

## Supplementary Methods

### Human CRC cancer cell lines

SW1116, LOVO, HT29 and DLD1 cells were purchased from the American Type Culture Collection (ATCC, Manassas, VA) and grown in DMEM supplemented with 10% fetal bovine serum (FBS) and antibiotics. All cell lines were cultured at 37°C in an incubator with 95% air and 5% CO<sub>2</sub> supply.

### RNA extraction, semi-quantitative RT-PCR, and real-time PCR analyses

Total RNA was extracted from cells and tissues using TRIzol<sup>TM</sup>. cDNA was synthesized from 1µg of total RNA using Transcriptor Reverse Transcriptase (Roche). Real-time PCR was performed using SYBR Green master mix (Roche) on LightCycler 480. Each sample was performed in triplicate.  $\Delta\Delta$ CT method was applied to determine the fold change in gene expression. Primer sequences are listed in **Table S4**.

### Plasmids and gene transfection

pCMV-SQLE (RC202008), pCMV control plasmids, SQLE shRNA (TL309122V), and control shRNA (pGFP-C-shLenti) plasmids were ordered from OriGene. pCMV control and pCMV-SQLE plasmids were transfected into CRC cells using Lipofectamine 2000 (Invitrogen). For stable cell lines selection, cells were passed into fresh growth medium with 800mg/ml neomycin (G418). After one month of selection, cells were harvested to detect SQLE expression. For shRNA, lentivirus was generated using HEK293T cells by transfection of shRNA plasmid with packaging plasmids. After 72h post-transfection,

the culture supernatants were used for transduction of CRC cells. For stable cell lines selection, cells were passed into fresh growth medium with 2 µg/ml puromycin. After 2 weeks of selection, cells were harvested to detect SQLE expression.

### **Western blot**

20-40 mg protein was loaded onto 12% SDS-PAGE and transferred to an equilibrated polyvinylidenedifluoride (PVDF) membrane (Invitrogen). After blocking in 5% BSA dissolved in TBST solution at room temperature for 1h, the membrane was incubated with primary antibody in 5% BSA overnight at 4°C. After washing steps, the membrane was incubated with secondary antibody for 1h at room temperature. After 4 washing steps, protein was detected by enhanced chemiluminescence reagent (ECL, Amersham Corporation, Heights, IL, USA). Antibodies used in this study are shown in **Table S5**.

### **Colony formation assay**

For the cell colony formation assay, stably transfected cells (1000/well) were plated in 6-well plates. After culturing for 7 days, cells were fixed with 70% ethanol and stained with 0.25% crystal violet solution. Colonies with over 50 cells per colony were counted. All experiments were conducted three times in triplicate.

### **Transmission Electron Microscopy**

Small pieces of colon tissues (1mm<sup>3</sup>) from germ-free mice (fed with stools from Sqle tg mice or wildtype mice) were collected and fixed in 2.0% glutaraldehyde in 0.1 mol/L

sodium cacodylate (Electron Microscopy Sciences, Hatfield, PA). The ultrathin sections were prepared on a Reichert Ultracut E ultramicrotome, and examined using a Philips CM100 transmission electron microscope.

### **Microbial DNA Extraction and Bacteria Quantification**

DNA was extracted from mice stools using the DNeasy PowerSoil Kit (QIAGEN) according to the manufacturer's instructions. For qPCR quantification of differentially expressed bacteria, genomic DNA was amplified by qPCR with universal SYBR Green PCR Master Mix (Takara) on the ABI QuantStudio 7 Flex Real-Time PCR System, and normalized to 16S gene abundance. Primers used for detecting specific bacteria species are listed in **Table S6**. For *Desulfovibrio Fairfieldensis*, *Rhodococcus erythropolis*, *Chlamydia muridarum*, *Streptomyces violaceusniger*, *Pseudomonas sp.* Leaf58, primer was designed using NCBI/Genebank. When target Genebank sequences were chosen, “pick primers” was applied (nr database and Bacteria organism). Specificity of primers was examined using NCBI/Primer-Blast to confirm that no off-target bacteria could be amplified by the primers. Genebank numbers and targeted proteins were listed in **Table S6**. For *Brucella abortus*, the primer sequences were reported previously [1] and its specificity was confirmed using NCBI/Primer-Blast.

### **Patient-derived CRC organoids**

Primary CRC organoids (CRC-816, CRC-828) were obtained from Princess Margaret Living Biobank (Canada), and cultured in DMEM/F12 with N2 and B27 supplements

(ThermoFisher), 10mmol/L HEPES, 1.25mmol/L N-acetyl cysteine (Sigma-Aldrich, St Louis, MO), 10mmol/L SB202190-monohydrochloride (Sigma-Aldrich), R-spondin-1 (RSPO1), Noggin, Wnt3a, 50ng/mL EGF (ThermoFisher) and penicillin/streptomycin (Sigma-Aldrich). To perform SQLE knockout, organoids were digested into single cells and infected with SQLE Lenti-sgControl or sgSQLE for 1h under centrifugation (600g), followed by 6h incubation at 37°C. Infected organoids were selected with puromycin (2µg/ml) for 2 weeks. SQLE knockout was confirmed by Western blot. For drug treatment, organoids were cultured in culture medium containing 60µM terbinafine for 6 days, with the medium renewed every 2 days.

#### **Cell invasion assay**

Cell invasion was performed using 8µm transwell plates and layered with 100µL of BD Matrigel™ matrix (1:6 dilution) for 30mins at 37°C. Cells ( $5 \times 10^4$ /ml in 100ul FBS free medium) were added to the upper chamber, while 700µl of complete DMEM was added to the lower chamber. 16-24h later, the cells were fixed with 4% paraformaldehyde for 30min and stained with 1% crystal violet solution for 15min. Cells were counted under a light microscope. All experiments were conducted three times in triplicate.

#### **Cell migration assay**

For cell migration assay, cells ( $5 \times 10^5$ ) were seeded into a 6-well plates overnight. Then, the monolayers were scratched with a sterile 200µl tip. The width of the wounded area was captured using a microscope microscopy (Nikon, Tokyo, Japan) at 0 and 48 h. The

wound closure was measured with ImageJ software. All experiments were conducted three times in triplicate.

### **MTT assay**

For cell viability assays, cells were seeded in a 96-well plate and MTT assay (5mg/ml; Promega) was performed according to the manufacturer's instruction. All experiments were conducted three times in triplicate.

### **Apoptosis and cell cycle analyses**

For cell apoptosis, cells were analyzed 48h after seeding into 6-well plates. Apoptosis was assessed using the annexin-phycoerythrin/7-aminoactinomycin D staining kit (BD Biosciences). For cell cycle analysis, cells were serum-starved overnight and stimulated with complete medium for 4 to 8 hours. Cells were fixed in 70% ethanol, stained with propidium iodide, and analyzed by flow cytometry. All experiments were conducted three times in triplicate.

### **Germ-free mice**

Germ-free C57/BL6 mice at 6 weeks old were divided into 2 groups (9 per group). One group of mice were fed with stools collected from *Sqle* tg mice and the other group of mice were fed with stools collected from corresponding wildtype mice. Briefly, stool samples (1g) was homogenized in 5mL of sterile PBS. Then, the germ-free mice were transplanted with 200 $\mu$ L of the suspension using gastric gavage (twice a week). One

month later mice were sacrificed for further analysis.

### **Immunohistochemistry staining**

Paraffin slides were used for analyzing SQLE and Ki67 (1:100 dilution) protein levels. The proportion of SQLE-positive cells was determined by counting at least 1000 cells in 5 random microscopic fields for each sample. SQLE immunoreactivity was analyzed in a semi-quantitative manner by two independent pathologists, who were blinded to outcome. The intensity of immunohistochemical stain was scored from 0-10 (0: 0-10% of positive cells; 1:10%-20% positive cells; 2: 20%-30% positive cells; 3: 30%-40% positive cells; 4: 40%-50% positive cells; 5: 50%-60% positive cells; 6: 60%-70% positive cells; 7: 70%-80% positive cells; 8: 80%-90% positive cells; 9: 90%-100% positive cells). To identify the strong or weak of SQLE expression in microarray tissues, nuclear staining higher than 10% was defined as moderate/strong (high), whereas lower than 10% was defined as negative/weak (low). The signal of Ki-67-positive cells was determined by counting positive cells in crypts (5 random microscopic fields).

### **TUNEL staining**

TUNEL staining was determined on paraffin slides using the DeadEnd Colorimetric TUNEL System (Promega). Briefly, paraffin in slides was removed by xylene and fixed in 4% paraformaldehyde in PBS. Next, TdT reaction mix was added to tissue sections and incubated for 60min at 37°C in a humidified chamber. Streptavidin horseradish peroxidase (1:500 in PBS) was then added, followed by 100ul of 3,3'-diaminobenzidine

for detection.

### **Metabolomics profiling**

Stool samples (100mg) were dissolved in 500 $\mu$ L of ice-cold water, vortexed and centrifuged for 15min with 10,000g. Pellets were further extracted with 500 $\mu$ L cold methanol. Supernatants were combined for analysis. Chromatographic analyses were performed on a Waters Acquity UPLC system equipped with a BEHC18 column (Milford, USA). The mobile phases consisted of solvents A and B. The elution gradient program for fecal samples was: 0–1% B for 0.5 min; 1–30% B from 0.5 to 5 min; 30–50% B from 5 to 13 min and 50–100% B from 13 to 17 min; the flow rate was 0.45 mL/min. Mass spectra was obtained in a Waters SYNAPT G2 HDMS (Waters Corp., Manchester, UK) TOF mass spectrometer combined with an ESI source with positive ion scan mode. TOF parameters are as follow: capillary voltage: 3.0 KV; sample and extraction cone voltage: 40 V and 4.0 V; desolvation gas rate and temperature: 800 L/h and 400 °C; cone gas rate: 40 L/h; source temperature: 100 °C; scan time and inter scan delay: 0.15 and 0.02s. The lock mass in all analyses was leucine-enkephalin ([M+H]<sup>+</sup> = 556.2771) with a concentration of 0.5 $\mu$ g/mL and a flow rate of 10 $\mu$ L/min. Raw data were acquired using centroid mode; the mass range was set from m/z 50 to m/z 1200. Metabolomic data was analyzed using R package Metabo Analyst<sup>20</sup>. Altered metabolites were determined by a pair-wise comparison using Wilcoxon rank-sum test, with p-values adjusted by Benjamini-Hochberg false-discovery rate. Metabolites with adjusted P<0.1 and Log<sub>2</sub>(fold change)> 1 were considered significantly altered. Pair-

wise species-metabolites correlation coefficients were computed using Spearman's rank correlation coefficients (Spearman's rho). The strong correlation was determined by the cutoff  $|\text{Spearman's rho}| > 0.6$  and  $P < 0.025$ . All the heatmaps were plotted using R Package Complex Heatmap.



**Table S1. Clinicopathological features of CRC patients in Hong Kong cohort**

<b>Variables</b>	<b>SQLE expression (log2 of mRNA expression)</b>	<b>P value</b>
<b>Age</b>		
≥60 (N=73) (50.6%)	1.158	0.377
<60 (N=71) (49.4%)	1.257	
<b>Gender</b>		
Male (N=81) (56.3%)	1.329	0.503
Female (N=63) (43.7%)	1.051	
<b>Tumor site</b>		
Colon (N=75) (52.1%)	1.023	0.051
Rectum (N=69) (47.9%)	1.409	
<b>Differentiation</b>		
High/moderate (N=104) (72.2%)	1.119	0.858
Low (N=40) (27.8%)	1.187	
<b>TNM stage</b>		
I+II (N=60) (41.7%)	0.977	0.019 *
III+IV (N=84) (58.3%)	1.782	
<b>Metastasis</b>		
Yes (N=45) (31.3%)	1.673	0.036 *
No (N=99) (68.7%)	1.051	

**Table S2. Clinicopathological features of CRC patients in tissue microarray cohort**

Variables	SQLE expression		P value
	Strong (n, %)	Weak (n, %)	
<b>Age</b>			
≥67 (N=106) (51.6%)	49 (46.2%)	57 (53.8%)	0.539
<67 (N=101) (48.4%)	51 (50.4%)	50 (49.6%)	
<b>Gender</b>			
Male (N=117) (56.5%)	57(48.7%)	60 (51.3%)	0.893
Female (N=90) (43.5%)	43 (47.8%)	47 (52.2%)	
<b>Grade</b>			
Well/Moderate (N=194) (93.7%)	94 (48.5%)	100 (51.5%)	0.872
Poor (N=13) (6.3)	6 (46.1%)	7 (53.9%)	
<b>LN metastasis</b>			
N0 (N=96) (46.4%)	19 (19.8%)	77 (81.2%)	0.002 **
N1 (n=68) (32.8%)	38 (55.9%)	30 (44.1%)	
N2 (n=43) (20.8%)	37 (86.0%)	6 (14%)	
<b>Distance Metastasis</b>			
M0 (N=142) (68.6%)	50 (35.2%)	92 (64.8%)	<0.0001****
M1 (N=65) (31.4%)	50 (76.9%)	15 (23.1%)	
<b>Size (cm in diameter)</b>			
≥4.5 (N=127) (61.4%)	62 (48.8%)	65 (51.2%)	0.111
<4.5 (N=80) (38.6%)	30 (37.5%)	50 (62.5%)	
<b>AJCC stage</b>			
1+2 (N=82) (39.6%)	26 (31.7%)	50 (68.3%)	0.002 **
3+4 (N=125) (60.4%)	74 (59.2%)	57 (40.8%)	
<b>MSI</b>			
MSS (n=194) (93.7%)	93 (93%)	101 (94.4%)	0.778
MSI-H (n=13) (6.3%)	7 (7%)	6 (5.6%)	

**Table S3 Clinicopathological features of CRC patients in GSE17538 cohort**

<b>Variables</b>	<b>SQLE expression (log<sub>2</sub> of mRNA expression)</b>	<b>P value</b>
<b>Age</b>		0.659
≥65 (n=122)	5.428	
<65 (n=110)	5.427	
<b>Gender</b>		0.219
Male (n=122)	5.492	
Female (n=110)	5.395	
<b>Differentiation</b>		0.553
High and moderate (n=183)	5.411	
Low (n=30)	5.487	
<b>AJCC stage</b>		0.025 *
I+II (n=100)	5.031	
III+IV (n=132)	5.639	

**Table S4. List of primers**

<b>Gene</b>	<b>Primer</b>	<b>Sequence</b>
Muc2	Forward	GCCCGTGGAGTCGTACGTGC
	Reverse	TTGGGGCAGAGTGAGGCGGT
Ki67	Forward	GTGCTGACCCTGATGGGGAAGG
	Reverse	GCTCTTGCCCTGCCTGACACC
Gapdh	Forward	CCCTTAAGAGGGATGCTGCC
	Reverse	ACTGTGCCGTTGAATTTGCC
SQLE	Forward	<a href="#">GATGATGCAGCTATTTTCGAGGC</a>
	Reverse	<a href="#">CCTGAGCAAGGATATTCACGACA</a>
GAPDH	Forward	CTTACCACCATGGAGGAGGC
	Reverse	GGCATGGACTGTGGTCATGAG
DHCR7	Forward	CGCAGGACTTTAGCCGGT
	Reverse	TGTCATTGGTGACGCCATCT
DHCR24	Forward	TGAAGACAAACCGAGAGGGC
	Reverse	CAGCCAAAGAGGTAGCGGAA
FDFT1	Forward	CCACCCCGAAGAGTTCTACAA
	Reverse	TGCGACTGGTCTGATTGAGATA
FDPS	Forward	CAGCTTTCTACTCCTTCTACCTCC
	Reverse	GCTCCTTCTCGCCATCAAT
HMGR	Forward	TGATTGACCTTCCAGAGCAAG
	Reverse	CTAAAATTGCCATTCCACGAGC
HMGS1	Forward	CATTAGACCGCTGCTATTCTGTC
	Reverse	TTCAGCAACATCCGAGCTAGA
IDI1	Forward	TGGATAAAACCCCTGTGGTG
	Reverse	CAACATCCGGCATAACTGTG
LDLR	Forward	TACAAGTGGGTCTGCGATGG
	Reverse	TGAAGTCCCCGGATTTGCAG
MVD	Forward	GTAAGTGGCTGTGGAGCTGG
	Reverse	GGAGTTGATGGGCAGAACCA
MVK	Forward	CTCTGATTGGCTGGCCTGAA
	Reverse	CCAACCTCCACAACCCAGAG
PMVK	Forward	GCTGATGTCTGTGCTGTCCT
	Reverse	GAAAGGCCTCCTTGTAGGTG

**Table S5. List of antibodies**

<b>Antibodies</b>	<b>Cat. Number</b>	<b>Company</b>
SQLE	12544-1-AP	Proteintech
JAM-C	Ab81331	Abcam
Occludin	33-1500	Invitrogen
Caspase-3	9665s	Cell signaling
Cleaved Caspase-3	9661s	Cell signaling
Caspase-7	9492s	Cell signaling
Cleaved Caspase-7	9491s	Cell signaling
P53	Sc126	Santa Cruz
Cyclin D1	2922s	Cell signaling
PCNA	2586	Cell signaling
P21	Sc6246	Santa Cruz
PARP	Sc8007	Santa Cruz
Cleaved PARP	5625	Cell signaling
CDK4	12790	Cell signaling
p27 <sup>Kip1</sup>	3686	Cell signaling
GAPDH	Sc4772	Santa Cruz
HMGCR	A1633	ABclonal
FDPS	A5744	ABclonal
FDFT1	A4651	ABclonal

**Table S6. List of primers for specific bacteria**

Gene name	Primers	Sequence
<i>Brucella abortus</i>	Forward	CATGCGCTATGTCTGGTTAC
	Reverse	GGCTTTTCTATCACGGTATTC
<i>Desulfovibrio fairfieldensis</i> (CP014229.1) Hypothetical protein	Forward	GGCAACAGCCTGGTTATGAT
	Reverse	ACGAAGATGGACGAAAGCAT
<i>Rhodococcus erythropolis</i> (CP050124.1) Non-ribosomal peptide synthetase	Forward	AGCACATATCGACCCAGGAC
	Reverse	CGAGGTGAAAGTGCAGATCA
<i>Chlamydia muridarum</i> (NZ_CP009760.1) ABC transporter substrate-binding protein	Forward	GCGACGCATATGTTCCGGAGA
	Reverse	TGTGCTGAGGTTGGGACTTG
<i>Streptomyces violaceusniger</i> (CP002994.1) TPR-repeat protein	Forward	AGGTCGCCCTCAGCTCTC
	Reverse	AACGGGATCAGGGGTAGGT
<i>Pseudomonas sp. Leaf58</i> (NZ_CP032677.1) MSF transporter	Forward	CGTGGTGCTGTCGCTACTAC
	Reverse	CAAGCCAATACCAATCGACA
Universe QPCR-Eub341F	Forward	ACTCCTACGGGAGGCAGCAGT
Universe QPCR-Eub534R	Reverse	ATTACCGCGGCTGCTGGC

**Reference:**

1. Redkar R, Rose S, Bricker B, DelVecchio V. Real-time detection of *Brucella abortus*, *Brucella melitensis* and *Brucella suis*. *Mol Cell Probes*. 2001 Feb;15(1):43-52.