Article

Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs

https://doi.org/10.1038/s41586-019-1668-3

Received: 3 December 2017

Accepted: 10 September 2019

Published online: 16 October 2019

William A. Flavahan[1,2,11], Yotam Drier[1,2,10,11]*, Sarah E. Johnstone[1,2], Matthew L. Hemming[3,4], Daniel R. Tarjan[1,2], Esmat Hegazi[1,2], Sarah J. Shareef[1,2], Nauman M. Javed[1,2], Chandrajit P. Raut[5], Benjamin K. Eschle[6], Prafulla C. Gokhale[6], Jason L. Hornick[7], Ewa T. Sicinska[8], George D. Demetri[3,4,9]* & Bradley E. Bernstein[1,2,9]*

Epigenetic aberrations are widespread in cancer, yet the underlying mechanisms and causality remain poorly understood[1–3]. A subset of gastrointestinal stromal tumours (GISTs) lack canonical kinase mutations but instead have succinate dehydrogenase (SDH) deficiency and global DNA hyper-methylation[4,5]. Here, we associate this hyper-methylation with changes in genome topology that activate oncogenic programs. To investigate epigenetic alterations systematically, we mapped DNA methylation, CTCF insulators, enhancers, and chromosome topology in *KIT*-mutant, *PDGFRA*-mutant and SDH-deficient GISTs. Although these respective subtypes shared similar enhancer landscapes, we identified hundreds of putative insulators where DNA methylation replaced CTCF binding in SDH-deficient GISTs. We focused on a disrupted insulator that normally partitions a core GIST super-enhancer from the *FGF4* oncogene. Recurrent loss of this insulator alters locus topology in SDH-deficient GISTs, allowing aberrant physical interaction between enhancer and oncogene. CRISPR-mediated excision of the corresponding CTCF motifs in an SDH-intact GIST model disrupted the boundary between enhancer and oncogene, and strongly upregulated *FGF4* expression. We also identified a second recurrent insulator loss event near the *KIT* oncogene, which is also highly expressed across SDH-deficient GISTs. Finally, we established a patient-derived xenograft (PDX) from an SDH-deficient GIST that faithfully maintains the epigenetics of the parental tumour, including hypermethylation and insulator defects. This PDX model is highly sensitive to FGF receptor (FGFR) inhibition, and more so to combined FGFR and KIT inhibition, validating the functional significance of the underlying epigenetic lesions. Our study reveals how epigenetic alterations can drive oncogenic programs in the absence of canonical kinase mutations, with implications for mechanistic targeting of aberrant pathways in cancers.

The human genome is partitioned into physical domains, often termed topologically associated domains (TADs), by chromosomal boundaries established by the DNA-binding insulator protein CTCF and cohesin[6–9]. Many proto-oncogenes and master regulators are isolated in such domains and thus protected from promiscuous enhancer interactions[10].

Mutations of tricarboxylic-acid-cycle-related enzymes, including SDH and isocitrate dehydrogenase (IDH), are initiating events in many tumour types[1,4,5]. These lesions cause accumulation of succinate and 2-hydroxyglutarate, respectively, which inhibit demethylases and are associated with DNA hypermethylation and other epigenetic alterations[5,11,12]. The CTCF insulator is methylation-sensitive and may be displaced by DNA methylation[13–15]. We previously showed that the *PDGFRA* oncogene is aberrantly activated by insulator defects in *IDH*-mutant glioma[16].

We hypothesized that SDH deficiency alters chromosome topology to drive GIST tumorigenesis (Fig. 1a). GISTs are the most common gastrointestinal tract sarcoma. They are typically caused by gain-of-function mutations of the *KIT* or *PDGFRA* oncogenes that render these receptor tyrosine kinases (RTKs) active and ligand-independent[17]. However, approximately 15% of GISTs lack these defining mutations, and have instead lost SDH expression due to mutation or transcriptional

[1]Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Center for Sarcoma and Bone Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [4]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School Boston, Boston, MA, USA. [5]Department of Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [6]Experimental Therapeutics Core, Belfer Center for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, MA, USA. [7]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [8]Department of Oncologic Pathology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA. [9]Ludwig Center at Harvard, Harvard Medical School, Boston, MA, USA. [10]Present address: The Lautenberg Center for Immunology and Cancer Research, IMRIC, Faculty of Medicine, The Hebrew University, Jerusalem, Israel. [11]These authors contributed equally: William A. Flavahan, Yotam Drier. *e-mail: yotam.drier@mail.huji.ac.il; George_Demetri@dfci.harvard.edu; Bernstein.Bradley@mgh.harvard.edu

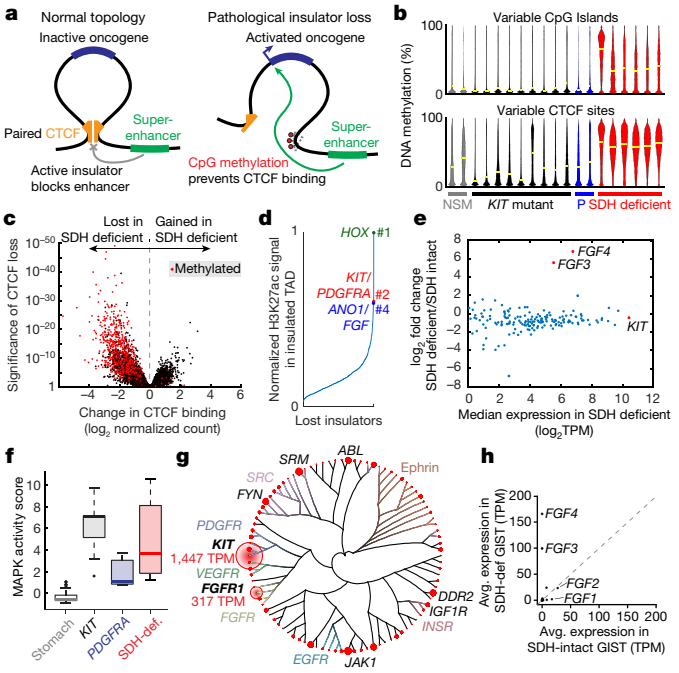


Fig. 1 | Insulator dysfunction in SDH-deficient GISTs. a, Proposed mechanism of epigenetic oncogene activation. Left, oncogene shielded from super-enhancer by CTCF insulator, which creates a topological boundary. Right, CTCF insulator displaced by DNA methylation, allowing the super-enhancer to contact and induce the oncogene. **b**, Violin plots depict DNA methylation levels over the 10,000 most variable CpG island promoters (top) and CTCF sites (bottom) in normal stomach muscle (NSM; $n = 2$), and *KIT* mutant ($n = 9$), *PDGFRA* mutant (P; $n = 2$) and SDH-deficient GISTs ($n = 6$). Yellow bars indicate mean. **c**, Volcano plot depicts differential CTCF occupancy between SDH-deficient ($n = 6$) and SDH-intact ($n = 8$) GISTs. Sites that gain DNA methylation in SDH-deficient GISTs are indicated in red (>25% increase, two-sided *t*-test false discovery rate (FDR) <5%). **d**, Plot depicts H3K27ac peaks near lost CTCF insulators (*y* axis) rank ordered by signal strength. **e**, Scatter plot depicts genes (points) separated from a super-enhancer by a CTCF loop anchor that is lost in SDH-deficient GIST. Genes are positioned according to their relative (*y* axis) and absolute median expression (*x* axis) in SDH-deficient GISTs. Potentially deregulated gene targets (outliers) include oncogenes *FGF3*, *FGF4* and *KIT* (red); see also Supplementary Information. TPM, transcripts per million. **f**, Box plot depicts average expression of MAPK signature genes in RNA-seq data for normal stomach ($n = 262$), and *KIT* mutant ($n = 10$), *PDGFRA* mutant ($n = 3$) and SDH-deficient GISTs ($n = 8$). Boxes depict 25th, 50th and 75th percentiles, and whiskers depict extreme values. **g**, Radial phylogenetic tree depicts tyrosine kinase gene expression in SDH-deficient GISTs. Each branch is one tyrosine kinase, arranged by similarity, and with major families depicted by colour. The area of each red circle is proportional to the average expression of the kinase. **h**, Scatter plot depicts average expression of FGF ligands in SDH-intact (*x* axis) and SDH-deficient (*y* axis) GISTs. *FGF3* and *FGF4* are highly expressed in SDH-deficient GISTs (bold). For all panels, *n* values indicate number of biologically independent specimens.

silencing of *SDH* subunit genes (*SDHA–D*)[18]. We collected an initial cohort of clinically defined specimens, including 11 *KIT*-mutant, 2 *PDGFRA*-mutant and 8 SDH-deficient tumours (Supplementary Table 1). We used hybrid-selection bisulfite sequencing to profile DNA methylation of over 160,000 CTCF sites and representative promoters in 17 of these tumours and 2 normal stomach muscle samples (see Methods). Consistent with previous reports[5], SDH-deficient GISTs exhibited CpG island hypermethylation (Fig. 1b). In addition, a substantial fraction of CTCF sites were methylated in this GIST subtype (Fig. 1b).

We next identified candidate insulators and enhancers in these tumours by mapping CTCF and histone H3 acetylated at lysine 27 (H3K27ac) by chromatin immunoprecipitation and sequencing (ChIP–seq). Overall patterns of enhancer acetylation were largely consistent across GISTs, relative to gastrointestinal carcinomas (Extended Data Fig. 1a). By contrast, comparison of genome-wide CTCF binding profiles revealed that approximately 5% of sites were specifically lost in SDH-deficient GISTs (Fig. 1c). CTCF loss was accompanied by notable increases in DNA methylation at these sites (Fig. 1c and Extended Data Fig. 1b, c). Given that DNA methylation has been established to prevent CTCF binding[13–15], this suggests that hypermethylation displaces CTCF from hundreds of candidate insulators in SDH-deficient tumours.

To investigate the impact of CTCF loss on genome topology, we used HiC to map TADs and TAD boundaries genome-wide in GIST-T1, a human cell line with an oncogenic *KIT* mutation and intact SDH expression[19]. We also used HiChIP to map CTCF loops and loop anchors, which correspond to TADs and boundaries, respectively[20] (Extended Data Fig. 1d). We used these maps to predict insulator losses likely to alter topology and gene expression. Of the 1,236 sites that lose CTCF and gain methylation in SDH-deficient GISTs, 688 corresponded to loop anchors. We reasoned that disruption of these loop anchors could alter topology and, in certain cases, permit aberrant enhancer–promoter interactions (Fig. 1a). Therefore, we further curated this list using enhancer maps and expression data. This highlighted 60 CTCF loop anchors that would normally have partitioned a large 'super-enhancer' from a gene, but that were lost in SDH-deficient GISTs (Fig. 1d, e and Supplementary Table 2). Top hits included lost CTCF insulators in the *FGF3* and *FGF4* locus (chromosome 11q13) and the *KIT* locus (chromosome 4q12) (Extended Data Fig. 2a, b).

Although SDH-deficient GISTs lack *KIT* or *PDGFRA* mutations[18], our insulator analysis raised the possibility that RTKs may instead be epigenetically deregulated. This prompted us to examine the expression of RTKs, ligands and downstream signalling programs. First, we found that a signature for mitogen-activated protein kinase (MAPK) targets is highly expressed and suggestive of active RTK signalling in SDH-deficient GISTs (Fig. 1f; see Methods). Second, a systematic analysis of tyrosine kinase gene expression revealed that *KIT* and FGF receptor 1 (*FGFR1*) are the most highly expressed RTKs in SDH-deficient GISTs (Fig. 1g). Third, we found that *FGF3* and *FGF4* were expressed at remarkably high levels, and were both specific to the SDH-deficient subtype (Fig. 1h). *FGF3* and *FGF4* are established oncogenes[21], and FGF signalling could help to explain the poor efficacy of KIT inhibitors in SDH-deficient GISTs[22,23]. We therefore investigated the mechanisms underlying this striking and specific upregulation of FGF ligands.

FGF3 and *FGF4* reside in an approximately 250 kb TAD flanked by boundaries that contain CTCF-binding sites (Fig. 2a). The adjacent TAD on the 11q side contains a large cluster of enhancers or super-enhancer. This super-enhancer overlaps the gene *ANO1*, which encodes the GIST clinical biomarker also known as DOG-1 ('discovered on GIST-1')[24]. The super-enhancer is highly acetylated and *ANO1* is highly expressed in all GIST subtypes (Extended Data Fig. 2a). Notably, the TAD boundary that partitions this super-enhancer from the FGF genes, which we refer to as the 'FGF insulator', contains several CTCF-binding sites (Fig. 2a).

We hypothesized that disruption of CTCF binding could compromise the FGF insulator and allow the *ANO1* super-enhancer to contact and activate the FGF genes. The FGF insulator contains two strong and several weak CTCF-binding sites. Although these sites are consistently bound in *KIT*- and *PDGFRA*-mutant tumours and normal stomach muscle control samples, all five are markedly reduced in SDH-deficient GISTs (Extended Data Fig. 2c). The strongest CTCF-binding site, which is closest to the *ANO1* super-enhancer, is almost completely lost in the SDH-deficient samples (Extended Data Fig. 2d). Consistently, it is methylated specifically in SDH-deficient tumours. This suggests that the FGF insulator has switched to a methylated state that occludes CTCF binding.

To assess the impact of CTCF loss on boundary integrity, we performed circularized chromatin conformation capture sequencing (4C-seq) on four SDH-intact and three SDH-deficient GISTs. We designed a 'viewpoint' primer that enabled us to quantify contacts between a central position in the *ANO1* super-enhancer and other genomic positions at high resolution (Fig. 2b). In SDH-intact tumours and stomach

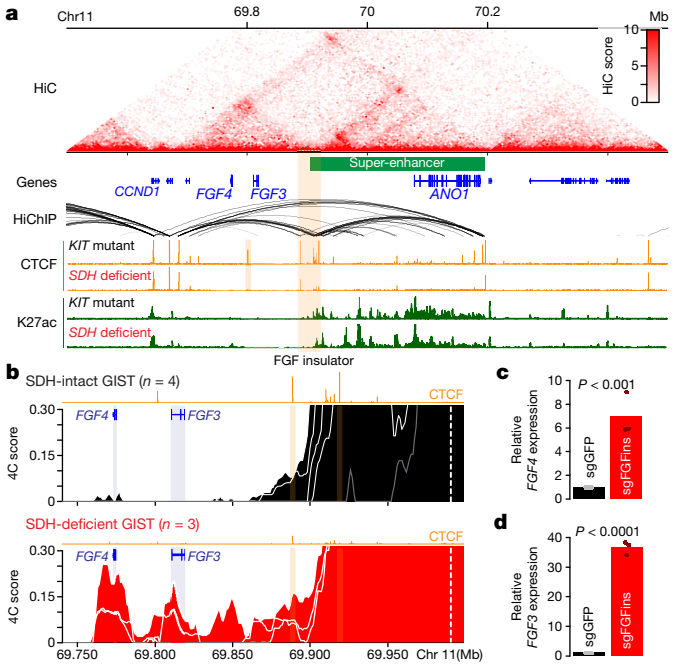


Fig. 2 | *FGF3–FGF4* locus topology reorganized in SDH-deficient GISTs. **a**, Genomic views of the *FGF3–FGF4* and *ANO1* loci depict baseline chromosome topology (HiC, red), genes (blue), CTCF–CTCF loop interactions (HiChIP, arcs, with darkness indicating significance), CTCF binding (ChIP–seq, orange) and candidate enhancers (H3K27ac ChIP–seq, green). HiC/HiChIP data are for the SDH-intact model GIST-T1, whereas CTCF and H3K27ac data are for representative clinical specimens (see also Extended Data Fig. 2). *ANO1* super-enhancer (green bar) and FGF insulator (orange shading) are indicated. **b**, Traces depict 4C-seq interaction frequency (*y* axis) between the *ANO1* super-enhancer viewpoint primer (dashed white line) and genomic positions in the *FGF3–FGF4–ANO1* locus (*x* axis). Data are shown for SDH-intact GISTs ($n = 4$; top), normal stomach muscle ($n = 1$; grey line, top) and SDH-deficient GISTs ($n = 3$; bottom). CTCF binding profiles for representative SDH-intact (top) and SDH-deficient (bottom) tumours are also shown (orange). Genes (blue) and CTCF sites in the FGF insulator (orange) are highlighted. **c**, **d**, Plots depict relative *FGF4* (**c**) and *FGF3* (**d**) expression in GIST-T1 cells expressing CRISPR–Cas9 and control sgRNA (black) or sgRNAs targeting the two CTCF sites in the FGF insulator (red). Bars indicate mean of three biologically independent replicates (dots). *P* values by two-sided *t*-test.

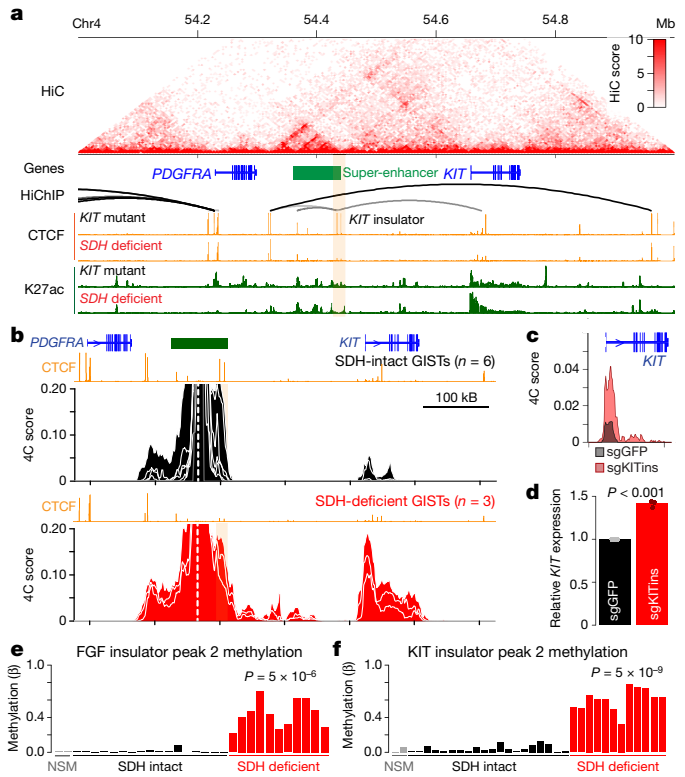


Fig. 3 | *KIT–PDGFRA* locus topology reorganized in SDH-deficient GISTs. **a**, Genomic views of *PDGFRA* and *KIT* loci depict baseline chromosome topology (HiC, red), genes (blue), CTCF–CTCF loop interactions (HiChIP, arcs), CTCF binding (ChIP–seq, orange) and candidate enhancers (H3K27ac ChIP–seq, green). HiC/HiChIP data are for the SDH-intact GIST model GIST-T1, whereas CTCF and H3K27ac data are for representative clinical specimens (see also Extended Data Fig. 2). KIT super-enhancer (green bar) and KIT insulator (orange shading) are indicated. **b**, Traces depict 4C-seq interaction frequency (*y* axis) between the KIT super-enhancer viewpoint primer (dashed white line) and genomic positions in the *KIT–PDGFRA* locus (*x* axis). Data are shown for SDH-intact GISTs ($n = 6$, top) and SDH-deficient GISTs ($n = 3$, bottom). CTCF profiles for representative SDH-intact (top) and SDH-deficient (bottom) tumours are also shown. Genes (blue) and CTCF-binding sites in the KIT insulator (orange) are highlighted. **c**, Traces depict 4C-seq interaction signal between the KIT super-enhancer viewpoint primer and the *KIT* gene in GIST-T1 cells expressing Cas9 and control (black) or KIT insulator targeting sgRNAs (red). **d**, Plot depicts relative *KIT* expression in GIST-T1 cells expressing Cas9 and control (black) or KIT insulator targeting sgRNAs (red). Bar indicates mean of three biologically independent replicates (dots). *P* values by two-sided *t*-test. **e**, **f**, FGF and KIT insulator fraction methylation (β-value) evaluated in an expanded cohort of GIST tumours by locus-specific bisulfite sequencing. **e**, Bar plot depicts average methylation levels across six CpGs within FGF insulator CTCF peak 2 in normal stomach muscle (NSM; $n = 2$), SDH-intact GISTs ($n = 17$) and SDH-deficient GISTs ($n = 11$). **f**, Bar plot depicts average methylation levels across four CpGs within KIT insulator CTCF peak 2 in normal stomach muscle ($n = 2$), SDH-intact GISTs ($n = 20$) and SDH-deficient GISTs ($n = 12$) (*n* values indicate number of biologically independent tumours).

muscle control samples, we detected robust interactions throughout the *ANO1* TAD, but not beyond its boundaries, consistent with robust FGF insulator function. In SDH-deficient tumours, however, the same super-enhancer viewpoint physically interacted with sequences well beyond the boundary, including with the *FGF3* and *FGF4* genes, which are ~200 kb from the viewpoint (Fig. 2b and Extended Data Fig. 3a–c). These data suggest that *FGF3–FGF4* locus topology is profoundly altered in SDH-deficient GISTs, with CTCF insulator loss allowing aberrant contacts between the *ANO1* super-enhancer and FGF ligand genes.

To assess directly whether FGF insulator loss affects FGF gene expression, we used genome editing to disrupt the insulator in GIST-T1 cells, which harbour a GIST-like enhancer landscape and retain CTCF binding and boundary function. We used CRISPR–Cas9 and short guide RNAs (sgRNAs) to edit the motifs underlying the two strongest CTCF sites in the FGF insulator (Extended Data Fig. 3d). This resulted in a sixfold induction of *FGF4*, and a 35-fold induction of *FGF3* (Fig. 2c, d). These data directly link insulator loss to the marked upregulation of FGF ligands in SDH-deficient GISTs.

Notably, a switch between CTCF-bound and DNA-methylated insulator states underlies genomic imprinting[13,14]. FGF insulator loss might therefore also represent a stable epigenetic event or 'epimutation' that effects a single allele. In five of our SDH-deficient samples, we identified heterozygous single nucleotide polymorphisms (SNPs) within an *FGF3* or *FGF4* exon. In four of these cases, analysis of the informative SNP in RNA-seq data revealed that the FGF ligand gene was mono-allelically expressed (Extended Data Fig. 4a–c). By contrast, *ANO1* was bi-allelically expressed, suggesting that the biased FGF expression reflected allele-specific insulator loss. Consistently, in one SDH-deficient tumour with a heterozygous SNP near the CTCF site, we confirmed that only one allele of the FGF insulator was methylated (Extended Data Fig. 4d). In a second tumour with an informative SNP near the *ANO1* super-enhancer

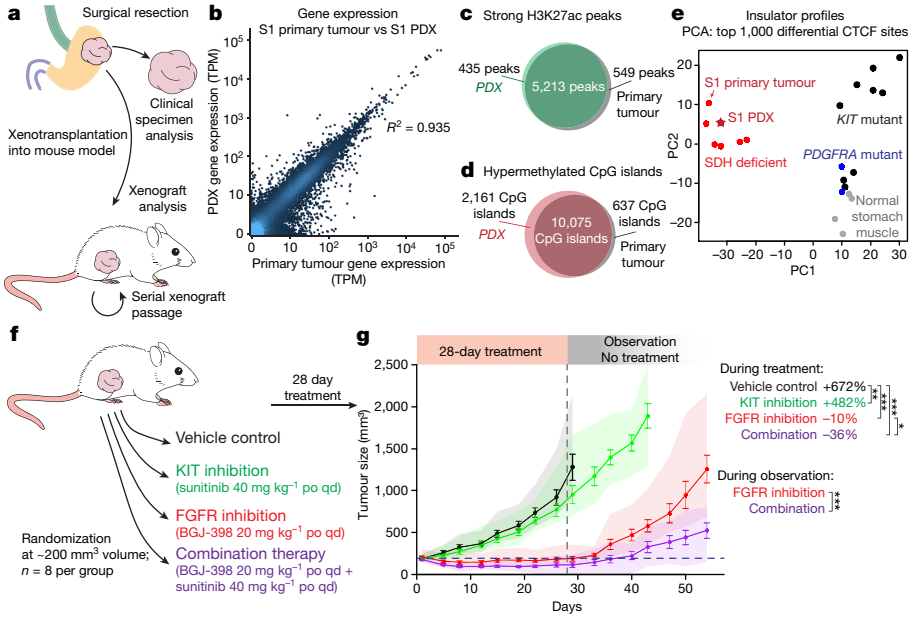


Fig. 4 | SDH-deficient GIST PDX trial confirms dependence on FGF signalling. **a**, Specimen collection and generation of PDX model. **b**, Scatter plot compares expression of genes (points) between primary tumour S1 (*x* axis) and PDX (*y* axis) per RNA-seq. Pearson correlation is indicated. **c**, Venn diagram depicts overlap between strong H3K27ac ChIP–seq peaks in primary tumour (black) and PDX (green). **d**, Venn diagram depicts overlap between hypermethylated CpG islands in primary tumour (black) and PDX (red) per bisulfite sequencing. **e**, Scatter plot depicts principal component analysis (PCA) on the top 1,000 differential CTCF sites for primary tumours (coloured by subtype) and PDX (star). PDX and originating tumour (S1) both cluster with SDH-deficient GISTs (red). **f**, Experimental design of the pre-clinical trial. Following xenograft implantation and growth, mice were randomized to four treatment groups and treated with the indicated regimen daily for 28 days (oral daily). Observation was continued until the clinical endpoint (tumour volume of 2,000 mm^3). **g**, Plot depicts tumour volume during treatment and observation periods. Points represent mean tumour volume, error bars represent s.e.m., and shading represents the range of tumour volumes for $n = 8$ biologically independent xenografts per group. Relative tumour volume immediately following treatment cessation (day 29) is indicated at the right of the panel. *P* values reflect the difference in tumour growth between group per two-way ANOVA (*$P < 0.05$, **$P < 0.001$, ***$P < 0.0001$).

4C-seq viewpoint primer, we confirmed that the aberrant interaction between super-enhancer and *FGF4* was also strongly biased to one allele (Extended Data Fig. 4e, f). These data suggest that insulator loss, topological reorganization and FGF induction reflect a stable epigenetic alteration propagated in the malignant clone.

In addition to the FGF insulator, our screen identified a top-ranked CTCF insulator loss in the *KIT* locus. This hit was of interest given that *KIT* is an established GIST oncogene, and given prior reports of cross-talk between FGF and KIT signalling[22,23]. HiC and HiChIP data reveal that the *KIT* gene is contained within a ~600 kb TAD (Fig. 3a). This large TAD contains within it a smaller insulated domain (~100 kb) that is flanked by CTCF sites. This smaller domain harbours a large super-enhancer that is highly acetylated in all GIST specimens examined (Extended Data Fig. 2b). It is partitioned from KIT by a topological boundary that we refer to as the KIT insulator.

The KIT insulator contains two strong CTCF sites separated by around 7 kb. Both sites gain methylation and lose CTCF binding in SDH-deficient GISTs (Extended Data Fig. 2e, f). To determine whether CTCF loss is associated with altered *KIT* locus topology, we performed 4C-seq using a viewpoint primer in the insulated super-enhancer (Fig. 3b). In SDH-intact tumours, the super-enhancer engages in robust interactions throughout the small insulated domain, but not beyond its boundaries (Extended Data Fig. 5a, b). In SDH-deficient tumours, however, the super-enhancer interacts with sequences well beyond the KIT insulator (Extended Data Fig. 5c, d), consistent with loss of CTCF binding and boundary function. Notably, quantification of interaction signals in SDH-deficient tumours indicated that approximately 15–20% of interactions made by this super-enhancer viewpoint are with the *KIT* promoter and gene, compared with 1–5% in SDH-intact tumours (Extended Data Fig. 5d).

To test directly whether CTCF loss alters *KIT* locus topology, we edited the two CTCF motifs in the KIT insulator in GIST-T1 cells (Extended Data Fig. 5e) and evaluated locus topology by 4C-seq. Although the KIT insulator boundary was clearly evident in control GIST-T1 cells (Fig. 3a), it was compromised in the edited cells, as demonstrated by frequent contacts between super-enhancer and KIT (Fig. 3c). We also considered the impact of insulator loss on *KIT* expression. Although GIST-T1 cells already highly express a constitutively active form of this oncogene, we found that insulator disruption further increased *KIT* expression by around 50% under culture conditions that partially mimic SDH deficiency (Fig. 3d and Extended Data Fig. 6). Although this proportional increase is modest, it corresponds to a substantial increase in transcriptional output given high baseline *KIT* expression in GIST-T1 cells.

Our hypothesis that FGF and KIT insulator losses drive SDH-deficient GIST predicts that these insulators should be recurrently disabled, and the corresponding oncogenes consistently expressed across tumours. We therefore examined insulator methylation across 32 GIST specimens from our original cohort and an additional validation cohort. Both insulators were highly methylated in all SDH-deficient cases, but not in any SDH-intact tumours or normal controls (Fig. 3e, f). Consistently, CTCF binding to these insulators was compromised in all six SDH-deficient GISTs evaluated, but retained in all SDH-intact tumours and normal stomach muscle controls (Extended Data Fig. 2c–f). Furthermore, these CTCF sites were consistently unmethylated across multiple non-malignant cell and tissue types, including a population enriched for interstitial cells of Cajal (ICCs), the presumed GIST cell of origin (Extended Data Fig. 7a). Finally, *FGF4* is consistently expressed across SDH-deficient tumours yet it is only expressed at very low or undetectable levels in *KIT*-mutant GISTs, *PDGFRA*-mutant GISTs and ICCs (Fig. 1h and Extended Data Fig. 7b). The recurrence and specificity of these insulator losses support their functional significance in SDH-deficient GISTs.

Finally, we evaluated directly whether signalling through FGFR and/or KIT is required for tumour growth. Although we are unaware of any in vitro SDH-deficient GIST models, we successfully established an in vivo PDX model from one of our SDH-deficient GIST specimens

(*KIT* and *PDGFRA* wild type) (Fig. 4a). Model and parental tumour have remarkably similar RNA expression, H3K27ac enhancer landscapes, methylation and CTCF binding profiles (Fig. 4b–e). The PDX also maintains characteristic enhancers and CTCF insulator losses in the FGF and *KIT* loci, and strongly expresses *FGF3*, *FGF4* and *KIT* (Extended Data Fig. 8a, b). These data support the fidelity of this SDH-deficient GIST model.

We therefore tested the efficacy of FGFR and KIT inhibitors in this model. We used BGJ-398, a potent and selective inhibitor of FGFR1–4 in clinical development[25], and sunitinib, a drug approved for GIST with potent activity against unmutated KIT[26]. We dosed PDX mice for 28 days with BGJ-398 (20 mg kg^{-1}), sunitinib (40 mg kg^{-1}) or a combination of the two (Fig. 4f). Single agent sunitinib minimally suppressed tumour growth, consistent with the drug-resistant phenotype of the SDH-deficient GIST subtype and prior reports that cross-talk between FGF and KIT signalling confers resistance to KIT inhibition[22,23]. By contrast, single agent BGJ-398 completely suppressed tumour growth throughout the dosing period, strongly supporting a critical role for FGF signalling in tumorigenesis (Fig. 4g and Supplementary Table 3). Sensitivity to FGFR inhibition is specific to this GIST subtype as BGJ-398 lacks efficacy against SDH-intact PDX models[23]. Notably, the combination of FGFR and KIT inhibitors resulted in the most durable response, with growth suppression well beyond the dosing period (Fig. 4g and Extended Data Fig. 8c, d). These pre-clinical data indicate that both RTK signalling pathways drive SDH-deficient GIST, and strongly support the significance of the underlying epigenetic lesions.

In conclusion, we identify multiple epigenetic lesions that converge to activate RTK signalling and proliferation in SDH-deficient GIST. We show that the characteristic hypermethylation in these tumours is associated with pervasive insulator losses, topological reorganization of the FGF and *KIT* loci, and particularly potent induction of the *FGF4* and *FGF3* oncogenes. Although our data do not address the precise cellular contexts in which these lesions arise, it is notable that KIT signalling regulates proliferation of the presumed GIST cell of origin, ICC[27]. Similarly, the *ANO1* gene, the super-enhancer of which aberrantly drives *FGF3* and *FGF4* expression, encodes an ion channel that is highly expressed and essential for ICC[28,29]. Hence, topological changes that deregulate FGF and *KIT* expression could lead to unchecked signalling in these precursors. Although the corresponding loci are genetically wild type in SDH-deficient GIST, the functional significance of their deregulation is supported by the prevalence of gain-of-function KIT mutations in SDH-intact GISTs and by a recent report that *FGF4* is genetically amplified in a rare subset of *KIT/PDGFRA/SDH/RAS* quadruple wild-type GIST[17,30]. Our pre-clinical PDX data substantiate their significance and establish proof of concept for therapeutic intervention. Given that few stable epigenetic events have been established as drivers of tumorigenesis[2], our nomination of two novel functional lesions represents an important advance.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1668-3.

1. Kaelin, W. G. Jr & McKnight, S. L. Influence of metabolism on epigenetics and disease. *Cell* **153**, 56–69 (2013).
2. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–692 (2007).
3. Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017).
4. Janeway, K. A. et al. Defects in succinate dehydrogenase in gastrointestinal stromal tumors lacking KIT and PDGFRA mutations. *Proc. Natl Acad. Sci. USA* **108**, 314–318 (2011).
5. Killian, J. K. et al. Succinate dehydrogenase mutation underlies global epigenomic divergence in gastrointestinal stromal tumor. *Cancer Discov.* **3**, 648–657 (2013).
6. Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013).
7. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
8. Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
9. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
10. Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
11. Xiao, M. et al. Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26**, 1326–1338 (2012).
12. Lu, C. et al. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).
13. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482–485 (2000).
14. Hark, A. T. et al. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**, 486–489 (2000).
15. Liu, X. S. et al. Editing DNA methylation in the mammalian genome. *Cell* **167**, 233–247.e17 (2016).
16. Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).
17. Hirota, S. et al. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* **279**, 577–580 (1998).
18. Boikos, S. A. & Stratakis, C. A. The genetic landscape of gastrointestinal stromal tumor lacking *KIT* and *PDGFRA* mutations. *Endocrine* **47**, 401–408 (2014).
19. Taguchi, T. et al. Conventional and molecular cytogenetic characterization of a new human cell line, GIST-T1, established from gastrointestinal stromal tumor. *Lab. Invest.* **82**, 663–665 (2002).
20. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
21. Arao, T. et al. FGF3/FGF4 amplification and multiple lung metastases in responders to sorafenib in hepatocellular carcinoma. *Hepatology* **57**, 1407–1415 (2013).
22. Javidi-Sharifi, N. et al. Crosstalk between KIT and FGFR3 promotes gastrointestinal stromal tumor cell growth and drug resistance. *Cancer Res.* **75**, 880–891 (2015).
23. Li, F. et al. FGFR-mediated reactivation of MAPK signaling attenuates antitumor effects of imatinib in gastrointestinal stromal tumors. *Cancer Discov.* **5**, 438–451 (2015).
24. West, R. B. et al. The novel marker, DOG1, is expressed ubiquitously in gastrointestinal stromal tumors irrespective of KIT or PDGFRA mutation status. *Am. J. Pathol.* **165**, 107–113 (2004).
25. Pal, S. K. et al. Efficacy of BGJ398, a fibroblast growth factor receptor 1-3 inhibitor, in patients with previously treated advanced urothelial carcinoma with *FGFR3* alterations. *Cancer Discov.* **8**, 812–821 (2018).
26. Janeway, K. A. et al. Sunitinib treatment in pediatric patients with advanced GIST following failure of imatinib. *Pediatr. Blood Cancer* **52**, 767–771 (2009).
27. Sircar, K. et al. Interstitial cells of Cajal as precursors of gastrointestinal stromal tumors. *Am. J. Surg. Pathol.* **23**, 377–389 (1999).
28. Gomez-Pinilla, P. J. et al. Ano1 is a selective marker of interstitial cells of Cajal in the human and mouse gastrointestinal tract. *Am. J. Physiol. Gastrointest. Liver Physiol.* **296**, G1370–G1381 (2009).
29. Singh, R. D. et al. Ano1, a Ca^{2+}-activated Cl^- channel, coordinates contractility in mouse intestine by Ca^{2+} transient coordination between interstitial cells of Cajal. *J. Physiol.* **592**, 4051–4068 (2014).
30. Urbini, M. et al. Gain of FGF4 is a frequent event in KIT/PDGFRA/SDH/RAS-P WT GIST. *Genes Chromosom. Cancer* **58**, 636–642 (2019).

Methods

Primary GIST specimens and cell culture models

Epigenetically characterized clinical samples were obtained as frozen specimens from Brigham and Women's Hospital or from the Massachusetts General Hospital Pathology Tissue Bank. The validation cohort was obtained as FFPE samples from the BWH tissue bank. All samples were acquired with Institutional Review Board approval, and were de-identified before receipt. *PDGFRA* and *KIT* mutational status were confirmed through Sanger sequencing for frozen specimens, while SDH status was determined by immunohistochemistry (details below).

Samples were also examined via RNA-seq and ChIP–seq input controls (details below) in order to look for mutations or copy number changes in all FGF ligand and receptor genes—no copy number alterations were found and no sequence variants were detected other than known annotated SNPs.

The GIST-T1 cell line was obtained from Cosmo Biosciences[19]. Cells were passaged in DMEM with 10% serum, 1× antibiotics and 1× Glutamax (Life Technologies). For pseudohypoxia experiments, cells were treated with 200 μM deferoxamine mesylate (Sigma) or vehicle control (water) for 72 h. For succinate conditions, cells were cultured in 20 μM dimethylsuccinate (Sigma), which was slowly added to a cell culture dish containing standard media and allowed to dissolve before addition of cells.

Chromatin immunoprecipitation

ChIP–seq was performed as described previously. In brief, cultured cells or minced frozen tissue were crosslinked in 1% formaldehyde and snap frozen in liquid nitrogen before storage at −80 °C for at least overnight. Sonication of samples were calibrated such that DNA was sheared to between 300 and 700 bp fragment length. CTCF was precipitated with a monoclonal rabbit CTCF antibody, clone D31H2 (Cell Signaling no. 3418). Histone H3K27 acetylation was immunoprecipitated with antibody from Active Motif (no. 39133). ChIP DNA was used to generate sequencing libraries by end repair (End-It DNA repair kit, Epicentre), 3′ A base overhang addition via Klenow fragment (NEB) and ligation of barcoded sequencing adapters. Barcoded fragments were amplified via PCR. Libraries were sequenced as 38-bp end reads on an Illumina NextSeq500 instrument. Processed genomic data has been deposited into GEO under accession number GSE107447, while raw sequencing data has been deposited into dbGaP (phs001906.v1.p1). See also Supplementary Table 4.

Reads were aligned to the reference genome (hg38) using BWA aln version 0.7.4[31], removing reads with mapping quality score <10. For H3K27 acetylation ChIP–seq and input controls, PCR duplicates were removed by Picard toolkit 2.9.2. Peaks were called with HOMER 4.9[32] against input controls. To call all H3K27ac peaks, we used 'histone' settings. To call super-enhancers[33], we used 'super' settings and no local filtering. CTCF peaks were called with 'factor' settings. To measure H3K27ac correlations, signal at the union of the peaks (5 kb window around the centre) was calculated by featureCounts 1.6.2[34]. We downloaded and reprocessed publicly available data of other gastrointestinal tumours for comparison (GSM1969645[35], GSM1969657[35], GSM2058055[36], GSM2058056[36], GSM2131266[37] and GSM2131280[37]). The dendrogram is based on unweighted average distance linkage of the Pearson correlations between the 10,000 most variable peaks, although analysis results were similar when comparing correlations over all peaks.

Hybrid selection bisulfite sequencing

Hybrid selection probes were designed to capture ~160,000 CTCF-binding sites, and ~5,000 promoters. CTCF bind sites lists were collated from ENCODE (as downloaded from UCSC genome browser, table wgEncodeRegTfbsClusteredV3, Release 4) as well as additional CTCF maps of primary cholangiocarcinoma and glioma[16]. Total genomic DNA was isolated using the DNAeasy Blood & Tissue Kit (Qiagen) and sheared using the Covaris LE220. Ampure XP beads (Agencourt) were used to size select gDNA fragments within 150–320 bp and sheared distribution was verified via BioAnalyzer (Agilient). One microgram of gDNA was end repaired, 3′ A base tailed (KAPA Hyper Prep Kit no. KK8502) and ligated to sequencing adaptor (Roche SeqCap Epi Enrichment System). Ligated products were purified using Ampure XP beads. Following bead clean-up, products were bisulfite-converted using the EZ DNA Methylation-Lightning Kit (Zymo Research) and then PCR amplified using KAPA HiFi U+ HotStart ReadyMix (KAPA no. KK2800). Equal concentrations of each library were then combined in sets of three or four along with SeqCap Epi universal and indexing oligos and bisulfite capture enhancer (SeqCap Epi Accessory Kit). Each pool was lyophilized using TOMY Micro-Vac (MV100), resuspended in hybridization buffer (SeqCap Epi Hybridization and Wash Kit), and then hybridized to SeqCap Epi Probe Pool (Roche) for 72 h at 47 °C in a thermocycler. Following the 72 h incubation, captured bisulfite-converted libraries were recovered (SeqCap Pure Capture Bead Kit) at 47 °C in a thermocycler for 45 min, with intermediate vortexing. Capture beads were washed (SeqCap Hybridization and Wash Kit) in a 47 °C water bath. Captured bisulfite-converted libraries were then amplified via PCR (SeqCap Epi Accessory Kit). Libraries were sequenced with 10% PhiX spike-in as 100-bp end reads on the HiSeq2500 in rapid run mode.

Hybrid-selection bisulfite sequencing (HSBS) data were processed by methylCtools 0.9.4[38], using BWA mem version 0.7.12, and aligned to human reference hg38. Owing to the sizes of the captured fragments, probe capture resulted in an effective coverage of about 600 bp around CTCF sites. PCR duplicates were removed by Picard toolkit 2.9.2. DNA methylation levels were called by methylCtools 0.9.4 in loci covered by at least five reads. Methylation at 36,281 CTCF-binding sites that are bound in GIST tumours and covered by the assay were used for downstream analysis.

HiC and HiChIP

In situ HiC was performed as described[9]. CTCF HiChIP was performed as described[20]. In brief, 3 tubes of ~5 million GIST-T1 cells were crosslinked in 1% formaldehyde (2 replicates for HiChIP, 1 for HiC). For HiChIP, cells were lysed using HiC lysis buffer as described. Chromatin was digested with 375 U MboI restriction enzyme (NEB, R0147). After heat inactivation of restriction enzyme and marking of ends with biotin-14-dATP (Life Technologies, 19524-016), DNA was ligated using T4 buffer (NEB, B0202). Chromatin was sheared by Covaris LE220. Chromatin immunoprecipitation was performed with 30 μl of monoclonal rabbit CTCF antibody, clone D31H2 (Cell Signaling 3418). Protein was bound by Protein G beads and after washing was incubated in DNA elution buffer. Eluant was treated with proteinase, crosslinks were reversed and the sample was Zymo purified (Zymo DNA Clean and Concentrator D4003). Biotinylated DNA was pulled down with M280 Streptavidin beads (Invitrogen 11205D) and the DNA was fragmented with Tn5 and libraries were constructed with Nextera kit (Illumina). For HiC, cells were lysed in HiC lysis buffer and chromatin was digested with 200 U MboI (NEB, R0147) overnight. Ends were marked and DNA was ligated as in HiChIP. DNA was precipitated and sheared by Covaris E220. Biotinylated DNA was pulled down by T1 Streptavidin beads (Life Technologies, 65602). End-repair, A-tailing and adaptor ligation were performed as described[9]. Libraries were prepared using Phusion High-Fidelity DNA Polymerase (NEB, M0530). HiChIP and HiC libraries were sequenced as 75-bp end reads on an Illumina NextSeq500 instrument. Data were processed using HiC-Pro[39] and visualized by the WashU EpiGenome Browser[40] and the R package Sushi[41].

The two HiChIP replicates showed high similarity, and therefore were merged for the rest of the analysis. CTCF–CTCF loops were called from HiChIP data with hichipper 0.7.3[42]. Only loops with FDR <5% and supported by at least 5 reads were considered for downstream analysis.

4C analysis

4C analysis was performed using methods adapted from published protocols[43]. In brief, ~10 million cells from culture or frozen minced tumour specimens were crosslinked in 2% formaldehyde. Fixed samples were lysed in lysis buffer containing protease inhibitor cocktail and mechanically disrupted using a Biomasher tissue grinder (Kimble Chase). Lysis was confirmed using methyl green-pyronin staining. Following lysis, nuclei were digested with NlaIII (NEB) overnight at 37 °C in a thermomixer set to 950 rpm. After heat inactivation of restriction enzyme, diluted nuclei were ligated using T4 DNA ligase (NEB) overnight at 16 °C, followed by RNase and proteinase K treatment. Isolated DNA was then digested overnight in Csp6I (Thermo) at 37 °C, diluted and ligated overnight at 16 °C in order to circularize fragments. Efficacy of each ligation and digestion step was verified via agarose gel electrophoresis. Purified circularized DNA was used as input in PCR reactions to create sequencing libraries. Sixteen reactions per sample were performed, each using 200 ng of circularized 4C DNA (3.2 µg total) in 50 µl using Q5 high-fidelity PCR mastermix (NEB). Primers contained sequencing adapters and barcodes, and annealing sections were as follows: *KIT* enhancer viewpoint primer: 5′-TTTCTATTTGCTCGTTCATG-3′; *KIT* non-viewpoint primer: 5′-GGAAACTTCCAAAGTAGGCT-3′; *ANO1* enhancer viewpoint primer: 5′-ATGTCGCCCTCCTGCATG-3′; *ANO1* non-viewpoint primer: 5′-AGACAAATGAGGCCTGGACG-3′; *ANO1* viewpoint primer for SNP detection: 5′-CTCAAACAGACACTCACATG-3′; *ANO1* non-viewpoint primer for SNP detection: 5′-TCTTTTTGGTTGGATTGTAGGAGT-3′. Standard 4C sequencing libraries were sequenced as 38-bp end reads on an Illumina NextSeq500, although only the first read (the viewpoint primer read) was used for further processing. 4C sequencing libraries for detecting SNPs were read as 75-bp end reads on the same machine, and the second read was used for SNP detection. Data were analysed via 4Cseqpipe[44], and median normalized data with a main trend resolution of 22.5 kb were visualized in R.

RNA-seq

Total RNA was isolated from clinical GIST samples using the RNeasy Plus Kit (Qiagen) and quality was determined via Bioanalyzer (Agilent). Libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina), and equimolar multiplexed libraries were sequenced with single-end 75 bp reads on an Illumina NextSeq 500.

Reads were aligned using STAR 2.5.3[45] to the human reference (hg19). RNA-seq data for SDH intact GISTs were previously published[46]. Gene expression was estimated by featureCounts 1.6.2[34]. TPM values were calculated for these data sets[47]. RNA-seq TPM values for normal stomach was downloaded from GTEx v7[48].

Statistical analysis and reproducibility

CTCF peaks of all GISTs were merged by bedtools merge 2.26[49], including only peaks with a score >50 and in the top 50,000 as reported by HOMER. Peaks were then centred around CTCF motif where found by FIMO (MEME 4.7)[50] at a 100 bp window around the peak centre, based on JASPAR 2014 CTCF motif MA0139.1 bases 4-19[51]. If multiple motifs were detected, we kept the one with the highest score. Reads were counted by HTSeq 0.6.1[52]. CTCF profiles were normalized by copy number estimates and across samples by standard median ratio. Copy number values were estimated from input by CNVnator 0.3[53], with 5 kb bins. CTCF sites bound in all samples were used for median ratio normalization as implemented by DESeq2[54], where a site is considered bound in any given sample if its signal is at least 0.25 of the median of the top 10,000 sites for that sample. Normalization factors were used to scale CTCF signal for visualization (Figs. 2, 4) and differential analysis. Differential CTCF binding was called by DESeq2[54], identifying 2,106 CTCF-binding sites that significantly lose CTCF binding in our cohort (FDR < 5%, fold change > 2). To estimate methylation at lost insulators sites we measured average methylation over a 250 bp window around the peak centre. We identified 4,502 sites that significantly gain methylation (FDR < 5% as determined by *t*-test, methylation increase >25%). Sites that both significantly lose CTCF binding and gain DNA methylation were considered 'perturbed' for downstream analysis. We focused on CTCF–CTCF loops that overlap with a perturbed CTCF site and either (1) the loop contains a promoter that is insulated from a super-enhancer (<500 kb away) by the perturbed CTCF site; or (2) the loop contains a super-enhancer (at least half of the super-enhancer resides in the loop) that is insulated from a promoter (<500 kb away) by the perturbed CTCF site. Here we only considered super-enhancers that scored in at least two SDH-deficient GISTs. This resulted in the identification of 60 putatively lost insulators with loss of CTCF and >50% methylation gain, and 167 putatively deregulated genes with >1 TPM median expression across SDH-deficient GIST, 25 of which with >100 TPM (Fig. 1e and Supplementary Table 2).

To test correlations between methylation and CTCF binding we focused only on peaks detected in at least one GIST sample and with an annotated CTCF motif (see above). To empirically estimate the null distribution of the correlation coefficient, CTCF-binding sites were permuted 100 times (Extended Data Fig. 1c).

To derive GIST MAPK activity score, we used previously identified MAPK biomarkers[55], and published data of Imatinib treatment of a GIST line[56]. We picked biomarkers that were downregulated by the Imatinib treatment ($P < 0.01$ by *t*-test, fold change >2), and expressed in all primary GISTs (>5 TPM). This yielded four biomarkers: DUSP6, ETV5, SPRY2 and SPRY4. Final MAPK score was computed by summing z-scores of the four genes and dividing by 2, as suggested[55].

For Figs. 2a and 3a, representative ChIP–seq traces were selected from ChIP–seq profiles for the 11 *KIT* mutant and 6 SDH-deficient GISTs characterized, all of which displayed similar results (see Supplementary Table 1 and Extended Data Fig. 2).

For all CRISPR insulator deletions, viral introduction of CRISPR–Cas9 + sgRNA vector was repeated three times with separate viral preparations and infections to generate biologically independent replicates.

For epigenomic and transcriptomic characterization of clinical tissue (for example, ChIP–seq, 4C, RNA-seq), multiple clinical specimens were analysed, but technical replicates could not be performed on individual samples due to limited availability of material.

For analysis of clinical tissue, no statistical methods were used to predetermine sample size; rather, all available SDH-deficient tumour specimens with validated SDH loss and enough material available for analysis were tested. For mouse studies, no specific statistical calculations were performed; rather, sample size was determined based on prior experience with similar PDX trials. The investigators were not blinded to allocation during experiments and outcome assessment.

Immunohistochemistry

Immunohistochemistry was performed on 4-µm-thick paraffin-embedded tissue sections using a mouse anti-SDHB monoclonal antibody (clone 21A11AE7; 1:200 dilution; Abcam), a rabbit anti-PDGFRA monoclonal antibody (clone D13C6; 1:300 dilution; Cell Signaling Technology), and a rabbit anti-KIT polyclonal antibody (1:150 dilution; Dako). Pressure cooker antigen retrieval in citrate buffer (pH 6.1; Dako Target Retrieval Solution) was used for PDGFRA and SDHB. Dako Envision+ secondary antibody was used. The sections were developed using 3,3′-diaminobenzidine as substrate and counterstained with Mayer's haematoxylin.

CRISPR–Cas9 insulator disruption

The following CRISPR sgRNAs were cloned into the LentiCRISPR vector[57]: sgGFP 5′-GAGCTGGACGGCGACGTAAA-3′; sgKIT_CTCFpeak1 5′-GTCTCTCTTCTGCCAGCAGG-3′; sgKIT_CTCFpeak2 5′-GACTTCCCTGACACTAGATG-3′; sgFGF_CTCFpeak1 5′-GTCCCACTGCCACCACAAGA-3′; sgFGF_CTCFpeak2 5′-GGGCCAGGCCCGCCGCCAGG-3′; sgSDHB 5′-GTGTCTCTTTCAGGCATCTG-3′. sgRNAs were designed to either the GG PAM in the consensus CTCF motifs for CTCF disruption, or to a PAM near the 5′ splice junction of exon 4 of the *SDHB* gene. GIST-T1 cells

were infected with the relevant lentivirus(es) for 48 h. Cells were then selected in 2 μg ml^{-1} puromycin for 4 days, with puromycin-containing media refreshed every 2 days. Cells were allowed to recover from puromycin for 1 week before analysis. Genomic DNA was then isolated and the region of interest was amplified using primers with sequencing adaptors and the following annealing regions: *KIT* CTCF Peak1 Forward 5′-TTTGGG ATTCGAGTGACCAC-3′; *KIT* CTCF Peak1 Reverse 5′-TTCAGGGCTCAACAG CTTCA-3′; *KIT* CTCF Peak2 Forward 5′-GGAAATAACCTCAACCGGTG-3′; *KIT* CTCF Peak2 Reverse 5′-GACTCGGTCTTGCTCCTCTAA-3′. Libraries were sequenced on an Illumina NextSeq500 as 38 bp end reads, and analysed for editing efficiency. Crosslinked cells were also harvested for ChIP analysis to verify loss of CTCF binding.

Quantitative real-time polymerase chain reaction

Total RNA was isolated from GIST-T1 cells using the RNeasy minikit (Qiagen) and used to synthesize cDNA with the SuperScriptIII system (Invitrogen). cDNA was analysed using the SYBR mastermix (Applied Biosystems) on a 7500 Fast Real Time system (Applied Biosystems). Gene expression primers were as follows: *KIT* forward 5′-GCACAATGGCACGG TTGAAT-3′; *KIT* reverse 5′-GGTGTGGGGATGGATTTGCT-3′; *KITLG* forward 5′-AGCGCTGCCTTTCCTTATGA-3′; *KITLG* reverse 5′-CCGGGGAC ATATTTGAGGGT-3′; *EPAS1* forward 5′-CCACCAGCTTCACTCTCTCC-3′; *EPAS1* reverse 5′-TCAGAAAAAGGCCACTGCTT-3′; *FGF4* set 1 forward 5′-CCAACAACTACAACGCCTACGA-3′; *FGF4* set 1 reverse 5′-CCCTTCTTGG TCTTCCCATTCT-3′; *FGF4* set 2 forward 5′-GCAGCAAGGGCAAGCTCT AT-3′; *FGF4* set 2 reverse 5′-CGGTTCCCCTTCTTGGTCTT-3′; *FGF3* forward 5′-ATGCTTCGGAGCACTACAGC-3′; *FGF3* reverse 5′-CACGTACCACAG TCTCTCGG-3′. All gene expression results were normalized to primers for ribosomal protein, large, P0 (RPLP0) as follows: forward 5′-TCCC ACTTGCTGAAAAGGTCA-3′; reverse 5′-CCGACTCTTCCTTGGCTTCA-3′.

Tyrosine kinome tree visualization

Tyrosine kinase phylogeny data were downloaded from kinase.com[58]. Phylogenetic tree was visualized using the R package ggtree[59]. Expression data of each tyrosine kinase were averaged across the SDH-deficient GISTs, and then plotted on the tree, with the area of the red circles corresponding to the average TPM value in SDH-deficient GISTs.

Interrogation of public normal tissue, GIST and ICC expression data

Data for normal tissue expression was obtained from ENCODE[60]. Mouse interstitial cell of Cajal expression data were previously processed and published[61]. Raw Affymetrix microarray data (CEL files) of human ICC and GIST samples were downloaded from GEO, under accessions GSE56670[62], GSE77839[63], GSE17743[64] and GSE20708[65]. CEL files were imported, normalized, and RMA values exported using the R/Bioconductor package affy[66].

Flow cytometry enrichment and analysis of ICCs

Fresh benign stomach muscle tissue was obtained from the MGH Pathology Tissue Bank and dissected from the gastric epithelium. Tissue was initially manually mechanically dissociated with a sterile scalpel, and then subjected to fine mechanical dissociation through three cycles of 1 min each in a Miltenyi gentleMACS dissociator, resulting in a single-cell suspension. A small portion was removed from the cell suspension to serve as the unlabelled and unpermeabilized control to set size gates and test viability. The remainder of the cell suspension was then incubated in a permeabilizing flow cytometry buffer and stained with ANO1-Alexa488 (Santa Cruz Biotechnology clone C-5), KIT-PE (Biolegend Clone 104D2) or CD45-APC (BD Biosciences clone 2D1) for 30 min at 4 °C. Non-permeabilized control cells were treated with propidium iodide immediately before analysis. Stained cells were analysed and collected on a Sony SH800S cell sorter. Compensation parameters were determined using single-labelled UltraComp eBeads (ThermoFisher). Approximately 1.5 million cells were sorted, of which 2,000 were collected as CD45$^-$KIT$^+$ANO1$^+$ (ICC enriched). Cells were lysed in a small volume of TAE/DTT and treated with Proteinase K. Genomic DNA in the lysed cell mixture was then bisulfite converted using the EZ DNA Methylation Gold Kit (Zymo), subjected to locus PCR, and then sequenced on an Illumina NextSeq500.

PDX model generation and efficacy studies

The PDX model was generated from surgical resection tissue from an SDH-deficient GIST patient who consented to research use of material under an IRB-approved protocol. The surgical sample was implanted subcutaneously in female NSG mice and allowed to grow. Tumour growth was monitored by caliper measurements. Once tumours grew to a size of 1,000 mm^3, tumours were isolated and cut into pieces of approximately 3 × 3 × 3 mm, dipped in Matrigel (Corning Life Science) and transplanted subcutaneously in additional NSG mice. Tumours were passaged for no more than 10 times. Tumour samples from all passages were banked by viably freezing in Bambanker freezing media (Fisher Scientific) and used for further studies. For efficacy studies, tumour fragments were implanted into 8-week-old NSG mice. Tumours were allowed to establish to 192 ± 35.7 mm^3 in size before randomization into various treatment groups with $n = 8$/group as: vehicle control (0.1 M citrate buffer, pH 4.5), 40 mg kg^{-1} sunitinib (LC Laboratories, 0.1 M citrate buffer, pH 4.5), 20 mg kg^{-1} BGJ-398 (LC Laboratories, acetate buffer, pH 4.6 and PEG300 in 1:1 ratio) or the combination of BGJ-398 and sunitinib. Mice were treated orally once daily for 28 days with these agents. In the BGJ-398 treatment group, 4 of 8 mice, and in the combination treatment group, 7 of 8 mice, lost >15% body weight requiring drug holidays (1–3 days of drug holidays in the single agent BGJ-398 group and 1–15 days of drug holidays in the combination group). Mice were re-started on treatment once body weight recovered to at least >85% of initial body weight. Treatment groups were censored when the tumour volume reached the maximum permissible size of 2,000 mm^3 in any single mouse in that group. Statistics were determined by two-way ANOVA.

All relevant ethical regulations regarding research in animal models were followed. All animal experiments and study protocols were approved by the Dana Farber Cancer Institute Institutional Animal Care and Usage Committee (IACUC). The endpoint criteria for mice were if the total tumour burden/size reaches 2 cm in any direction or tumour volume exceeds 2,000 mm^3, and/or if the tumour mass interferes with basic/vital bodily functions or becomes persistently ulcerated, and these criteria were followed for all mice in the study.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Sequencing data that support the findings of this study have been deposited in GEO with the accession code GSE107447.

31. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
32. 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).
33. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
34. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
35. Ooi, W. F. et al. Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat. Commun.* **7**, 12983 (2016).
36. Cohen, A. J. et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat. Commun.* **8**, 14400 (2017).
37. McDonald, O. G. et al. Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nat. Genet.* **49**, 367–376 (2017).
38. Hovestadt, V. et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).

39. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
40. Zhou, X. et al. Exploring long-range genome interactions using the WashU Epigenome Browser. *Nat. Methods* **10**, 375–376 (2013).
41. Phanstiel, D. H. *Sushi: Tools for visualizing genomics data*. R package version 1.16.0. https://www.bioconductor.org/packages/release/bioc/html/Sushi.html (2019).
42. Lareau, C. A. & Aryee, M. J. hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat. Methods* **15**, 155–156 (2018).
43. Splinter, E., de Wit, E., van de Werken, H. J., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58**, 221–230 (2012).
44. van de Werken, H. J. et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972 (2012).
45. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
46. Hemming, M. L. et al. Gastrointestinal stromal tumor enhancers support a transcription factor network predictive of clinical outcome. *Proc. Natl Acad. Sci. USA* **115**, E5746–E5755 (2018).
47. Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285 (2012).
48. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
49. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
50. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
51. Mathelier, A. et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42**, D142–D147 (2014).
52. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
53. Abyzov, A., Urban, A. E., Snyder, M. & Gerstein, M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**, 974–984 (2011).
54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
55. Wagle, M. C. et al. A transcriptional MAPK pathway activity score (MPAS) is a clinically relevant biomarker in multiple cancer types. *NPJ Precis. Oncol.* **7**, 2 (2018).Correct publication information for this reference?
56. Chi, P. et al. ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours. *Nature* **467**, 849–853 (2010).
57. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
58. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
59. Yu, G. et al. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
60. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
61. Lee, M. Y. et al. Transcriptome of interstitial cells of Cajal reveals unique and selective gene signatures. *PLoS One* **12**, e0176031 (2017).
62. Killian, J. K. et al. Recurrent epimutation of SDHC in gastrointestinal stromal tumors. *Sci. Transl. Med.* **6**, 268ra177 (2014).
63. Tang, C. M. et al. Hedgehog pathway dysregulation contributes to the pathogenesis of human gastrointestinal stromal tumors via GLI-mediated activation of KIT expression. *Oncotarget* **7**, 78226–78241 (2016).
64. Ostrowski, J. et al. Functional features of gene expression profiles differentiating gastrointestinal stromal tumours according to KIT mutations and expression. *BMC Cancer* **9**, 413 (2009).
65. Astolfi, A. et al. A molecular portrait of gastrointestinal stromal tumors: an integrative analysis of gene expression profiling and high-resolution genomic copy number. *Lab. Invest.* **90**, 1285–1294 (2010).
66. Gautier, L., Cope, L., Bolstad, B. M. & Irizarry, R. A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
67. Li, Z. et al. Hypoxia-inducible factors regulate tumorigenic capacity of glioma stem cells. *Cancer Cell* **15**, 501–513 (2009).
68. Murakami, A. et al. Hypoxia increases gefitinib-resistant lung cancer stem cells through the activation of insulin-like growth factor 1 receptor. *PLoS One* **9**, e86459 (2014).
69. Bosch-Marce, M. et al. Effects of aging and hypoxia-inducible factor-1 activity on angiogenic cell mobilization and recovery of perfusion after limb ischemia. *Circ. Res.* **101**, 1310–1318 (2007).

Acknowledgements We thank S. Gillespie, M. Miri, F. Najm, P. van Galen and E. Choy for assistance with clinical samples and analysis, H. Gu and A. Gnirke for sequencing assistance, and E. Gaskell for discussions. W.A.F. is supported by an F32 from the National Cancer Institute. Y.D. is supported by the Tosteson Postdoctoral Fellowship. B.E.B. is the Bernard and Mildred Kayden Endowed MGH Research Institute Chair and an American Cancer Society Research Professor. This research was supported by the National Cancer Institute, the NIH Common Fund, the Starr Cancer Consortium, and the Ludwig Center at Harvard.

Author contributions Conception and experimental design: W.A.F., Y.D., S.E.J., M.L.H., E.T.S., G.D.D. and B.E.B. Methodology and data acquisition: W.A.F., Y.D., S.E.J., M.L.H., D.R.T., E.H., S.J.S., N.M.J., C.P.R., B.K.E., P.C.G., J.L.H., E.T.S., G.D.D. and B.E.B. Analysis and interpretation of data: W.A.F., Y.D., M.L.H., S.J.S., N.M.J., G.D.D. and B.E.B. Manuscript writing and revision: W.A.F., Y.D. and B.E.B.

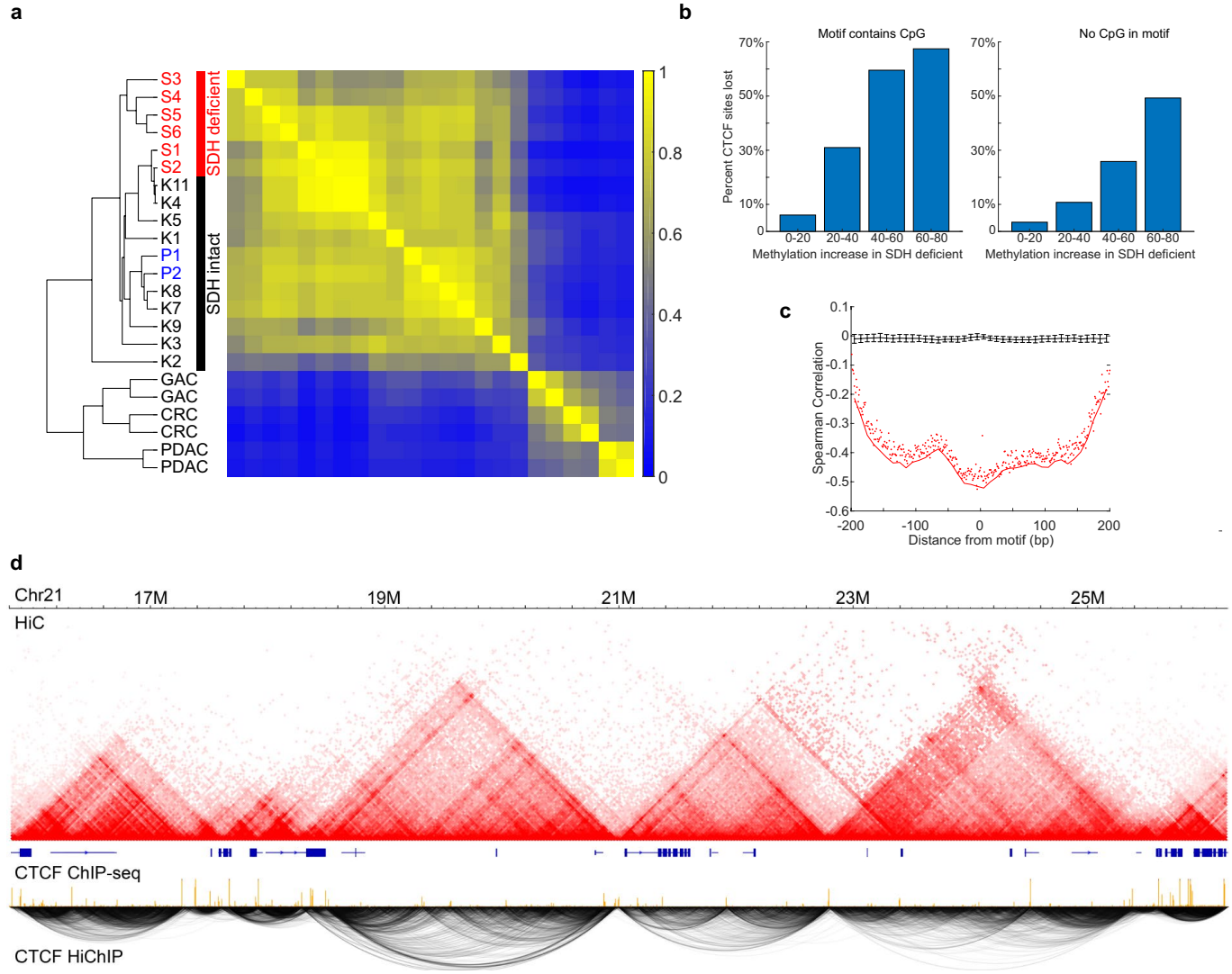
Competing interests B.E.B. is an advisor and equity holder for Fulcrum Therapeutics, 1CellBio, HiFiBio and Arsenal Biosciences, is an advisor for Cell Signaling Technologies, and has equity in Nohla Therapeutics. G.D.D. reports relationships with Novartis, Bayer, Pfizer, EMD-Serono, Sanofi, Ignyta, Roche, Loxo Oncology, AbbVie, Mirati Therapeutics, Epizyme, Daiichi-Sankyo, WIRB Copernicus Group, ZioPharm, Polaris Pharmaceuticals, M.J. Hennessey / OncLive, Adaptimmune, GlaxoSmithKline, Blueprint Medicines, Merrimack Pharmaceuticals, G1 Therapeutics, CARIS Life Sciences, Bessor Pharmaceuticals, ERASCA Pharmaceuticals, CHAMPIONS Oncology, Janssen, PharmaMar; in addition, G.D.D. has a Use patent on imatinib for GIST, licensed to Novartis, with royalties paid to the Dana-Farber Cancer Institute.

Additional information

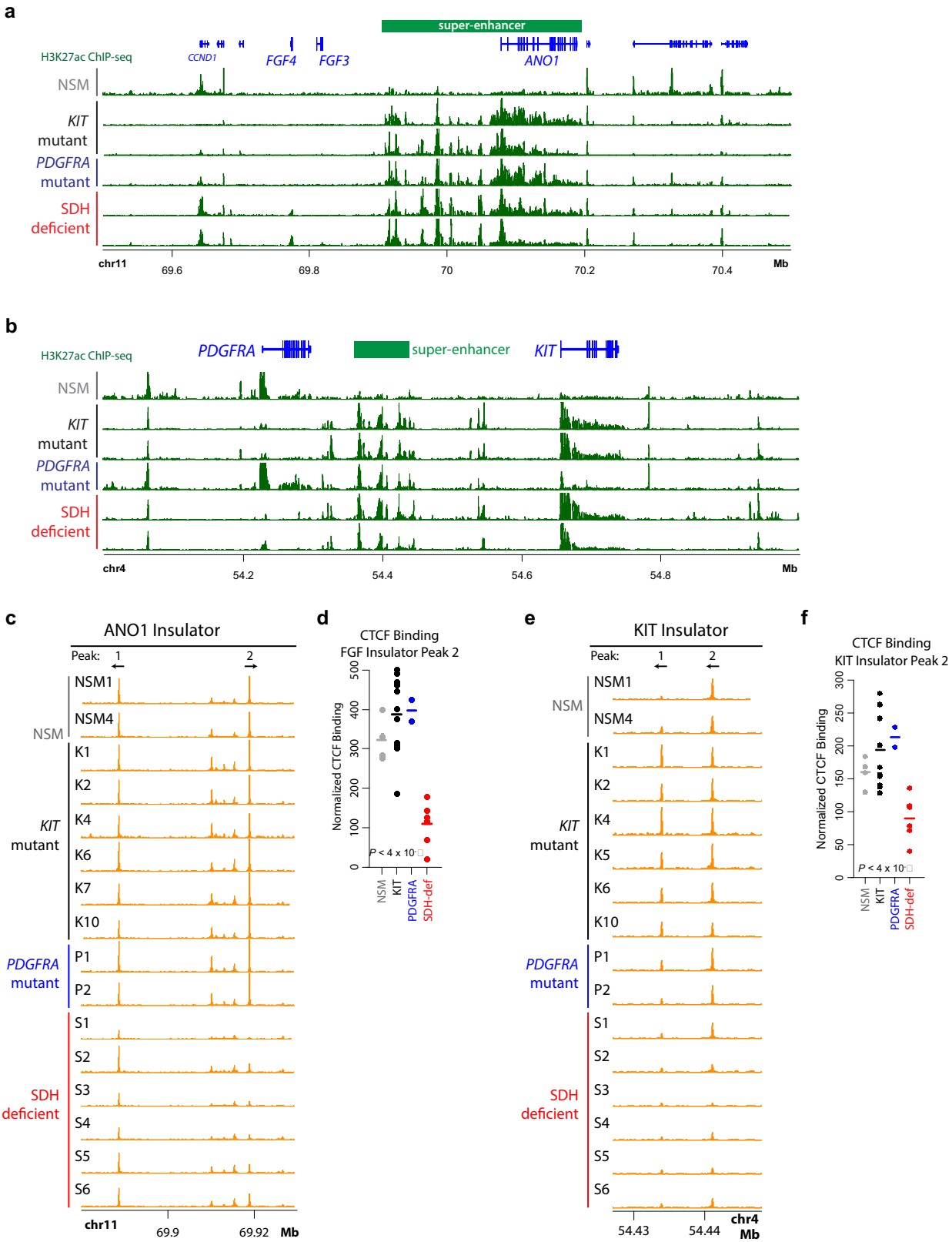
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1668-3.

Correspondence and requests for materials should be addressed to Y.D., G.D.D. or B.E.B.

Peer review information *Nature* thanks Christian Frezza, Michael Heinrich, Michael Rehli and Peter Scacheri for their contribution to the peer review of this work.

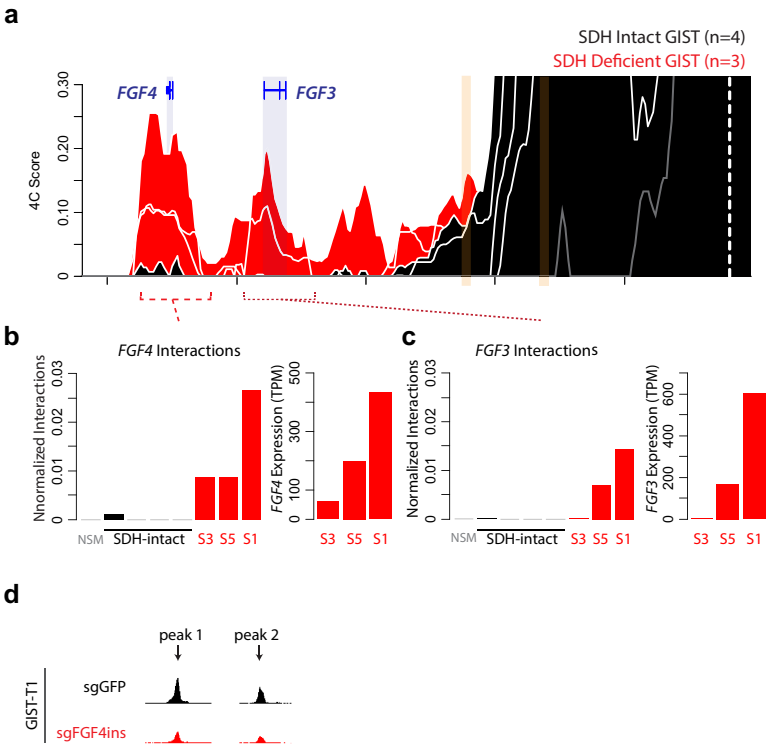


Extended Data Fig. 1 | Epigenomic characterization of GIST. a, ChIP–seq profiles for H3K27ac were compared for GIST specimens and other gastrointestinal tract tumour specimens (GAC, gastric adenocarcinoma; CRC, colorectal cancer; PDAC, pancreatic ductal adenocarcinoma). Heat map depicts pairwise Pearson correlations between the top 10,000 most variable peaks (yellow indicates high correlation; blue indicates low correlation). The dendrogram (left) was derived by unweighted average distance linkage. Enhancer patterns are relatively consistent across GIST subtypes, compared to other tumour types. **b**, DNA methylation levels in the vicinity of CTCF sites were profiled genome-wide by hybrid-selection bisulfite sequencing. CTCF sites are binned according to the amount their methylation increased in SDH-deficient GISTs, relative to SDH-intact GISTs (methylation change computed over a 250 bp window centred on the motif). For each bin, bar graphs depict the percentage of sites that lose CTCF binding in SDH-deficient GISTs, per ChIP–seq. Separate plots are shown for CTCF sites for which motifs do or do not contain a CpG. Increased methylation over CTCF sites is associated with more frequent loss of CTCF binding, even when the CTCF motif lacks a CpG. **c**, Plot depicts correlation between CTCF occupancy and DNA methylation in SDH-deficient GISTs. Red points show Spearman correlations between CTCF ChIP–seq signal and methylation of CpGs at indicated positions relative to the centre of the CTCF motif. Red line reflects correlation to average methylation over 10 bp windows. Randomly permuted data (black) are shown for comparison. Anti-correlation between CTCF occupancy and methylation is evident over a ~250-bp binding footprint. **d**, Genomic views of a representative 10 Mb region on chromosome 21 depict chromosome topology (HiC, red), CTCF binding (ChIP–seq, orange) and CTCF–CTCF loop interactions (HiChIP, black) for the SDH-intact GIST model, GIST-T1. TADs are visible as triangles of enhanced interaction in HiC data, flanked by boundaries that correspond to loop interactions in HiChIP data. Genes (blue) are also indicated.

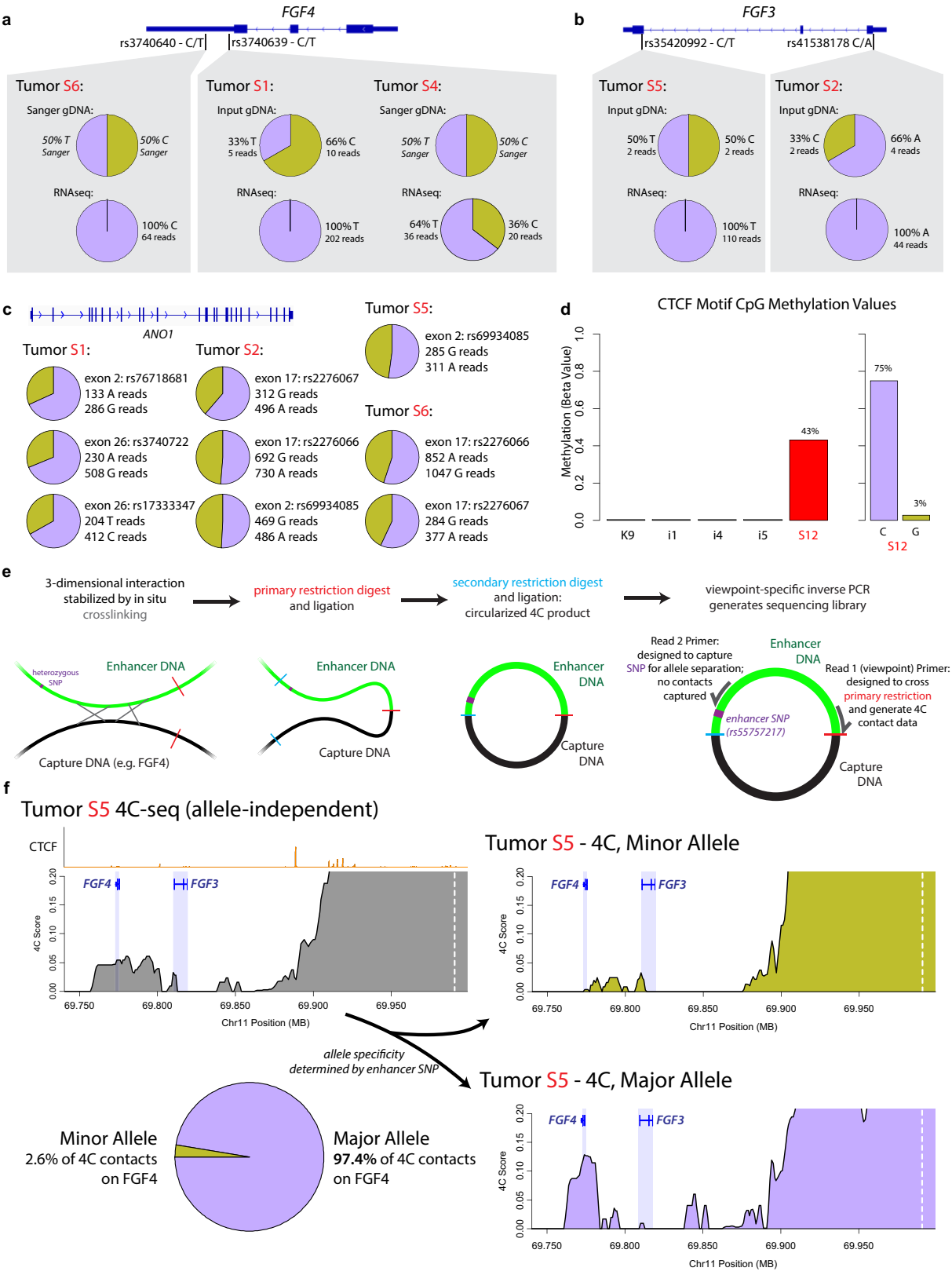


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Super-enhancers and insulators in GIST. a, Traces depict H3K27ac ChIP–seq signal for normal stomach muscle (NSM) and GISTs of the indicated subtype over the *FGF–ANO1* locus. **b**, Traces depict H3K27ac ChIP–seq signal for NSM and GISTs of the indicated subtype over the *PDGFRA–KIT* locus. Genes are indicated in blue, and super-enhancer locations are indicated by green bars. For **a**, **b**, traces are representative of 11 *KIT*-mutant and 6 SDH-deficient tumours with similar results. **c**, Traces depict CTCF binding over the FGF insulator in normal stomach muscle (NSM) and GIST clinical specimens. **d**, Plot depicts CTCF ChIP–seq signal over the strongest CTCF peak in the FGF insulator in normal stomach muscle (NSM, $n = 4$), and *KIT* mutant ($n = 11$), *PDGFRA* mutant ($n = 2$) and SDH-deficient GISTs ($n = 6$). **e**, Traces depict CTCF binding over the KIT insulator in normal stomach muscle (NSM) and GIST clinical specimens. **f**, Plot depicts CTCF ChIP–seq signal over the strongest CTCF peak in the KIT insulator in normal stomach muscle (NSM, $n = 4$), and *KIT* mutant ($n = 11$), *PDGFRA* mutant ($n = 2$) and SDH-deficient GISTs ($n = 6$). For **d** and **f**, horizontal bars reflect mean values and *P* values indicate significance of CTCF loss in SDH-deficient GIST, as determined by the Walt test (via DEseq2[54]). All *n* values represent the number of biologically independent clinical specimens.

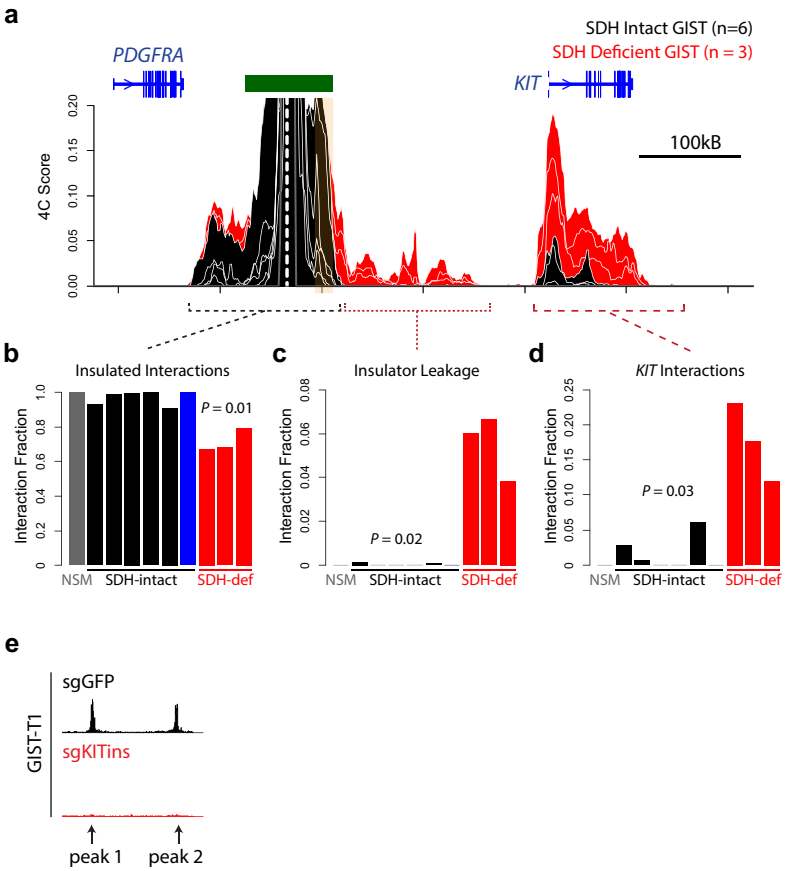


Extended Data Fig. 3 | FGF locus 4C-seq data and insulator deletion. a, Traces depict 4C-seq data at FGF locus, as in Fig. 2b, except graphed on the same axis to allow for direct comparison. **b**, **c**, Bar plots quantify 4C-seq interactions between the super-enhancer viewpoint and *FGF4* (**b**) or *FGF3* (**c**). Expression of these genes in the corresponding SDH-deficient GIST specimens is also shown. **d**, Traces depict CTCF ChIP–seq signal in GIST-T1 cells infected with CRISPR–Cas9 and either a control sgRNA directed at GFP (black, top) or sgRNAs directed against the two indicated CTCF motifs in the FGF insulator (second row, red).

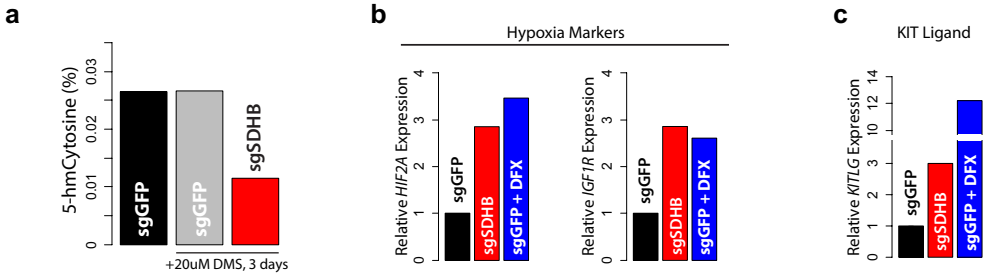


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Allelic imbalance in *FGF3* and *FGF4* activation. **a**, Two heterozygous SNPs in *FGF4* (both 3′ UTR) enabled us to evaluate allelic expression in three SDH-deficient GISTs (tumours S6, S1 and S4). Both alleles for each SNP were detected in DNA sequencing data for these tumours, but only one allele was detected in RNA-seq data of tumours S1 and S6, indicative of mono-allelic *FGF4* expression. Both alleles are detected in tumour S4, indicating bi-allelic expression of *FGF4*. **b**, Heterozygous SNPs in *FGF3* exons (both synonymous base substitutions) enabled us to evaluate allelic expression in the SDH-deficient GISTs (tumours S2 and S5). In both cases, DNA sequencing confirmed heterozygosity at the genome level (C/A and T/C, respectively), but RNA-seq data demonstrated mono-allelic FGF3 expression. **c**, Both alleles of heterozygous SNPs in *ANO1* exons were found in the RNA-seq data derived from SDH-deficient GIST samples, confirming bi-allelic expression of *ANO1*. Similarly, both alleles of heterozygous SNPs were found in the histone H3K27ac ChIP–seq data, confirming the bi-allelic nature of the super-enhancer (not shown). **d**, One SDH-deficient GIST sample was heterozygous for a SNP (rs386829467) located about 50 bp from the CTCF motif of Peak 2 in the FGF insulator. Allele-agnostic methylation data confirmed 43% methylation of the CTCF peak in this tumour, while essentially no methylation was detected in the SDH-intact tumours (left). Separation of the two alleles using the heterozygous SNP revealed strong allelic bias in the SDH-deficient tumour: one allele was largely unmethylated (~3% methylation), while the other was highly methylated (~75% methylation), consistent with mono-allelic methylation of the CTCF site (right). **e**, Schematic depicts 4C-seq experimental protocol and primer design for detecting SNPs. DNA elements in close physical proximity are crosslinked and restricted with an enzyme that leaves nucleotide overhangs. These overhangs are then proximity ligated to crosslinked fragments. A second restriction enzyme (with different restriction sites) is then used to circularize the ligated fragments, allowing for inverse PCR. Here we selected restriction enzymes and designed a custom read 2 primer to capture a heterozygous SNP within the super-enhancer. This second read is normally non-informative as contact frequencies are determined through the viewpoint primer (read 1), but in this case enabled us to detect the SNP and assign each ligated fragment to a specific allele. **f**, The left trace (grey) depicts standard 4C-seq data (allele agnostic), which demonstrates strong interaction between super-enhancer viewpoint and *FGF4*. However, the SNP covered in the non-viewpoint read enabled us to distinguish interactions involving the minor (top right) or major (bottom right) allele. This revealed that the major allele (purple) is responsible for ~97% of super-enhancer–FGF4 interactions.

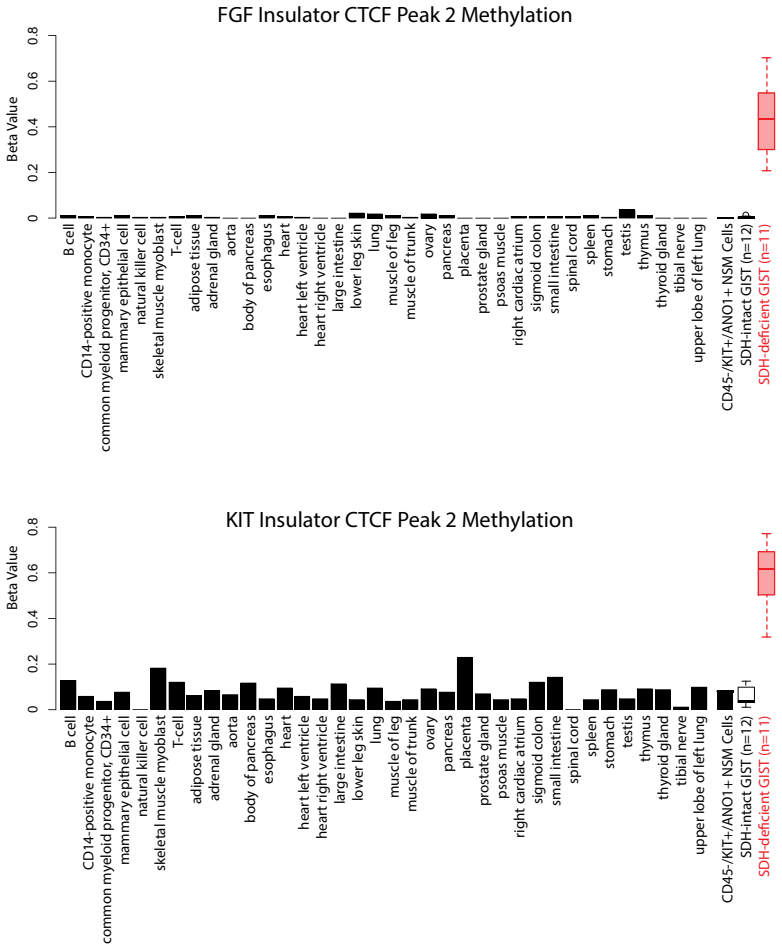


Extended Data Fig. 5 | *KIT* locus 4C-seq data and insulator deletion. a, Traces depict *KIT* locus 4C-seq data, as in Fig. 3b, except graphed on the same axis to allow for direct comparison. **b**–**d**, Bar plots quantify 4C-seq interactions (top, reproduced from Fig. 3b) between the super-enhancer viewpoint and positions within the super-enhancer TAD (**b**), sequences just beyond the KIT insulator (**c**), or the *KIT* gene itself (**d**). *P* values indicate significance of difference between SDH-intact and SDH-deficient, by two-sided *t*-test. **e**, Traces depict CTCF ChIP–seq signal in GIST-T1 cells infected with Cas9 and either a control sgRNA directed at GFP (black, top) or sgRNAs directed against the two bound CTCF motifs in the KIT insulator (second row, red).



Extended Data Fig. 6 | Hypoxia marker induction in GIST-T1 cells. a, Bar plot depicts levels of the TET product, 5-hydroxymethyl-cytosine (5-hmC), measured by ELISA in GIST-T1 cells infected with CRISPR–Cas9 and either a short guide RNA targeting GFP (sgGFP) or SDHB. Cells were cultured in either control media or media supplemented with 20 μM of dimethylsuccinate (DMS), a membrane-permeable ester of succinate, for 3 days, as indicated. **b**, **c**, Plots show relative expression of pseudo-hypoxia-associated genes *EPAS1* (also known as *HIF2A*)[67] and *IGF1R*[68] (**b**), and *KITLG* (also known as *SCF*)[69] (**c**) in control GIST-T1 cells (black), SDH-deficient GIST-T1 cells generated by CRISPR–Cas9 knockout of SDHB and cultured with exogenous succinate (red), or GIST-T1 cells treated with the iron chelator DFX to simulate hypoxia (blue). Upregulation of KIT ligand due to pseudo-hypoxia or tumour hypoxia may supplement FGF ligands in promoting RTK signalling in SDH-deficient GIST.

a

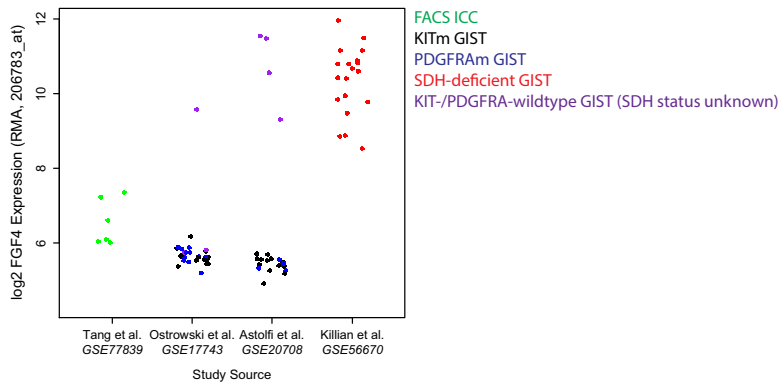


b

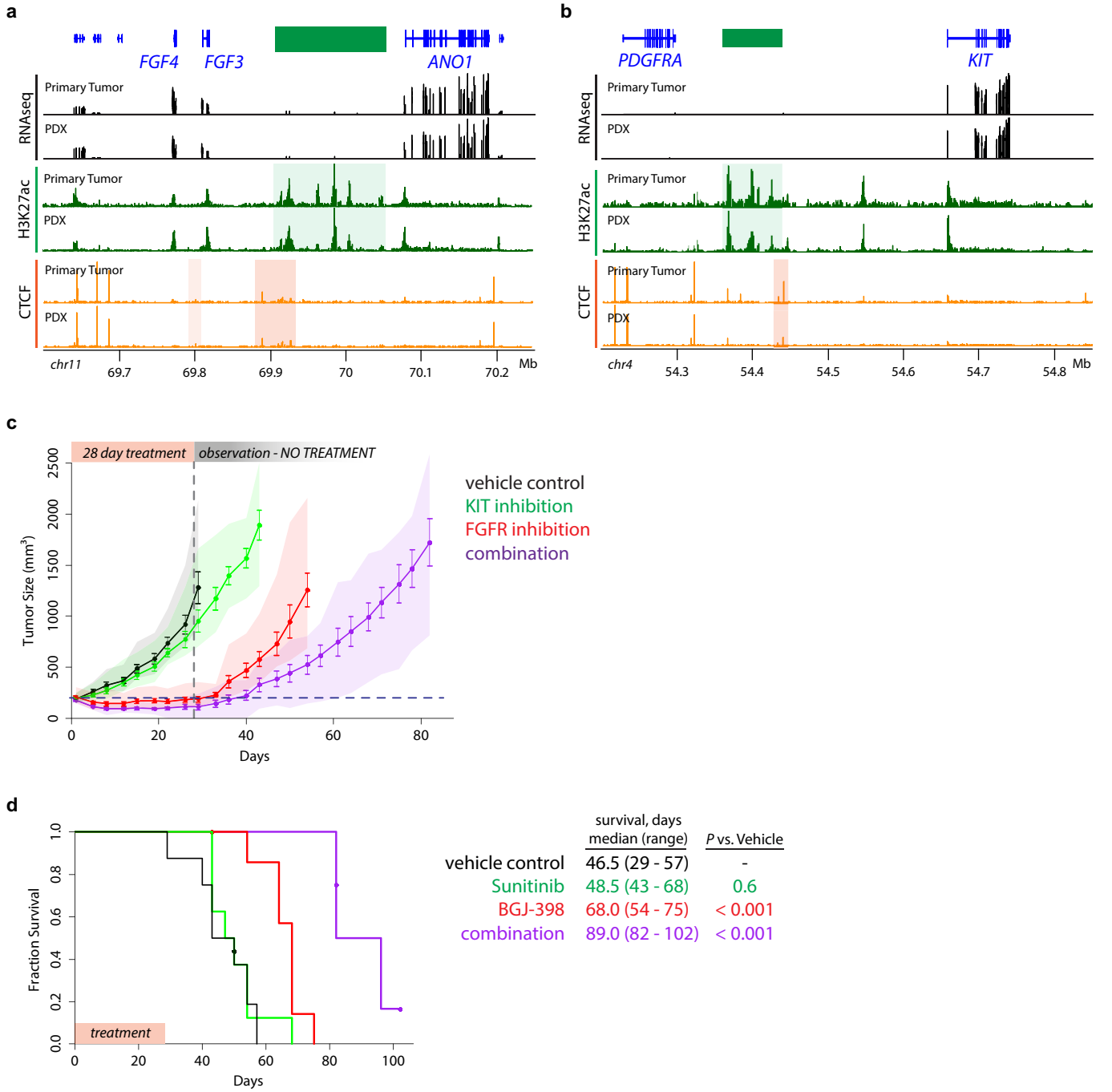
Mouse ICC Transcriptomic Data (FPKM)

Gene	Jejunal ICC	Colonic ICC
Fgf3	**0.07**	**0.07**
Fgf4	**0.07**	**0**
Ano1	702.72	1068.03
Kit	468.03	792
Kitl	13.48	57.33
Fgfr1	61.48	214

Human ICC/GIST U133 Microarray Data



Extended Data Fig. 7 | FGF and KIT insulator methylation and expression in GIST subtypes and non-malignant cells. a, Bar plot depicts methylation of FGF insulator CTCF peak 2 (top) and KIT insulator CTCF peak 2 (bottom) in 34 tissues and primary cells available through ENCODE[60]. Values are average methylation of CpGs nearest the CTCF motifs, determined by whole genome bisulfite sequencing (WGBS) (KIT insulator, 3 CpGs; FGF insulator, 6 CpGs). Methylation of these sites is also shown for SDH-intact and SDH-deficient GISTs (see Fig. 3e, f), and for flow-sorted $CD45^{-}ANO1^{+}KIT^{+}$ (ICC enriched) cells from normal stomach muscle (NSM) tissue (*n* values represent biologically independent specimens). **b**, Left, table depicts FPKM (fragments per kilobase of transcript per million mapped reads) values of relevant genes in mouse ICCs isolated from jejunum or colon[61]. Right, dot plot depicts expression of FGF4 in either flow sorted ICCs (green) or GISTs of the indicated subtype: *KIT* mutant in black, *PDGFRA* mutant in blue, SDH-deficient in red, and KIT-/PDGFRA wild type in purple[62–66]. SDH status of the latter group is unknown, but SDH-deficient GIST represent a significant portion of KIT-/PDGFRA wild-type tumours. Data are drawn from the indicated GEO series and publications.



Extended Data Fig. 8 | PDX trial of FGFR and KIT combination therapy in SDH-deficient GIST. a, **b**, Genomic views depict RNA expression (black), H3K27ac (green) and CTCF occupancy (orange) over the FGF (**a**) and KIT (**b**) loci for the S1 primary tumour and PDX. Genes (blue), super-enhancers (green bar and shade) and lost CTCF insulators (orange shade) are indicated. **c**, Plot depicts tumour volume during treatment and observation periods of experiment, as in Fig. 4g, except with time axis extended until final group reached censor point (one tumour in the group >2,000 mm^3). Points represent mean tumour size, error bars represent s.e.m., and shading represents range of tumour sizes for $n = 8$ biologically independent xenograft-bearing mice per group. For statistics, see Fig. 4g. **d**, Kaplan–Meier plot depicts survival until clinical endpoint (tumour size >2,000 mm^3) for the same PDX trial. Median and range survival are indicated for each group. *P* values reflect difference in survival between groups as calculated by logrank test.

nature research

Corresponding author(s): Bradley Bernstein

Last updated by author(s): Jul 27, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist .

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used specifically for data collection; see below for data analysis software.
Data analysis	Software usage and parameters are detailed in methods section of manuscript. Briefly, sequencing reads were aligned with BWA v. 0.7.4, Methylation data analyzed with methylCtools v. 0.9.4. ChIP-seq peaks were called and analyzed via Homer v. 4.9, MEME v. 4.7, HTSeq v. 0.6.152 and DESeq2 v. 1.16.1, general data analysis and graphing was performed in Matlab v. 9.1.0.441655, R v. 3.5.3, and IGV v. 2.5.3.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data that support the findings of this study have been deposited in GEO with the accession code GSE107447.

Raw sequencing data generated through this project may contain identifiable human genetic information, as such it requires IRB approval to access (data has been deposited into a dbGaP dataset connected to the GEO database).

Raw data for the mouse xenograft trial (i.e. measured tumor volumes) are available as supplementary information table 3.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For analysis of clinical tissue, no statistical methods were used to predetermine sample size; rather, all available SDH-deficient tumor specimens with validated SDH loss and enough material available for analysis were tested. For mouse studies, no specific statistical calculations were performed; rather, sample size was determined based on prior experience with similar PDX trials.
Data exclusions	No data were excluded from analyses.
Replication	For all CRISPR insulator deletions, viral introduction of CRISPR/Cas9+sgRNA vector was repeated three times with separate viral preparations and infections to generate biologically independent replicates. For epigenomic and transcriptomic characterization of clinical tissue (e.g. ChIP-seq, 4C, RNA-seq), multiple clinical specimens were analyzed, but technical replicates could not be performed on individual samples due to limited availability of material. All biological replicates (i.e. tumors of a given driver) were similar at reported sites (i.e. insulator loss/enhancer presence was consistent within driver subgroups).
Randomization	For PDX trial, xenograft-bearing mice were randomized to treatment group at ~200mm3 tumor volume, with 8 mice per treatment group.
Blinding	For PDX trial, blinding was not possible due to preparation and delivery methods of tested drugs. This is thought to have minimal impact on the studies, as no subjective (e.g. behavioral) criteria were measured.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	CTCF antibody is from Cell Signaling technologies, clone D31H2, catalog number 3418. Antibody was validated by manufacturer for ChIP in human cells, has been previously utilized by the authors (Flavahan et al., Nature 2016), and was validated as part of the ENCODE project via Western Blot for CTCF, as well as motif analysis to confirm enrichment for the known CTCF motif. H3K27ac is a rabbit polyclonal antibody available from Active Motif, catalog number 39133, lot 31814008. Antibody was validated by the manufacturer for ChIP in human cells, and was validated as part of the ENCODE project.
Validation	Both antibodies were validated by manufacturers (including statements on the websites), and by the investigators and colleagues as part of the ENCODE project - including, but not limited to, western blot, immunoprecipitation, and ChIP motif finding and analysis of control cell lines and known peaks/motifs.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	One cell line, GIST-T1, was used, and was obtained from the commercial vendor Cosmo Biosciences.
Authentication	The cell line was not authenticated via STR testing, however locus sequencing for the KIT gene confirmed the presence of the known and published KIT mutation present in GIST-T1 cells and the parental tumor the cell line was derived from.
Mycoplasma contamination	The cell line was tested for mycoplasma via PCR-based method and confirmed to be mycoplasma-free.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were utilized.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	For the PDX trial, female 8-week old NSG mice were utilized.
Wild animals	No wild animals were utilized in this study.
Field-collected samples	No field-collected samples were utilized in this study.
Ethics oversight	All animal experiments and study protocols were approved by the Dana Farber Cancer Institute Institutional Animal Care and Usage Committee (IACUC), and this is noted in the manuscript.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	The study involved the collection of deidentified and anonymized tumor material from patients at either Brigham and Women's Hospital, The Dana Farber Cancer Institute, or Massachusetts General Hospital. As such, no information about the patients, other than disease pathology, is known.
Recruitment	Patient tissue was obtained from tissue banks at either MGH or DFCI. Standard of care for GIST includes surgical resection of the tumor bulk - following treatment, excess surgical material from consenting patients was deposited into these banks. As the collection of these tissues occurs as part of the normal disease treatment, it is unlikely there is significant self-selection bias.
Ethics oversight	The study protocol was approved by the Massachusetts General Hospital IRB and the Dana Farber Cancer Institute IRB.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	GEO database accession GSE107447 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107447
Files in database submission	See included file, ChIP-seq_2017-12-16052D_file_info.xlsx
Genome browser session (e.g. UCSC)	no longer applicable

Methodology

Replicates	For epigenomic and transcriptomic characterization of clinical tissue (e.g. ChIP-seq, 4C, RNA-seq), multiple clinical specimens were analyzed, but technical replicates could not be performed on individual samples due to limited availability of material. All clinical samples within a driver subgroup (SDH-intact vs. SDH-deficient) were highly similar at tested locations (e.g. KIT/FGF insulator loss and superenhancer presence). See extended data figure 2 for more info.

Sequencing depth	See included file, ChIP-seq_2017-12-16052D_file_info.xlsx, which includes sequencing depth for each experiment.
Antibodies	CTCF antibody is from Cell Signaling technologies, clone D31H2, catalog number 3418. Antibody was validated by manufacturer for ChIP in human cells, has been previously utilized by the authors (Flavahan et al., Nature 2016), and was validated as part of the ENCODE project via Western Blot for CTCF, as well as motif analysis to confirm enrichment for the known CTCF motif. H3K27ac is a rabbit polyclonal antibody available from Active Motif, catalog number 39133, lot 31814008. Antibody was validated by the manufacturer for ChIP in human cells, and was validated as part of the ENCODE project.
Peak calling parameters	Peaks were called with HOMER 4.9 against input controls. To call all H3K27ac peaks, we used 'histone' settings. To call super-enhancers, we used 'super' settings and no local filtering. CTCF peaks were called with 'factor' settings.
Data quality	All reported peaks are detected with FDR < 0.1% and fold change > 4
Software	Homer 4.9, DESeq2 1.16.1, FIMO/MEME 4.7, HTSeq 0.6.1, featureCounts 1.6.2, and CNVnator 0.3