 14). Until now, one barrier to success in BET inhibitor clinical trials has been the lack of predictive biomarkers which serve to stratify good responders from poor responders to the therapy. Thus, some 20q13.2 amplifications could additionally act as a way of identifying subsets of patients that are likely to respond to BET inhibitor therapy, in combination with other cancer drugs. Moreover, gastric cancers are not the only cancers with non-coding 20q13.2 co-amplifications as similar amplifications are present in breast and prostate cancers (data not shown). Further research will confirm whether our candidate enhancers are linked to ZNF217 expression as well as determining whether this relationship has the potential to

extend to other, non-gastric cancers with 20q13.2 amplifications, such as breast and prostate cancers.

CHAPTER 5: Conclusion and Future Directions

5.1 Conclusion

In this dissertation, we have sought to analyze characteristics of enhancers in the 20q13.2, ZNF217-co-amplified enhancer cluster—identified by our lab—which may be indicative of their functional importance within the cluster. This included analysis of HiChIP data from multiple breast and gastroesophageal lines which showed several highly recurrent loops that overlapped with strong signals from enhancer-associated features such as H3K27ac that were common to most of the HiChipped cancer cell lines. From this, along with an analysis of overlapping breast cancer GWAS SNPs with candidate enhancers, we designed and cloned sgRNAs against this priority list of seven candidate enhancers from the larger co-amplified region. We were also able to generate stable CRISPRi cell lines in MKN74 and MKN1—gastric cancer lines which can be infected a second time with a vector containing sgRNAs of interest. We successfully knocked down a positive control gene—validating the functionality of the CRISPRi system—but encountered inadequate knockdown efficiencies at another locus of interest: the ZNF217 promoter. We further performed knockdown experiments for three candidate enhancers, which appeared to have little to no impact on ZNF217 expression. Further experiments will be required to determine whether these enhancers have only a minimal functional relationship with ZNF217 expression or whether optimization of our CRISPRi protocol will be necessary for more reliable and robust knockdowns.

5.2 Future Directions

5.2.1 Enhancer-Associated Epigenetic Data in Cell Lines and Tumors

Our enhancers were chosen from epigenetic data collected in breast and gastric cancer tumors and cell lines; however, we do not have this data specifically in our cell lines, MKN1 and MKN74. Currently in the lab we are implementing H3K27ac Chipmentation in MKN74 and MKN1, which will provide more relevant data on the relative strength of enhancer-associated epigenetic marks in our cell lines. We are also interested in performing this in tumor samples. In the future, we may also set up HiChIP in our lab to visualize the exact looping configuration of our locus in our cell lines of interest.

5.2.2 Candidate Enhancer and ZNF217 Promoter Knockdown

There are a few different avenues for further experimentation with respect to knockdown of our candidate enhancers. There are 2-3 guides already cloned against the remaining candidate enhancers and which are ready to test in our cell lines. As well, we have designed 6-8 more guides against enhancer e6 and the ZNF217 promoter to evaluate whether using more guides per functional element is worthwhile. In addition, we will attempt simultaneous, pairwise targeting of two enhancers at a time to see if this produces a more striking effect on ZNF217 expression.

5.2.3 Recapitulation of Metastatic-Related Phenotypes in Gastric Cancer Cell Lines

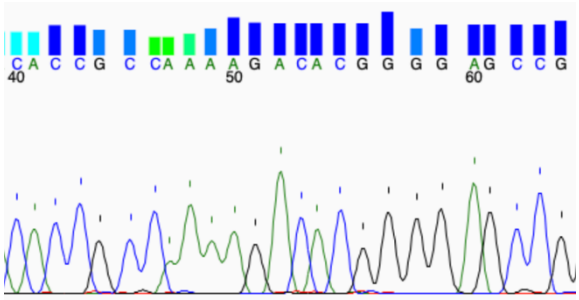
ZNF217 knockdown has been associated with multiple metastasis-related phenotypes which have been captured by *in vitro* assays, largely in breast cancer cell lines. Our lab has several of these set up, including the MTS Proliferation Assay (Abcam, ab197010), as well as the scratch wound and Boyden chamber assays for cell migration and invasion.

CHAPTER 6: SUPPLEMENTARY FIGURES AND TABLES

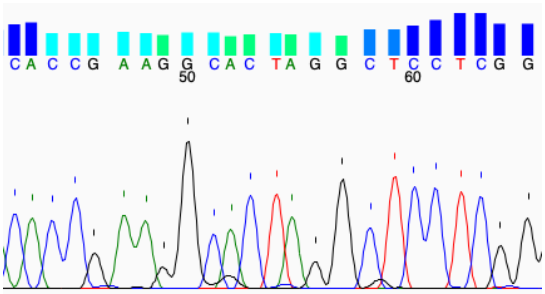
Fig. 13 Sanger sequencing traces for sgRNAs cloned into lentiGuide-Puro vector.

a

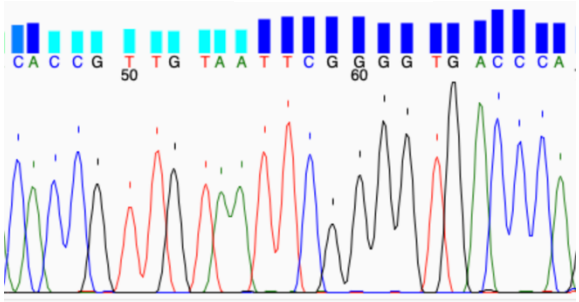
ZNF217 #1



ZNF217 #2

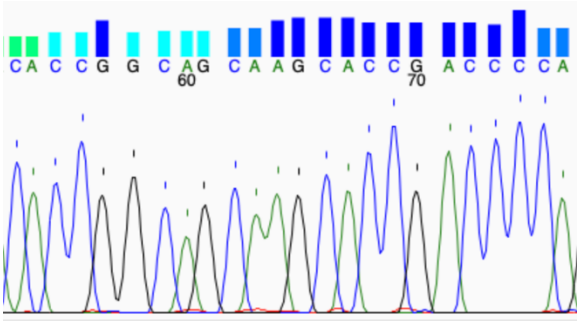


ZNF217 #3

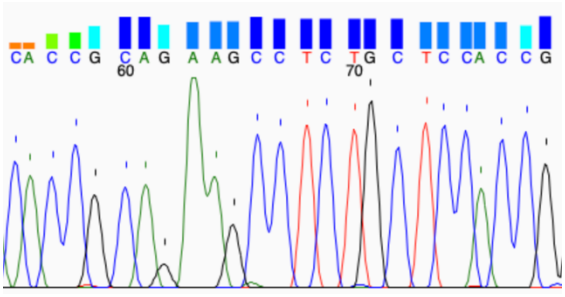


b

NET1 #1

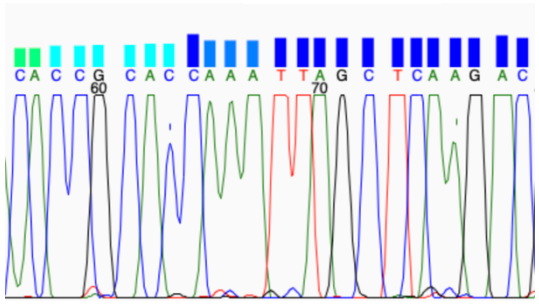


NET1 #2

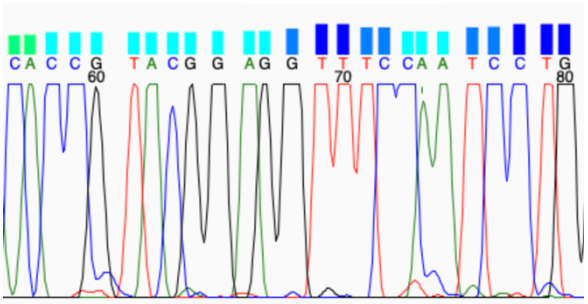


c

e1 #1

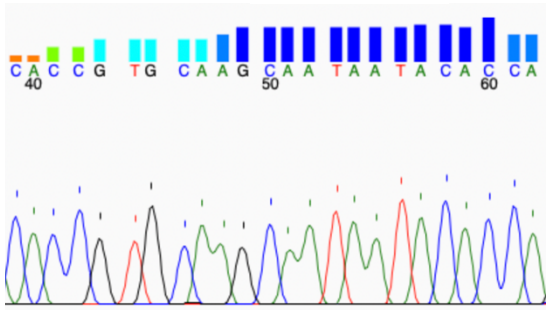


e1 #2

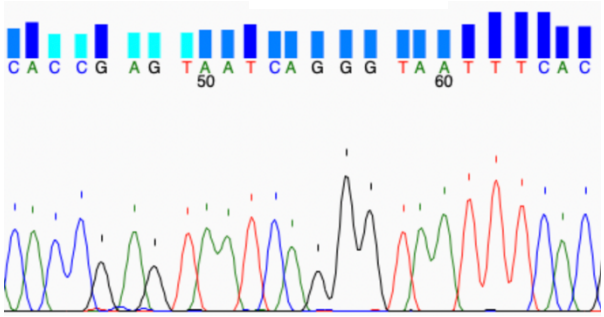


d

e2 #1

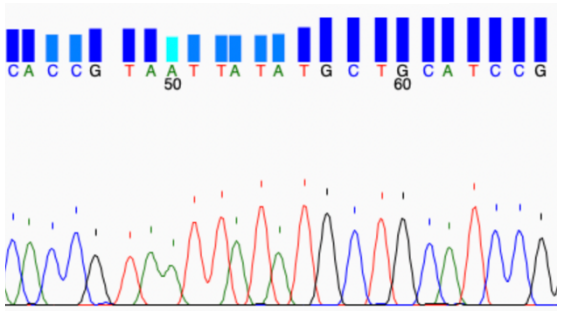


e2 #2

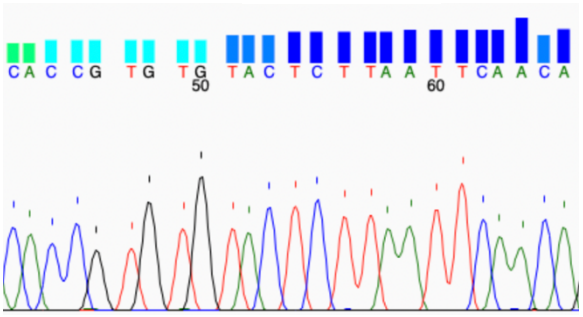


e

e3 #1

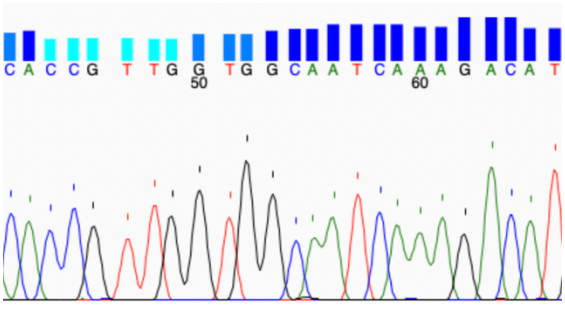


e3 #2

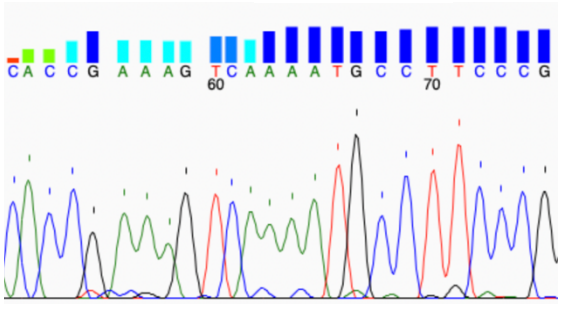


f

e4 #1

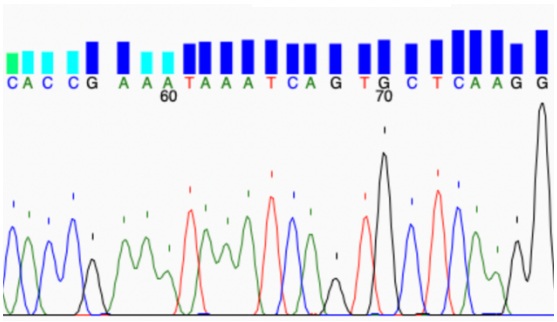


e4 #2

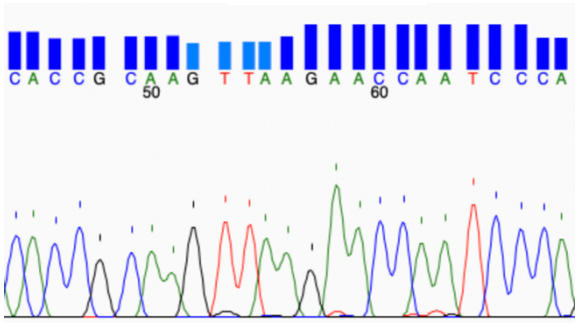


g

e5 #1

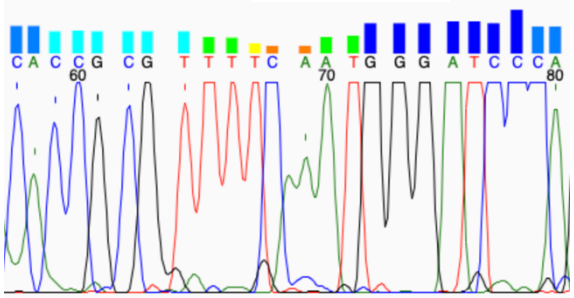


e5 #2

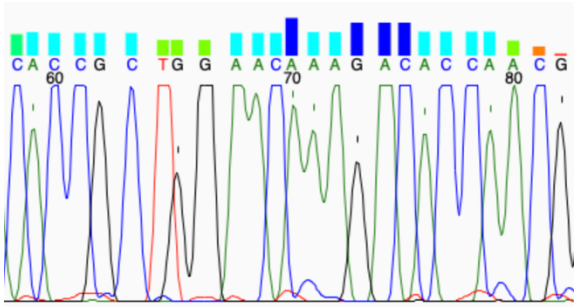


h

e6 #1

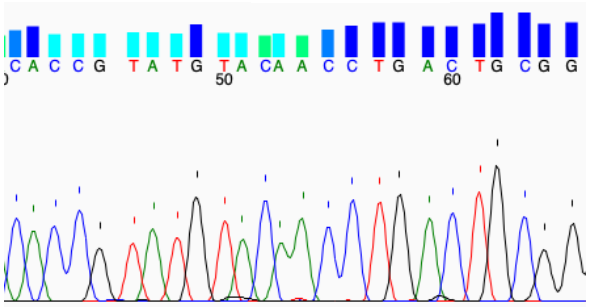


e6 #2



i

e7 #1



e7 #2

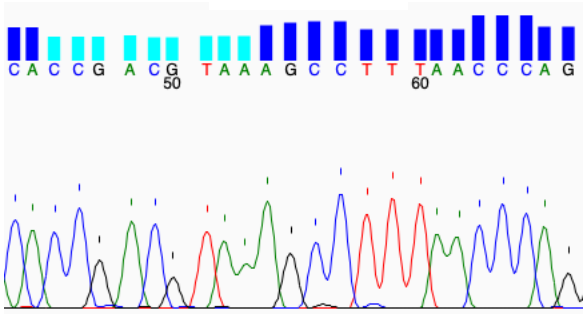


Fig. 14 Calling of 20q13.2 super-enhancer (SE) peaks in MCF7, T47D, and ZR75 breast cancer cell lines using HOMER.

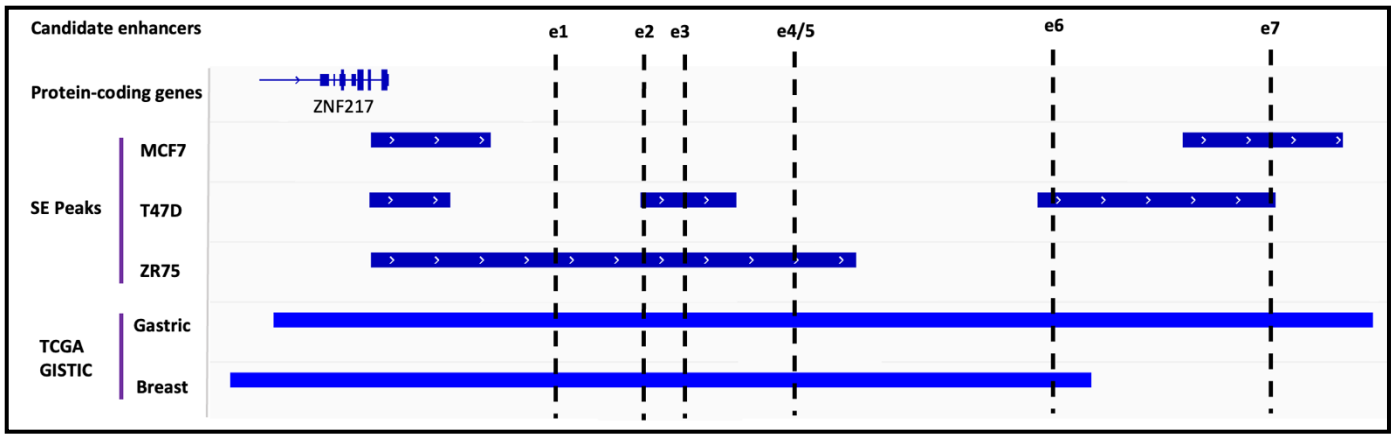


Table 1. Candidate enhancers

Enhancer	Chromosomal Location (hg19)	Notes/SNPs
e1	chr20:52237847-52238724	strong loop enhancer
e2	chr20:522259238-52259738	rs6068599 (in LD with rs2041278)
e3	chr20:52268812-52269312	rs2041278
e4	chr20:522943362-52294959	rs13036868 (in LD with rs13039563)
e5	chr20:52296701-52296850	rs13039563
e6	chr20:52354896-52355390	strong loop enhancer
e7	chr20:52402608-52403108	

Table 2. sgRNA sequences

sgRNA	Forward Sequence	Reverse Sequence
NET1 #1	CACCGGCAGCAAGCACCGACCCCA	AAACTGGGGTCGGTGCTTGCTGCC
NET1 #2	CACCGCAGAAGCCTCTGCTCCACCG	AAACCGGTGGAGCAGAGGCTTCTGC
ZNF217 #1	CACCGCCAAAAGACACGGGGAGCCG	AAACCGGCTCCCCGTGTCTTTTGGC
ZNF217 #2	CACCGAAGGCACTAGGCTCCTCGG	AAACCCGAGGAGCCTAGTGCCTTC
ZNF217 #3	CACCGGAGTCGCCGAGCGAGAACT	AAACAGTTCTCGCTCGGCGACTCC
e1 #1	CACCGCACCAAATTAGCTCAAGAC	AAACGTCTTGAGCTAATTTGGTGC
e1 #2	CACCGTACGGAGGTTTCCAATCCTG	AAACCAGGATTGGAAACCTCCGTAC
e1 #3	CACCGTGGTTATTACTAACACCAGT	AAACACTGGTGTTAGTAATAACCAC
e2 #1	CACCGTGCAAGCAATAATACACCA	AAACTGGTGTATTATTGCTTGCAC
e2 #2	CACCGAGTAATCAGGGTAATTTCAC	AAACGTGAAATTACCCTGATTACTC
e3 #1	CACCGTAATTATATGCTGCATCCG	AAACCGGATGCAGCATATAATTAC
e3 #2	CACCGTGTGTACTCTTAATTCAACA	AAACTGTTGAATTAAGAGTACACAC
e4 #1	CACCGTTGGTGGCAATCAAAGACAT	AAACATGTCTTTGATTGCCACCAAC
e4 #2	CACCGAAAGTCAAAATGCCTTCCCG	AAACCGGGAAGGCATTTTGACTTTC
e5 #1	CACCGAAATAAATCAGTGCTCAAGG	AAACCCTTGAGCACTGATTTATTTC
e5 #2	CACCGCAAGTTAAGAACCAATCCCA	AAACTGGGATTGGTTCTTAACTTGC
e6 #1	CACCGCGTTTTCAAATGGGATCCCA	AAACTGGGATCCCATTTGAAAACGC
e6 #2	CACCGAGACACCAACGGGGCCACCT	AAACAGGTGGCCCCGTTGGTGTCTC
e6 #3	CACCGCTGGAACAAAGACACCAACG	AAACCGTTGGTGTCTTTGTTCCAGC
e7 #1	CACCGTATGTACAACCTGACTGCGG	AAACCCGCAGTCAGGTTGTACATAC
e7 #2	CACCGACGTAAAGCCTTTAACCCAG	AAACCTGGGTTAAAGGCTTTACGTC

Table 3. qPCR primers

Gene	Forward	Reverse	Notes
NET1 short isoform	GTGGCACATGATGAGACTGG	CAAAACGTCTGACAGCTCCA	From Xiaoyang Zhang
ZNF217	CTTCAAGAGAGTGTAGTTATTGTGG	CAGAGATGTCTTCTGGGCTGCA	From Origene
ACTB	AGAGCTACGAGCTGCCTGAC	AGCACTGTGTTGGCGTACAG	
GAPDH	CGAGATCCCTCCAAAATCAA	TTCACACCCATGACGAACAT	From Xiaoyang Zhang
HPRT1	GACCAGTCAACAGGGGACAT	CCTGACCAAGGAAAGCAAAG	From Xiaoyang Zhang
Cas9 #1	AAGCCGGCTTCATCAAAAGG	TGGTGTTCATGCGTGAATCG	
Cas9 #2	AAAACCTCATCGCACAGCTC	TTGCTCAGTTGAAGCTTGGC	

CHAPTER 7: REFERENCES

1. Sitarz, R. *et al.* Gastric cancer: epidemiology, prevention, classification, and treatment. *Cancer Manag. Res.* **Volume 10**, 239–248 (2018).

2. Morgan, E. *et al.* The current and future incidence and mortality of gastric cancer in 185 countries, 2020–40: A population-based modelling study. *eClinicalMedicine* **47**, 101404 (2022).

3. Parsonnet, J., Friedman, G. D., Orentreich, N. & Vogelman, H. Risk for gastric cancer in people with CagA positive or CagA negative Helicobacter pylori infection. *Gut* **40**, 297–301 (1997).

4. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).

5. Shinozaki-Ushiku, A., Kunita, A. & Fukayama, M. Update on Epstein-Barr virus and gastric cancer (Review). *Int. J. Oncol.* **46**, 1421–1434 (2015).

6. Yamaguchi, N. & Kakizoe, T. Synergistic interaction between Helicobacter pylori gastritis and diet in gastric cancer. https://reader.elsevier.com/reader/sd/pii/S1470204500002254?token=96D5E72095F77EB986A46A49BF68DA36642D555B7E2FDEB44BD02E486D775016B3B6838CB9C39D649D3B3C70BE3E5204&originRegion=us-east-1&originCreation=20220915172259 doi:10.1016/S1470-2045(00)00225-4.

7. González, C. A. *et al.* Smoking and the risk of gastric cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *Int. J. Cancer* **107**, 629–634 (2003).

8. Duell, E. J. *et al.* Alcohol consumption and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Am. J. Clin. Nutr.* **94**, 1266–1275 (2011).

9. Machlowska, J., Baj, J., Sitarz, M., Maciejewski, R. & Sitarz, R. Gastric Cancer: Epidemiology, Risk Factors, Classification, Genomic Characteristics and Treatment Strategies. *Int. J. Mol. Sci.* **21**, 4012 (2020).

10. Lee, J. H. *et al.* Lauren Histologic Type Is the Most Important Factor Associated With Pattern of Recurrence Following Resection of Gastric Adenocarcinoma. *Ann. Surg.* **267**, 105–113 (2018).

11. Laurén, P. The Two Histological Main Types of Gastric Carcinoma: Diffuse and so-Called Intestinal-Type Carcinoma. *Acta Pathol. Microbiol. Scand.* **64**, 31–49 (1965).

12. Nagtegaal, I. D. *et al.* The 2019 WHO classification of tumours of the digestive system. *Histopathology* **76**, 182–188 (2020).

13. Di Giammartino, D. C., Polyzos, A. & Apostolou, E. Transcription factors: building hubs in the 3D space. *Cell Cycle* **19**, 2395–2410 (2020).

14. Snetkova, V. & Skok, J. A. Enhancer talk. *Epigenomics* **10**, 483–498 (2018).

15. Lim, B. & Levine, M. S. Enhancer-promoter communication: hubs or loops? *Curr. Opin. Genet. Dev.* **67**, 5–9 (2021).

16. Wagh, K., Garcia, D. A. & Upadhyaya, A. Phase separation in transcription factor dynamics and chromatin organization. *Curr. Opin. Struct. Biol.* **71**, 148–155 (2021).

17. Panigrahi, A. & O'Malley, B. W. Mechanisms of enhancer action: the known and the unknown. *Genome Biol.* **22**, 108 (2021).

18. Moorthy, S. D. *et al.* Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. *Genome Res.* **27**, 246–258 (2017).

19. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).

20. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).

21. Blobel, G. A., Higgs, D. R., Mitchell, J. A., Notani, D. & Young, R. A. Testing the super-enhancer concept. *Nat. Rev. Genet.* **22**, 749–755 (2021).

22. Dukler, N., Gulko, B., Huang, Y.-F. & Siepel, A. Is a super-enhancer greater than the sum of its parts? *Nat. Genet.* **49**, 2–3 (2017).

23. Hay, D. *et al.* Genetic dissection of the α-globin super-enhancer in vivo. *Nat. Genet.* **48**, 895–903 (2016).

24. Hong, J.-W., Hendrix, D. A. & Levine, M. S. Shadow Enhancers as a Source of Evolutionary Novelty. *Science* **321**, 1314–1314 (2008).

25. Lam, D. D. *et al.* Partially Redundant Enhancers Cooperatively Maintain Mammalian Pomc Expression Above a Critical Functional Threshold. *PLOS Genet.* **11**, e1004935 (2015).

26. Huang, J. *et al.* Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat. Commun.* **9**, 943 (2018).

27. Shin, H. Y. *et al.* Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nat. Genet.* **48**, 904–911 (2016).

28. Maurano, M. T. *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 (2012).

29. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).

30. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).

31. Kim, H. *et al.* Extrachromosomal DNA is associated with oncogene amplification and

poor outcome across multiple cancers. *Nat. Genet.* **52**, 891–897 (2020).

32. Paolella, B. R. *et al.* Copy-number and gene dependency analysis reveals partial copy loss of wild-type SF3B1 as a novel cancer vulnerability. *eLife* **6**, e23268 (2017).

33. Liang, L., Fang, J.-Y. & Xu, J. Gastric cancer and gene copy number variation: emerging cancer drivers for targeted therapy. *Oncogene* **35**, 1475–1482 (2016).

34. Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat. Genet.* **48**, 176–182 (2016).

35. Morton, A. R. *et al.* Functional Enhancers Shape Extrachromosomal Oncogene Amplifications. *Cell* **179**, 1330-1341.e13 (2019).

36. Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).

37. Zhu, Y. *et al.* Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell* **39**, 694-707.e7 (2021).

38. Hung, K. L. *et al.* ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature* **600**, 731–736 (2021).

39. Tanner, M. *et al.* Amplification of Chromosomal Region 20q13 in Invasive Breast Cancer: Prognostic Implications. *Clin. Cancer Res.* **1**, 1455–1461 (1995).

40. Courjal, F. *et al.* DNA amplifications at 20q13 and MDM2 define distinct subsets of evolved breast and ovarian tumours. *Br. J. Cancer* **74**, 1984–1989 (1996).

41. Kallioniemi, A. *et al.* Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Med. Sci.* 5 (1994).

42. Collins, C. *et al.* Positional cloning of ZNF217 and NABC1: Genes amplified at 20q13.2 and overexpressed in breast carcinoma. *Proc. Natl. Acad. Sci.* **95**, 8703–8708 (1998).

43. Rahman, M. T. *et al.* Prognostic and therapeutic impact of the chromosome 20q13.2 ZNF217 locus amplification in ovarian clear cell carcinoma. *Cancer* **118**, 2846–2857 (2012).

44. Bar-Shira, A. *et al.* Multiple Genes in Human 20q13 Chromosomal Region Are Involved in an Advanced Prostate Cancer Xenograft. 5.

45. Rooney, P. H. *et al.* The candidate oncogene ZNF217 is frequently amplified in colon cancer. *J. Pathol.* **204**, 282–288 (2004).

46. Mao, X. Overexpression of ZNF217 in glioblastoma contributes to the maintenance of glioma stem cells regulated by hypoxia-inducible factors. *Lab. Invest.* **91**, 11 (2011).

47. Weiss, M. M. *et al.* Determination of amplicon boundaries at 20q13.2 in tissue samples of human gastric adenocarcinomas by high-resolution microarray comparative genomic hybridization. *J. Pathol.* **200**, 320–326 (2003).

48. Nunez et al. - 2011 - The Multi-zinc Finger Protein ZNF217 Contacts DNA .pdf.

49. Lee, D.-F., Walsh, M. J. & Aguiló, F. ZNF217/ZFP217 Meets Chromatin and RNA. *Trends Biochem. Sci.* **41**, 986–988 (2016).

50. Krig, S. R. *et al.* Identification of Genes Directly Regulated by the Oncogene ZNF217 Using Chromatin Immunoprecipitation (ChIP)-Chip Assays. *J. Biol. Chem.* **282**, 9703–9712 (2007).

51. Thillainadesan, G. *et al.* Genome Analysis Identifies the p15 ^{ink4b} Tumor Suppressor as a Direct Target of the ZNF217/CoREST Complex. *Mol. Cell. Biol.* **28**, 6066–6077 (2008).

52. Krig, S. R. *et al.* ZNF217, a candidate breast cancer oncogene amplified at 20q13, regulates expression of the ErbB3 receptor tyrosine kinase in breast cancer cells. *Oncogene* **29**, 5500–5510 (2010).

53. Forneris, F., Binda, C., Battaglioli, E. & Mattevi, A. LSD1: oxidative chemistry for

multifaceted functions in chromatin regulation. *Trends Biochem. Sci.* **33**, 181–189 (2008).

54. Perillo, B., Tramontano, A., Pezone, A. & Migliaccio, A. LSD1: more than demethylation of histone lysine residues. *Exp. Mol. Med.* **52**, 1936–1947 (2020).

55. Aguilo, F. *et al.* Coordination of m 6 A mRNA Methylation and Gene Transcription by ZFP217 Regulates Pluripotency and Reprogramming. *Cell Stem Cell* **17**, 689–704 (2015).

56. Nonet, G. H. *et al.* The ZNF217 Gene Amplified in Breast Cancers Promotes Immortalization of Human Mammary Epithelial Cells1. *Cancer Res.* **61**, 1250–1254 (2001).

57. Huang, G. *et al.* ZNF217 suppresses cell death associated with chemotherapy and telomere dysfunction. *Hum. Mol. Genet.* **14**, 3219–3225 (2005).

58. Sun, Y. *et al.* The eukaryotic translation elongation factor eEF1A2 induces neoplastic properties and mediates tumorigenic effects of ZNF217 in precursor cells of human ovarian carcinomas. *Int. J. Cancer* **123**, 1761–1769 (2008).

59. Thollet, A. *et al.* ZNF217 confers resistance to the pro-apoptotic signals of paclitaxel and aberrant expression of Aurora-A in breast cancer cells. *Mol. Cancer* **9**, 291 (2010).

60. Vendrell, J. A. *et al.* ZNF217 Is a Marker of Poor Prognosis in Breast Cancer That Drives Epithelial–Mesenchymal Transition and Invasion. *Cancer Res.* **72**, 3593–3606 (2012).

61. Littlepage, L. E. *et al.* The Transcription Factor ZNF217 Is a Prognostic Biomarker and Therapeutic Target during Breast Cancer Progression. *Cancer Discov.* **2**, 638–651 (2012).

62. Nijjar, T. *et al.* Accumulation and altered localization of telomere-associated protein TRF2 in immortally transformed and tumor-derived human breast cells. *Oncogene* **24**, 3369–3376 (2005).

63. Bellanger, A. *et al.* Exploring the Significance of the Exon 4-Skipping Isoform of the ZNF217 Oncogene in Breast Cancer. *Front. Oncol.* **11**, 647269 (2021).

64. Kong, D., Li, Y., Wang, Z. & Sarkar, F. Cancer Stem Cells and Epithelial-to-Mesenchymal Transition (EMT)-Phenotypic Cells: Are They Cousins or Twins? *Cancers* **3**, 716–729 (2011).

65. Tang, Y., Wang, S., Jiang, J. & Liang, X. Links between cancer stem cells and epithelial– mesenchymal transition. *OncoTargets Ther.* 2973 (2015) doi:10.2147/OTT.S91863.

66. Zhang, C. *et al.* Hypoxia-inducible factors regulate pluripotency factor expression by ZNF217- and ALKBH5-mediated modulation of RNA methylation in breast cancer cells. *Oncotarget* **7**, 64527–64542 (2016).

67. Shida, A. *et al.* Prognostic Significance of ZNF217 Expression in Gastric Carcinoma. *ANTICANCER Res.* 5 (2014).

68. Vendrell, J. A. *et al.* Evaluating ZNF217 mRNA Expression Levels as a Predictor of Response to Endocrine Therapy in ER+ Breast Cancer. *Front. Pharmacol.* **9**, 1581 (2019).

69. Doroshow, D. B., Eder, J. P. & LoRusso, P. M. BET inhibitors: a novel epigenetic approach. *Ann. Oncol.* **28**, 1776–1787 (2017).

70. Lovén, J. *et al.* Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. *Cell* **153**, 320–334 (2013).

71. Liu, Y. *et al.* A predominant enhancer co-amplified with the SOX2 oncogene is necessary and sufficient for its expression in squamous cancer. 30.

72. Boija, A., Klein, I. A. & Young, R. A. Biomolecular Condensates and Cancer. *Cancer Cell* **39**, 174–192 (2021).

73. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).

74. Han, X. *et al.* Roles of the BRD4 short isoform in phase separation and active gene transcription. *Nat. Struct. Mol. Biol.* **27**, 333–341 (2020).

75. Shrinivas, K. *et al.* Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell* **75**, 549-561.e7 (2019).

76. Purshouse, K. *et al. Oncogene expression from extrachromosomal DNA is driven by copy number amplification and does not require spatial clustering*. http://biorxiv.org/lookup/doi/10.1101/2022.01.29.478046 (2022) doi:10.1101/2022.01.29.478046.

77. Shorstova, T., Foulkes, W. D. & Witcher, M. Achieving clinical success with BET inhibitors as anti-cancer agents. *Br. J. Cancer* **124**, 1478–1490 (2021).

78. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

79. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183 (2013).

80. Yeo, N. C. *et al.* An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat. Methods* **15**, 611–616 (2018).

81. Urrutia, R. KRAB-containing zinc-finger repressor proteins. *Genome Biol.* 8 (2003).

82. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).

83. Evers, B. *et al.* CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34**, 631–633 (2016).

84. Klein, J. C., Chen, W., Gasperini, M. & Shendure, J. Identifying Novel Enhancer Elements with CRISPR-Based Screens. *ACS Chem. Biol.* **13**, 326–332 (2018).

85. Aguirre, A. J. *et al.* Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov.* **6**, 914–929 (2016).

86. Munoz, D. M. *et al.* CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. *Cancer Discov.* **6**, 900–913 (2016).

87. Cohen, P. A., Donini, C. F., Nguyen, N. T., Lincet, H. & Vendrell, J. A. The dark side of ZNF217, a key regulator of tumorigenesis with powerful biomarker value. *Oncotarget* **6**, 41566–41581 (2015).

88. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).

89. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

90. Vandesompele, J., Preter, K. D., Roy, N. V. & Paepe, A. D. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. 12.

91. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).

92. Lareau, C. A. & Aryee, M. J. hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat. Methods* **15**, 155–156 (2018).

93. kConFab Investigators *et al.* Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat. Genet.* **52**, 572–581 (2020).

94. Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci.

Nature **551**, 92–94 (2017).

95. Yuen, G. *et al.* CRISPR/Cas9-mediated gene knockout is insensitive to target copy number but is dependent on guide RNA potency and Cas9/sgRNA threshold expression level. *Nucleic Acids Res.* **45**, 12039–12053 (2017).

96. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490–493 (2010).

97. Whyte, W. A. *et al.* Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* **153**, 307–319 (2013).

98. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).