This is a repository copy of *AQP5 enriches for stem cells and cancer origins in the distal stomach*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/155846/

Version: Accepted Version

1 **AQP5 Enriches for Stem Cells and Cancer Origins in the Distal Stomach**

2 Si Hui Tan[1$], Yada Swathi[1$], Shawna Tan[1], Jasmine Goh[1], Ryo Seishima[1], Kazuhiro

3 Murakami[2], Masanobu Oshima[3], Toshikatsu Tsuji[4], Phyllis Phuah[1], Liang Thing Tan[1], Esther

4 Wong[1], Aliya Fatehullah[1], Taotao Sheng[5,6], Shamaine Wei Ting Ho[5,7], Heike Grabsch[8,9],

5 Supriya Srivastava[10,11], Ming Teh[10], Simon L.I.J. Denil[12], Seri Mustafah[13], Patrick Tan[5], Asim

6 Shabbir[14], Jimmy So[14], Khay Guan Yeoh[11], and Nick Barker[1, 2, 15, *]

7

8 [1] A*STAR Institute of Medical Biology, Singapore, 138648 Singapore

9 [2] Division of Epithelial Stem Cell Biology, Cancer Research Institute, Kanazawa University,

10 Kakuma-machi Kanazawa, 920-1192, Japan

11 [3] Division of Genetics, Cancer Research Institute, Nano-Life Science Institute, Kanazawa

12 University, Kakuma-machi, Kanazawa 920-1192, Japan

13 [4] Department of Gastroenterological Surgery, Ishikawa Prefectural Central Hospital,

14 Kuratsuki-Higashi 2-1, Kanazawa 920-8530, Japan

15 [5] Cancer and Stem Cell Biology Program, Duke-NUS Graduate Medical School, Singapore,

16 169857 Singapore

17 [6] Department of Biochemistry, National University of Singapore, Singapore, 117596 Singapore

18 [7] Cancer Science Institute of Singapore, National University of Singapore, Singapore, 117599

19 Singapore

20 [8] Department of Pathology, GROW School for Oncology and Developmental Biology,

21 Maastricht University Medical Center+, Maastricht, NL

22 [9] Pathology and Data Analytics, Leeds Institute of Medical Research at St. James's, University

23 of Leeds, Leeds, UK

24 [10] Department of Pathology, National University Health System, Singapore, 119228 Singapore

25   [11] Department of Medicine, National University Health System, Singapore, 119228 Singapore

26   [12] Skin Research Institute of Singapore, Singapore 138648 Singapore

27   [13] A*STAR Singapore Immunology Network, Singapore, 138648 Singapore

28   [14] Department of Surgery, National University Health System, Singapore, 119228 Singapore

29   [15] School of Biological Sciences, Nanyang Technological University, Singapore, 308232

30   Singapore

31   [$] Equal Contribution

32   * Corresponding author

33   **Running title: AQP5 Enriches Stem Cells in Distal Stomach**

34

35

Lgr5 marks resident adult epithelial stem cells at the gland base in the mouse pyloric stomach[1], but the identity of the equivalent human stem cell population remains elusive due to a lack of surface markers facilitating its prospective isolation and validation. Lgr5+ intestinal stem cells are major sources of cancer following Wnt pathway hyperactivation in mice[2]. However, the contribution of pyloric Lgr5+ stem cells to gastric cancer following Wnt pathway dysregulation, a frequent event in human gastric cancer[3], is unknown. Here, we employed comparative profiling of Lgr5+ stem cell populations along the mouse gastrointestinal tract to identify, then functionally validate the membrane protein AQP5 as a marker that enriches for mouse and human adult pyloric stem cells. We show that stem cells within the Aqp5+ compartment are a source of Wnt-driven, invasive gastric cancer *in vivo* using new *Aqp5*-CreERT2 mouse models. Additionally, tumour-resident Aqp5+ cells can selectively initiate organoid growth *in vitro*, indicating that this population harbours potential cancer stem cells. In human gastric cancer, AQP5 is robustly expressed in primary and metastases of intestinal and diffuse subtypes, often displaying altered cellular localization compared to healthy tissue. These new markers and mouse models will be an invaluable resource for deciphering early gastric cancer formation and for isolating and characterizing human stomach stem cells as a prerequisite to potentially harnessing their regenerative medicine potential in the clinic.

Comparative profiling yields new markers

To identify markers specific to Lgr5[Hi] pyloric stem cells within the gastrointestinal tract, we profiled the transcriptomes of qPCR-validated Lgr5-EGFP[Hi] (Lgr5+ stem cells), Lgr5-EGFP[Lo] (immediate progeny) and unfractionated populations from the small intestine, colon, gastric pylorus of Lgr5-EGFP-IRES-CreERT2 mice[4] by microarray, and identified genes selectively enriched in Lgr5[Hi] pyloric stem cells (Fig. 1a; Extended Data Fig. 1a-d, Supplementary Table 1). This dataset also revealed the transcriptional signature of Lgr5[Hi] colon stem cells (Supplementary Table 2). Profiling of Lgr5-EGFP[Hi], Lgr5-EGFP[Lo] and Lgr5-EGFP[Neg] pylorus

63   populations from another Lgr5-reporter model, Lgr5-DTR-EGFP[5], revealed 67 overlapping

64   genes (Fig. 1b).

65   Candidate markers were empirically validated by qPCR, *in situ* hybridization (ISH) and

66   immunostaining. Optimal candidates presented enriched expression in the Lgr5-EGFP[Hi]

67   pyloric population by qPCR, robust, localized mRNA/protein expression within the Lgr5+

68   pyloric gland base[1], minimal expression in intestines and gastric corpus and co-expression

69   with *Lgr5* (Extended Data Fig. 1j-n). Six candidates were selected: Alpha-1,4-N-

70   Acetylglucosaminyltransferase (A4gnt), Aquaporin-5 (Aqp5), Gastric intrinsic factor (Gif),

71   Mucin6 (Muc6), Solute carrier protein 9a3 (Slc9a3/Nhe3) and Secreted phosphoprotein 1

72   (Spp1/Osteopontin) (Fig. 1c-m; Extended Data Fig. 1e-n, Supplementary Table 3).

73   To validate the six candidates as markers of pyloric stem cell-enriched populations, we

74   generated EGFP-IRES-CreERT2 mouse models for *Aqp5*, *A4gnt*, *Spp1* and *Slc9A3*

75   (Extended Data Fig. 2a), and Aqp5-2A-EGFP, Slc9A3-2A-EGFP, Aqp-2A-CreERT2 and

76   Slc9A3-2A-CreERT2 mice in which endogenous gene expression is unaffected (Extended

77   Data Fig. 2b). GFP expression in the pylori of the 2A-EGFP models recapitulated endogenous

78   gene expression in the pylorus and small intestine (Extended Data Fig. 2c-f). Additionally,

79   97.4% concurrence between the Aqp5+ and Slc9a3+ pyloric populations by co-staining

80   (Extended Data Fig. 2g-h) reaffirms that they effectively label the same population.

81   The *in vivo* contribution of the gland populations expressing the candidate genes to epithelial

82   renewal was evaluated via lineage tracing in CreERT2;Rosa26-tdTomato[LSL] lines. tdTomato

83   (dTom) expression was first observed at the gland bases in all lines 20-48 hours post-

84   Tamoxifen induction, confirming the expected Cre expression domain (Extended Data Fig. 3a-

85   d, q-r). After several months (tissue turnover spans 7-10 days[1]), multiple entirely dTom+

86   glands were evident throughout the pylorus, documenting the long-term self-renewal and

87   multipotency of the cells expressing *Aqp5*, *Slc9A3*, *Spp1* or *A4gnt* (Extended Data Fig. 3e-h,

88   s-t). Importantly, no intestinal tracing was observed, except for transient reporter expression

89   within the villi of Slc9A3-Cre models (Extended Data Fig. 2f, 3i-x). These observations confirm

90   that populations expressing *Aqp5*, *Spp1*, *Slc9a3* and *A4gnt* contain pyloric stem cells.

91

92   **Aqp5 enriches for stem cells in mice**

93   Aqp5 water channel protein[6] emerged as a promising candidate for isolating pyloric stem cells

94   from mice and humans. EGFP expression was restricted to pyloric gland bases in Aqp5-

95   EGFP-IRES-CreERT2 and Aqp5-2A-EGFP mouse models (Extended Data Fig. 2c,w). Sorted

96   EGFP+ cells from adult Aqp5-EGFP-IRES-CreERT2 mice (Fig. 2a) showed a 9-fold and 15-

97   fold enrichment of *Aqp5* and *Lgr5* transcripts respectively over EGFP- cells (Extended Data

98   Fig. 2u-v). By immunostaining, endogenous Aqp5 protein co-localized with EGFP (Extended

99   Data Fig. 2w). Thus, the Aqp5 models faithfully report *Aqp5* expression in pyloric gland bases.

100   Next, we found that Aqp5+ population overlapped with Gif and Ki67 but not Gastrin (Gast),

101   Chromogranin A (Chga) or Mucin5ac (Muc5ac), whilst the Lgr5-EGFP^{Hi} population expressed

102   major gastric lineage markers (Gif, Gast, Chga) and Ki67 (Fig. 2b-e, Extended Data Fig. 2k-

103   m, x). This observation was confirmed via single-cell analysis of Lgr5-EGFP^{Hi} pyloric stem

104   cells by CEL-Seq/RaceID[7]. The Lgr5-EGFP^{Hi} compartment comprised three subpopulations:

105   the major subpopulation co-expressed some of the new pyloric markers – Aqp5, Gif and Muc6

106   (Extended Data Fig. 2n-q), while the two minor subpopulations expressed Chga/Gast and

107   Krt8/Krt18 (Extended Data Fig. 2r-s). Proliferation marker expression was significantly

108   enriched in the major subpopulation (Extended Data Fig. 2t). Aqp5 staining on Lgr5-DTR-

109   EGFP pylori revealed 94.1% overlap between the two populations (Extended Data Fig. 2i-j),

110   underscoring the CEL-Seq finding that Aqp5 marks the major subpopulation of Lgr5^{Hi} cells.

111   Lineage tracing using adult Aqp5-EGFP-IRES-CreERT2;Rosa26-tdTomato^{LSL} (Aqp5-IRES-

112   CreERT2;dTom) mice detailed the homeostatic behaviour of Aqp5+ cells – dTom+ cells

113   appeared exclusively at the gland bases 20 hours post-tamoxifen induction (Fig. 2f, Extended

114   Data Fig. 3y), expanded to clones reaching gland surfaces by 5 days (Fig. 2g), and persisted

115   for 1 year, demonstrating self-renewal of Aqp5+ cells (Fig. 2h-i, Extended Data Fig. 3z,c').

116   Uninduced controls presented negligible dTom+ clones (Extended Data Fig. 3a'-b'). Six

117   months post-induction, dTom+ clones comprised the major pyloric lineages expressing Gif,

118   Gast, Chga and Muc5ac (Fig. 3j-m), confirming multipotency in the Aqp5+ population. Ablating

119   endogenous Aqp5+ cells using a new Aqp5-2A-DTR model severely disrupted the gland

120   bases (Extended Data Fig. 4k-r).

121   Tracing was absent from corpus, small intestine and colon, except for Brunner's glands

122   (Extended Data Fig. 3d'-h'). Our Aqp5-2A-EGFP and Aqp5-CreERT2 models also faithfully

123   reported endogenous Aqp5 expression in tissues such as cornea, lung, mammary gland and

124   salivary gland[13,14] (data not shown).

125   To evaluate Aqp5's utility as a marker for isolating enriched pyloric stem cell populations, we

126   sorted Aqp5+ and Aqp5- cells from adult wild-type mice using anti-Aqp5 antibody (Fig. 2n,

127   Extended Data Fig. 4a-b) to profile their transcriptomes and evaluate their *in vitro* organoid-

128   forming capacity. *Aqp5* and *Lgr5* were markedly enriched in Aqp5+ population relative to

129   Aqp5- population by qPCR and microarray (Fig. 2o-p, Extended Data Fig. 4d-e). Aqp5+ and

130   Lgr5-EGFP[Hi] transcriptomes are highly correlated by GSEA analysis (FDR p-value<0.001,

131   Extended Data Fig. 4c), with the Aqp5+ population presenting strong enrichment of our new

132   markers (Fig. 2q; Extended Data Fig. 4f) and published gland base markers, *Lrig1*[8] and

133   *Runx1*[9] (Extended Data Fig. 4g). *Axin2*[10], *Cck2r*[11] and *Sox2*[12] were not enriched in the Aqp5+

134   population, consistent with their relatively broad expression within pyloric glands (Extended

135   Data Fig. 4g). Concurring with immunostaining, Aqp5+ cells presented high *Gif*, moderate

136   *Ki67* and no *Muc5ac* expression; there was no significant difference in *Gast* and *Chga*

137   expression between Aqp5+ and Aqp5- populations, likely reflecting the limited numbers of

138   Gastrin+ G cells and Chga+ endocrine cells within the Aqp5- population (Extended Data Fig.

139   4h). Therefore, this antibody-based strategy facilitates enrichment of mouse pyloric stem cells,

140   independent of fluorescent reporters.

141   Compared to Aqp5- cells, Aqp5+ cells generated three-fold more organoids (0.64% vs 2.58%

142   respectively) that could be maintained long-term, while the few organoids derived from Aqp5-

143   cells died within 3 weeks (Fig. 2r-t). Organoid initiation frequencies from Aqp5+ cells (2.58%)

144   and Lgr5-EGFP+ cells (3.09%) from Lgr5-2A-EGFP mice were similar (Extended Data Fig.

145   4s-u), indicating high functional overlap. Aqp5+ cell-derived organoids showed heterogeneous

146   Aqp5 expression, which partially overlapped with Ki67 (Extended Data Fig. 4i-j), similar to its

147   *in vivo* pattern. Withdrawal of WNT3A, FGF10 and NOGGIN (Fig. 2u) resulted in

148   downregulation of stem cell markers *Lgr5*, *Aqp5* and upregulation of differentiation marker

149   *Muc5ac* in the organoids after 3 days (Fig. 2v-x). Therefore, Aqp5 is a useful marker for the

150   prospective isolation of enriched murine pyloric stem cells.

151

152   **AQP5 enriches for stem cells in humans**

153   We sought to evaluate AQP5 as a marker facilitating enrichment of human pyloric stem cells.

154   AQP5 was exclusively expressed at human pyloric gland bases, overlapping with *LGR5* and

155   other mouse pyloric markers, *MUC6*, *A4GNT* and *SLC9A3* (Fig. 3a-d, Extended Data Fig. 5a-

156   f'). A minor proportion of the human AQP5+ cells were KI67+ (Fig. 3e-f), reminiscent of the

157   murine pylorus (Extended Data Fig. 2x). Human *AQP5+* cells overlapped with *PEPC+* and

158   *MUC6+* populations, but, not *GIF+* parietal cells or *MUC5AC+* foveolar cells (Extended Data

159   Fig. 5g-j).

160   We next sorted AQP5+ cells from healthy human pyloric specimens using anti-AQP5 antibody

161   and verified 10.2-fold enriched *AQP5* expression in AQP5+ cells by qPCR (Fig. 3g-h). AQP5+

162   cells routinely established organoids that were passaged for >3 months, whereas AQP5- cells

163   never initiated organoids (Fig. 3i-j). The human pyloric organoids expressed AQP5

164   heterogeneously, partially overlapping with KI67 (Extended Data Fig. 5k-l). Withdrawal of

165   WNT3A, NOGGIN and FGF10 resulted in reduced *LGR5* and *AXIN2*, and elevated *MUC5AC*

166   and *TFF2* expression, indicating differentiation towards mucous lineages (Fig. 3k-l, Extended

167   Data Fig. 5m-o).

168   RNASEQ was then performed on FACS-isolated AQP5+ pyloric cells from healthy human

169   pylori, and the top hits validated by qPCR. We identified >500 differentially expressed genes

170   that were significantly up/down-regulated by >4-fold in AQP5+ versus AQP5- populations (Fig.

171   3m, top candidates in Supplementary Table 3). Q-PCR analysis validated 18 candidates as

172   being enriched in the AQP5+ population (Extended Data Fig. 6a-e). Selected murine pyloric

173   stem cell markers, including *AQP5*, *A4GNT* and *MUC6*, and an intestinal stem cell marker,

174   *SMOC2*[15], were upregulated in the AQP5+ fraction by RNASEQ and qPCR (Extended Data

175   Fig. 6a). ISH confirmed *SMOC2* as being expressed in a subset of gland base cells (Extended

176   Data Fig. 6f-f"). Gene Onotology analysis revealed that approximately half of the candidates

177   were membrane expressed and had protein-binding activity (Fig. 3n, Extended Data Fig. 6b),

178   suggesting that additional markers could be identified from our list to potentially further enrich

179   for pyloric stem cells. Moreover, components of key signalling pathways (e.g. Wnt and Notch),

180   chemokine signalling and extracellular matrix, and some uncharacterized genes, were also

181   enriched in the AQP5+ population, highlighting the additional biological insight to be gained

182   from the profile (Fig. 3n; Extended Data Fig. 6c-g).

183   Collectively, these data demonstrate AQP5's utility as a marker for isolating enriched

184   populations of endogenous stem cells from human stomach epithelia for downstream

185   purposes.

186

187   **Aqp5+ cells as source of gastric cancer**

188   We targeted conditional oncogenic mutations to pyloric stem cell-enriched populations using

189   our new CreERT2 mouse models to evaluate their contribution to gastric cancer and

190   circumvent the rapid lethality of Lgr5+ cell-driven cancer models[2]. To determine how often

191   pathways are co-dysregulated in gastric cancer, we analysed transcriptomes of gastric cancer

192   patients from TCGA (n=155) and East Asian cohorts (n=42) for Wnt/β-catenin, PI3K and Ras

193   signalling activities (Extended Data Fig. 7a-b), whose components are frequently mutated in

194   gastric cancer[3,16]. Using published gene signatures[17–19], we found that Wnt/β-catenin

195   signalling was commonly hyper-activated in human gastric cancer (>80% in both patient

196   cohorts), and frequently co-occurred with hyper-activation of PI3K and/or Ras signalling (57.1-

197   64.3%)(Extended Data Fig. 7). We thus recapitulated these co-dysregulated pathways by

198   crossing our pyloric CreERT2 drivers to conditional APC, PTEN and Kras^{G12D} alleles and

199   induced hyper-activation of the pathways at adulthood.

200

201   All mouse models developed sizeable tumours exclusively in the pylorus, with latencies

202   ranging from 1-11 months post-induction (n=30/34, Extended Data Fig. 8a-g). In the tumours,

203   which were classified as tubular-type gastric adenocarcinomas (WHO), malignant structures

204   surrounded by stroma and inflammatory cells replaced normal glands. Sole hyper-activation

205   of Kras^{G12D} did not produce pyloric tumours (Extended Data Fig. 8h-i). Pylori of uninduced

206   mice of the same cancer genotype lacking Cre were normal (Extended Data Fig. 8j).

207   Across all oncogene combinations, the tumor incidence in 2A-CreERT2 models was 100%

208   compared to 82.6% in IRES-CreERT2 models (Fig. 4a), reflecting differences in Cre activation

209   efficacies. The 2A-CreERT2 models displayed almost contiguous tumour growth, contrasting

210   the multi-focal lesions in IRES-CreERT2 models (Extended Data Fig. 8b-g). Hyper-activation

211   of Wnt/β-catenin signalling alone [APC$^{fl/fl}$ (A)] was sufficient to drive tumourigenesis, and co-

212   activation of PI3K and/or Kras pathways [APC$^{fl/fl}$;PTEN$^{fl/fl}$ (AP); APC$^{fl/fl}$;PTEN$^{fl/fl}$; Kras^{G12D}

213   (APK)] accelerated tumour development and progression (Extended Data Fig. 8a-g). We also

214   observed focal invasions through the muscularis mucosae in A and AP models (IRES-

215   CreERT2: 15.8%, 2A-CreERT2: 66.7%; Fig. 4a). As expected, intestinal tumors were never

216   observed (Extended Data Fig. 8k-l).

217   We characterized all tumours to detail pathway activation, proliferation status, lineage marker

218   expression and epithelia/stroma constitution. As there is no major phenotypic differences

219   between the different models, we present data from Aqp5-IRES-CreERT2;APK tumours, and

220   focal invasions from Slc9a3-2A-CreERT2;AP tumours (Fig. 4b-c, i). The pyloric tumours and

221   focal invasions were predominantly dTom+, confirming Aqp5+ or Slc9A3+ cells as their origins

222   (Extended Data Fig. 9b, j). In contrast to the normal pylorus (Extended Data Fig. 9p-w), the

223   gastric adenocarcinomas presented hyper-activation of Wnt/ß-catenin, MAP kinase and

224   phospho-AKT pathways, evidenced by elevated levels of nuclear/cytoplasmic ß-catenin,

225   MAPK and phospho-AKT expression respectively (Fig. 4f-h, m-o; Extended Data Fig. 9a).

226   These regions were also highly proliferative, lacked Gif, Gast and Muc5ac expression

227   (Extended Data Fig. 9d-g, l-o). In the Aqp5-IRES-CreERT;APK model, dTom+ cells that

228   retained E-cadherin expression were also found throughout the tumour stroma (Fig. 4e). In

229   the Slc9a3-2A-CreERT2;AP model, E-cadherin+ cells infiltrated through the muscularis

230   mucosae (Fig. 4l). Immunostaining for Aqp5-GFP reporter expression revealed *Aqp5*

231   expression in a subpopulation of the pyloric tumours, some of which were Ki67+ (Fig. 4d, k,

232   p). Many of the Aqp5+ cells in the tumours co-expressed *Lgr5*, with elevated *Aqp5* expression

233   in tumours compared to adjacent normal mucosa (Extended Data Fig. 9h-h"). There was a

234   low incidence of tumours within non-gastrointestinal organs, such as the salivary gland (<25%

235   in Aqp5-IRES-CreERT2-driven cancer models), which did not impact survival to preclude

236   gastric adenocarcinoma development (Extended Data Fig. 9i-i'). Thus, our mouse models

237   support pyloric stem cell-enriched populations as being a source of invasive, Wnt-driven

238   gastric cancer and are valuable for modelling gastric cancer *in vivo*.

239

240   **Aqp5+ tumour cells display *ex vivo* stemness**

241   We then determined if Aqp5+ tumour cells behave differently from their Aqp5- counterparts.

242   After confirming that the Aqp5-GFP reporter recapitulated endogenous Aqp5 expression in

243   the tumours (Extended Data Fig. 9x-x'), we cultured sorted Aqp5-GFP+ and GFP- cells

244   (Extended Data Fig. 9y-z"). Aqp5-GFP+ tumour cells reproducibly generated organoids that

245   could be serially propagated in the absence of exogenous growth factors, while Aqp5-GFP-

246   tumour populations never produced organoids, despite containing Ki67+ cells (Fig. 4p-s).

247   Moreover, though Aqp5-GFP+ normal cells could initiate organoids with growth factors, they

248   died upon growth factor removal (Fig. 4r-s). These data suggest that the stem potential of the

249   tumour is found within the Aqp5+ compartment.

250

251   **AQP5 expression in human gastric cancer**

252   We surveyed AQP5 expression in human gastric cancer by immunostaining on a tissue

253   microarray of 145 distal gastric cancer samples comprising intestinal, diffuse and mixed

254   subtypes with variable grades of differentiation. AQP5 was expressed in most intestinal,

255   diffuse and mixed cases (Extended Data Fig. 10a-e). Contrasting normal pylorus, 96.1% of

256   the tumour samples displayed cytoplasmic AQP5 expression, 37.9%, 3.9% and 37.9% had

257   membranous, nuclear and multiple localizations of AQP5 respectively (Extended Data Fig.

258   10a-e). While intracellular AQP5 localization has been reported in other cancers[20–22], its

259   functional significance is unknown.

260   In full sections of 54 advanced human distal gastric adenocarcinomas and 12 metastatic

261   lesions, most expressed AQP5 (Extended Data Fig. 10f). Cytoplasmic AQP5 was observed in

262   all AQP5+ samples, while membranous/luminal, nuclear and multiple sites of AQP5

263   localization were found in 53.7%, 9.8% and 53.7% of the samples respectively (Extended Data

264   Fig. 10f-j). In 67.7% of the sections, submucosal malignant cells expressed more AQP5 than

265   their mucosal counterparts (Extended Data Fig. 10g-j, o, p). AQP5 was expressed in poorly

266   cohesive tumour cells in 70% of the cohort (Extended Data Fig. 10j, o). All AQP5+ cells in the

267   submucosa retained E-Cadherin expression, and a subset co-expressed KI67 (Extended Data

268   Fig. 10k-n). 46.2% of intestinal metaplasia cases, which is strongly correlated with gastric

269   cancer[23,24], displayed mild to moderate AQP5 expression (Extended Data Fig. 10o, q). AQP5

270   was also weakly expressed in a subset of signet ring cells in 66.7% of the samples (Extended

271   Data Fig. 10o, r). 83.3% of the metastatic lesions with AQP5+ primary tumours harboured

272   AQP5+ tumour cells in the lymph node (Extended Data Fig. 10o, s).

273   Our broad survey shows that AQP5 is commonly expressed in primary and metastases of

274   intestinal and diffuse subtypes of gastric cancer.

275

276   Efforts to exploit the therapeutic potential of stem cells require functionally validated markers

277   for their prospective isolation. For the first time, we purify enriched populations of human

278   pyloric stem cells and demonstrate the direct contribution of murine Aqp5+ cells to gastric

279   carcinogenesis. Though the expression of secretory markers like Gif and Muc6 may not

280   conform to the classical "undifferentiated" stem cell dogma, it is well established in liver and

281   lung that more specialized cells can serve as homeostatic stem cells[25–28], an emerging trend

282   in epithelial organs[29]. While Aqp5 is known for regulating water transport in healthy tissues[6],

283   it is increasingly implicated in cancers as a driver of proliferation and invasiveness *in vitro*[30–

284   [34]. Various human cancers, including gastric, breast, soft tissue sarcoma, lung, oesophageal

285   and colorectal cancers, present high AQP5 expression[21,22,33,35–37]. We show that AQP5 is

286   expressed in most human primary tumours and metastases of intestinal and diffuse gastric

287   cancer subtypes. Interestingly, we found that tumour-resident Aqp5+ cells in our gastric cancer

288   mouse model selectively exhibited *ex vivo* stem potential, indicating that the Aqp5+ tumour

289   population harbours cancer stem cells. Future evaluation of the stem potential of AQP5+ in

290   human gastric cancers using our new antibody-based isolation protocols will potentially reveal

291   novel opportunities for developing more effective cancer therapeutics.

292

References

294   1.   Barker, N. *et al.* Lgr5+ve Stem Cells Drive Self-Renewal in the Stomach and Build
295   Long-Lived Gastric Units In Vitro. *Cell Stem Cell* **6**, 25–36 (2010).

296   2.   Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature*
297   (2009). doi:10.1038/nature07602

298   3.   Network, T. C. G. A. R. *et al.* Comprehensive molecular characterization of gastric
299   adenocarcinoma. *Nature* **513**, 202 (2014).

300   4.   Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker
301   gene Lgr5. *Nature* **449**, 1003–1007 (2007).

302   5.   Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive
303   cells dispensable. *Nature* **478**, 255–9 (2011).

304   6.   Direito, I., Madeira, A., Brito, M. A. & Soveral, G. Aquaporin-5: from structure to
305   function and dysfunction in cancer. *Cell. Mol. Life Sci.* **73**, 1623–1640 (2016).

306   7.   Grun, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell
307   types. *Nature* **525**, 251–255 (2015).

308   8.   Choi, E. *et al.* Lrig1+ gastric isthmal progenitor cells restore normal gastric lineage
309   cells during damage recovery in adult mouse stomach. *Gut* gutjnl-2017-313874
310   (2017). doi:10.1136/gutjnl-2017-313874

311   9.   Matsuo, J. *et al.* Identification of Stem Cells in the Epithelium of the Stomach Corpus
312   and Antrum of Mice. *Gastroenterology* **152**, 218-231.e14 (2017).

313   10.   Sigal, M. *et al.* Stromal R-spondin orchestrates gastric epithelial stem cells and gland
314   homeostasis. *Nature* **548**, 451–455 (2017).

315   11.   Hayakawa, Y. *et al.* CCK2R identifies and regulates gastric antral stem cell states and
316   carcinogenesis. *Gut* **64**, 544–553 (2015).

317   12.   Arnold, K. *et al.* Sox2 + adult stem and progenitor cells are important for tissue
318   regeneration and survival of mice. *Cell Stem Cell* (2011).
319   doi:10.1016/j.stem.2011.09.001

320   13.   Funaki, H. *et al.* Localization and expression of AQP5 in cornea, serous salivary
321   glands, and pulmonary epithelial cells. *Am. J. Physiol.* (1998).

322   14.   Matsuzaki, T. *et al.* Expression and immunolocalization of water-channel aquaporins
323   in the rat and mouse mammary gland. *Histochem. Cell Biol.* (2005).
324   doi:10.1007/s00418-005-0753-x

325   15.   Muñoz, J. *et al.* The Lgr5 intestinal stem cell signature: Robust expression of
326   proposed quiescent ′ +4′ cell markers. *EMBO J.* (2012). doi:10.1038/emboj.2012.166

327   16.   Muratani, M. *et al.* Nanoscale chromatin profiling of gastric adenocarcinoma reveals
328   cancer-associated cryptic promoters and somatically acquired regulatory elements.
329   *Nat. Commun.* (2014). doi:10.1038/ncomms5361

330   17.   Zhang, Y. *et al.* A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway
331   Alterations. *Cancer Cell* (2017). doi:10.1016/j.ccell.2017.04.013

332   18.   Pek, M. *et al.* Oncogenic KRAS-associated gene signature defines co-targeting of
333   CDK4/6 and MEK as a viable therapeutic strategy in colorectal cancer. *Oncogene*
334   (2017). doi:10.1038/onc.2017.120

335   19.   Van der Flier, L. G. *et al.* The Intestinal Wnt/TCF Signature. *Gastroenterology* (2007).
336   doi:10.1053/j.gastro.2006.08.039

337   20.   Lambertz, N. *et al.* Expression of aquaporin 5 and the AQP5 polymorphism A(-1364)C
338   in association with peritumoral brain edema in meningioma patients. *J. Neurooncol.*
339   (2013). doi:10.1007/s11060-013-1064-z

340   21.   Shimizu, H. *et al.* The expression and role of Aquaporin 5 in esophageal squamous
341   cell carcinoma. *J. Gastroenterol.* (2014). doi:10.1007/s00535-013-0827-9

342   22.   Zhu, Z. *et al.* Expression of AQP3 and AQP5 as a prognostic marker in triple-negative
343   breast cancer. *Oncol. Lett.* **16**, 2661–2667 (2018).

344   23.   Kinoshita, H., Hayakawa, Y. & Koike, K. Metaplasia in the stomach—precursor of
345   gastric Cancer? *International Journal of Molecular Sciences* (2017).
346   doi:10.3390/ijms18102063

347   24.   Jencks, D. S. *et al.* Overview of current concepts in gastric intestinal metaplasia and
348   gastric cancer. *Gastroenterol. Hepatol.* (2018).

349   25.   Rawlins, E. L., Clark, C. P., Xue, Y. & Hogan, B. L. M. The Id2+ distal tip lung
350   epithelium contains individual multipotent embryonic progenitor cells. *Development*
351   (2009). doi:10.1242/dev.037317

352   26.   Barkauskas, C. E. *et al.* Type 2 alveolar cells are stem cells in adult lung. *J. Clin.*
353   *Invest.* (2013). doi:10.1172/JCI68782

354   27.   Desai, T. J., Brownfield, D. G. & Krasnow, M. A. Alveolar progenitor and stem cells in
355   lung development, renewal and cancer. *Nature* (2014). doi:10.1038/nature12930

356   28.   Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2(+)
357   cells fuel homeostatic renewal of the liver. *Nature* **524**, 180–5 (2015).

358   29.   Post, Y. & Clevers, H. Defining Adult Stem Cell Function at Its Simplest: The Ability to
359   Replace Lost Cells through Mitosis. *Cell Stem Cell* **25**, 174–183 (2019).

360   30.   Kang, S. K. *et al.* Aquaporin 5 (AQP5) is a novel signaling molecule triggering
361   Ras/ERK/retinoblastoma (Rb) signaling pathway in colon cancer cell lines. *Mol.*
362   *Cancer Ther.* **6**, B115 LP-B115 (2007).

363   31.   Woo, J. *et al.* The effect of aquaporin 5 overexpression on the Ras signaling pathway.
364   *Biochem. Biophys. Res. Commun.* (2008). doi:10.1016/j.bbrc.2007.12.073

365   32.   Woo, J. *et al.* Overexpression of AQP5, a putative oncogene, promotes cell growth
366   and transformation. *Cancer Lett.* (2008). doi:10.1016/j.canlet.2008.01.029

367   33.   Chae, Y. K. *et al.* Expression of aquaporin 5 (AQP5) promotes tumor invasion in
368   human non small cell lung cancer. *PLoS One* (2008).
369   doi:10.1371/journal.pone.0002162

370   34.   Huang, Y.-H. *et al.* Aquaporin 5 promotes the proliferation and migration of human
371   gastric carcinoma cells. *Tumour Biol.* (2013). doi:10.1007/s13277-013-0712-4

372   35.   Shimasaki, M., Kanazawa, Y., Sato, K., Tsuchiya, H. & Ueda, Y. Aquaporin-1 and -5
373   are involved in the invasion and proliferation of soft tissue sarcomas. *Pathol. - Res.*
374   *Pract.* **214**, 80–88 (2018).

375   36.   Wang, W. *et al.* Expression of AQP5 and AQP8 in human colorectal carcinoma and
376   their clinical significance. *World J. Surg. Oncol.* (2012). doi:10.1186/1477-7819-10-
377   242

378   37.   Shen, L. *et al.* Expression profile of multiple aquaporins in human gastric carcinoma
379   and its clinical significance. *Biomed. Pharmacother.* (2010).
380   doi:10.1016/j.biopha.2009.12.003

381

Fig. 1: Comparative profiling of Lgr5 populations in gastrointestinal tissues yields new pyloric specific markers

(a) Heatmap of transcriptomes of Lgr5-EGFP^{Hi}, Lgr5-EGFP^{Lo} and unsorted populations from mouse pylorus, small intestine and colon (n=4, 2, 3, biological replicates respectively) from Lgr5-EGFP-IRES-CreERT2 mice. Candidates are enriched only in pylorus-GFP^{Hi} population (black box).

(b) Overlap of candidate markers from Lgr5-EGFP-IRES-CreERT2 (above) and Lgr5-DTR-EGFP (n=4 biological replicates) models (statistical significance assessed by hypergeometric distribution).

(c) Shortlisted pylorus-specific markers.

(d) Relative *Aqp5* expression (qPCR) in gastrointestinal populations (n=2 technical replicates of pooled sample from 8 mice, means represented).

(e-j) *In situ* hybridization (ISH) expression of candidate markers in pylorus (n=3 mice).

(k-m) *Aqp5* expression with *Lgr5* in pylorus by immunostaining (k,m) and ISH (l) (n=3 mice).

Scale bars represent 25 μm.

Fig. 2: Pyloric Aqp5+ population contains stem cells *in vivo*

(a) Aqp5-EGFP-IRES-CreERT2 transgene.

(b-e) EGFP co-immunostaining in Aqp5-EGFP-IRES-CreERT2 pylorus with Gif (b), Gastrin (c), Mucin5ac (d) and Chromogranin A (ChgA, e) (n=3 mice).

(f-i) Lineage tracing of Aqp5-EGFP-IRES-CreERT2;tdTomato^{LSL} pylori at 20 hours (f), 5 days (g), 1 month (h) and 6 months (i) (n=3 mice).

(j-m) Co-immunostaining of tdTomato with ChgA (j), Gastrin (k), GIF (l) and Mucin5ac (m) (n=3 mice per marker).

(n) FACS gating for sorting Aqp5+ and Aqp5- cells (means±s.e.m.; n=6 biological replicates).

409   (o-p) Relative *Aqp5* (o) and *Lgr5* (p) expression (qPCR) in sorted populations (means± s.e.m.;

410   n=10 biological replicates).

411   (q) *Lgr5* and candidate marker enrichment in volcano plot of differentially expressed genes in

412   Aqp5+ population by microarray (n=4 biological replicates, p-value from one-way ANOVA in

413   Partek Analysis Software).

414   (r-s) Outgrowth efficiency of single antibody-mediated FACS-sorted Aqp5+ and Aqp5- cells (r)

415   with representative images (s) (n=5 biological replicates).

416   (t) Longevity and highest passage number of organoids from Aqp5+ and Aqp5- cells (n=3

417   biological replicates).

418   (u) Organoid differentiation protocol.

419   (v-x) Relative *Aqp5* (v), *Lgr5* (w) and *Muc5ac* (x) expression (qPCR) in Aqp5+ cell-derived

420   organoids on day 4 of differentiation (n=5 biological replicates).

421   Scale bars represent 25 µm in (b-m), and 100 µm in (s). Graphs represent means±s.e.m. with

422   two-sided t-test (r), (t), (v-x).

423

424   Fig. 3: Human pyloric AQP5+ cells are stem cells *ex vivo*

425   (a) Immunostaining of AQP5 in normal human pylorus (n=3 biological replicates).

426   (b-d) *AQP5* and *LGR5* expression in pylorus (b) near gland surface (c) and base (d) by ISH

427   (n=3 biological replicates).

428   (e-f) AQP5 and KI67 co-localization (arrowheads, n=3 biological replicates).

429   (g) FACS gating for sorting human AQP5+ and AQP5- cells (means±SEM; n=4 biological

430   replicates).

431   (h) Relative *AQP5* expression (qPCR) in FACS-sorted AQP5+ and AQP5- cells (n=3 biological

432   replicates).

433   (i-j) Outgrowth efficiency of single AQP5+ and AQP5- cells *in vitro* (i) with representative image

434   (j) (n=3 biological replicates).

435   (k-l) Relative *LGR5* (k) and *MUC5AC* (l) expression in (qPCR) AQP5+ cell-derived organoids

436   on day 4 of differentiation (n=3 biological replicates).

437　(m) Heatmap of 200 genes with highest variation between normal human pylorus AQP5+ and

438　AQP5- populations by RNASEQ (n=8 biological replicates).

439　(n) GO terms associated with genes significantly enriched in AQP5+ population.

440　Scale bars represent 100µm in (a),(b),(e),(j) and 50µm in (c),(d),(f). Graphs are presented as

441　means± s.e.m.; two-sided t-test.

442

443　Fig. 4: Aqp5+ population is a source of Wnt-driven, invasive gastric cancer

444　(a) Characteristics of gastric cancer models.

445　(b-h) Characterization of Aqp5-EGFP-IRES-CreERT2;APK intramucosal gastric

446　adenocarcinoma. H&E stain of entire pyloric region (b) and tumour (green box) (c).

447　Immunostaining of the tumour region in (b) for Aqp5-GFP (d), E-cadherin (e), phospho-AKT

448　(f), MAPK (g) and β-catenin (h) in the tumour region (b). (h) is a higher magnification of

449　Extended Data Fig. 12a.

450　(i-o) Characterization of Slc9a3-2A-CreERT2;AP gastric adenocarcinoma. H&E of entire

451　pyloric region (i) and the focal invasion (j). Immunostaining for Aqp5 (k), E-cadherin (l),

452　phospho-AKT (m), MAPK (n) and β-catenin (o) in the focal invasion.

453　(p) Co-staining of Aqp5-GPF, Ki67 and E-cadherin in Aqp5-CreERT2;APK pyloric tumour.

454　(s-u) Organoid assay with Aqp5-GFP+ GFP- tumour cells (n=3 biological replicates). (q)

455　Initiation frequencies of Aqp5-GFP+ tumour cells (APK-GFP+) versus Aqp5-GFP- tumour cells

456　(APK-GFP-). Paired two-sided t-test, means±s.e.m. (r) Means and individual longevities of

457　organoids derived from APK-GFP+, APK-GFP- and Norm-GFP+ (Aqp5-GFP+ cells from

458　normal pylorus) cells seeded with exogenous growth factors (GF) for a week and without GF

459　for the next 4 weeks. (s) Representative images of organoids derived from APK-GFP+ cells

460　and Norm-GFP+ cells in the respective GF conditions.

461　Scale bars represent 1 mm in (b),(i), 100 µm in (d-h), (j-o), 20 µm in (p), and 500 µm in (s).

462

Methods

Mice

For exon 1-knockins, an EGFP-IRES-CreERT2 cassette was inserted immediately downstream of the start codons of *Aqp5, A4gnt, Spp1* and *Slc9a3* gene loci by homologous recombination in embryonic stem cells as illustrated in Extended Data Fig. 3a. For 3' UTR-knockins, a 2A-CreERT2, 2A-EGFP or 2A-DTR cassette was inserted immediately before the stop codon of *Aqp5, Lgr5* and/or *Slc9a3* gene loci by homologous recombination in embryonic stem cells as illustrated in Extended Data Fig. 3b, thereby preserving the intact protein-coding region and endogenous expression of the genes. Rosa26-tdTomato[LSL] (Ai14) (JAX ID 007914) mice[38] were obtained from Jackson Labs. Lgr5-EGFP-IRES-CreERT2[4] (JAX ID 008875), Lgr5-DTR-EGFP (MGI ID 5294798)[5], Kras[LSL-G12D] (JAX ID 019104)[39], APC[fl/fl] (MGI ID 1857966)[40] and PTEN[fl/fl] (MGI ID 2182005)[41] mice have been described previously. All Cre and EGFP lines are bred as heterozygotes except A4gnt-IRES-CreERT2 and Spp1-IRES-CreERT2 mice, which are bred as homozygotes. All animal experiments were approved by the "Institutional Animal Care and Use Committee" of Singapore and performed in compliance with all relevant ethical regulations. Maximum tumour size allowed by IACUC is 20mm in any dimension and none of the experiments exceeded this limit. For all experiments, adult animals (not selected for gender) with a minimum age of 7-8 weeks were used. The experiments were not randomized, and there was no blinded allocation during experiments and outcome assessment. No statistical method was used to pre-determine sample size. Genotyping primers are collated in Supplementary Table 4. Mouse lines are available upon request.

Human material

Normal human pylorus for FACS was provided by K.G. Yeoh, J. So and A. Shabbir, NUS Department of Medicine and Pathology (granted under protocol 11-167E) and N. Inaki and T. Tsuji, Ishikawa Prefectural Central Hospital. Informed consent was obtained from all patients and experiments were performed in compliance with all relevant ethical regulations. Human distal cancer FFPE sections were provided by NUS Department of Medicine and Pathology

490 (granted under protocol-11-167E) and Leeds Teaching Hospitals NHS Trust, Leeds, UK

491 (granted under protocol CA01_122).

492 **Animal treatment**

493 Mice were each injected with tamoxifen dissolved in sunflower oil intraperitoneally at 4 mg

494 tamoxifen/30g body weight. Diphtheria Toxin (DT)-treated mice were injected with a single

495 dose of DT dissolved in PBS intraperitoneally at 0.5 ug DT/30g body weight.

496 **Gland isolation, cell dissociation and flow cytometry**

497 *Murine pylorus*

498 Murine pylorus was incubated in chelation buffer (5.6 mM sodium phosphate, 8 mM potassium

499 phosphate, 96.2 mM sodium chloride, 1.6 mM potassium chloride, 43.4 mM sucrose, 54.9 mM

500 D-sorbitol, 1 mM dithiothreitol) with 5 mM EDTA at 4∘C for 2 hours. Glands were isolated by

501 repeated pipetting of finely chopped pylorus tissue in cold chelation buffer. Chelation buffer

502 containing isolated glands was filtered through 100 µm filter mesh, and centrifuged at 720*g* at

503 4∘C for 3 min. The pellet was resuspended in TrypLE (Life Technologies) with DNaseI

504 (0.8U/µL) (Sigma) and incubated at 37∘C for 10 min with intermittent trituration for digestion

505 into single cells. Digestion was quenched by dilution with cold HBSS buffer. The suspension

506 was centrifuged at 720*g* at 4∘C for 3 min. For anti-Aqp5 antibody stain, the pellet was

507 resuspended in HBSS with 2% fetal bovine serum (FBS, Hyclone) with anti-Aqp5-AF647

508 (Abcam, ab215225) at 1:500 dilution and incubated on ice at 30min in the dark. The pellet was

509 subsequently washed twice with cold HBSS and spun at 800g for 3min at 4∘C. The pellet was

510 resuspended in HBSS with 2% FBS. Before sorting, 1 µg/ml propidium iodide (Life

511 Technologies) was added to the cell suspensions, filtered through a 40µm strainer, and sorted

512 on BD Influx Cell Sorter (BD Biosciences). Cells were collected in RLT Plus buffer (Qiagen)

513 for RNA extraction or HBSS with 2% FBS and 1% PenStrep (Gibco) for organoid culture.

514

515

516   *Human pylorus*

517   Human pylorus was collected in Advanced DMEM/F-12 media with 10 mM HEPES, 2 mM

518   Glutamax (incubation buffer, all from Life Technologies), supplemented with 1X Anti-Anti (Life

519   Technologies) and 1mM N-acetylcysteine (Sigma). After at least three washes in HBSS, the

520   pylorus was finely chopped and digested in incubation buffer supplemented with 1mg/mL

521   Collagenase (Gibco) and 2mg/mL Bovine Serum Albumin (Sigma) for 30 min at 37◦C with

522   intermittent mixing. The remainder of the processing protocol is identical to that for the murine

523   tissue described earlier. Cells for organoid culture were collected in organoid culture medium

524   with growth factors and 0.2% Growth factor-reduced Matrigel (Corning) (v/v).

525   **Organoid culture**

526   Organoid culture of FACS-isolated single human and murine pylorus cells were performed as

527   described previously[42]. Briefly, single cells were resuspended in growth factor-reduced

528   Matrigel (Corning) and cultured in basal media [Advanced DMEM/F-12 media with 10mM

529   HEPES, 2mM Glutamax, 1X N2, 1X B27 (all Invitrogen), N-acetyl-cysteine (Sigma), Primocin

530   (Invivogen)] supplemented with these growth factors - EGF (Invitrogen), Gastrin (Sigma),

531   FGF10 (Peprotech), Noggin (Peprotech), Wnt3a (Millipore), R-spondin and ROCK inhibitor

532   Y27632 (Sigma). A83-01 (Tocris) is also added to human pyloric cultures. Mouse cancer

533   organoids were grown in only basal media after first week of culture. Organoids were

534   passaged when confluent, at least once a week. Only organoids beyond 100 µm and 200 µm

535   in diameter with a clear central lumen are scored as organoids for Fig. 4 and Fig. 5

536   respectively.

537   **RNA isolation and qPCR**

538   Tissues were lysed in Trizol (Qiagen) and single cells were lysed in RLT Plus buffer (Qiagen).

539   RNA was subsequently isolated with RNeasy Universal Plus kit (Qiagen) and cDNA was

540   generated with Superscript III (Life Technologies) according to manufacturer's instructions.

541   qPCR was performed with a minimum of three biological replicates per gene using SYBR

542   Green dye (Promega) according to the manufacturer's instructions and ran on StepOne or

543   Quantstudio7 qPCR machines (Applied Biosystems). Analysis was carried out using double

544   CT method on Step One Software on the respective qPCR machines (Applied Biosystems).

545   qPCR validation of top candidates from RNASEQ was performed on 2-4 ng of SPIA-amplified

546   cDNA derived from the Ovation Pico WTA system (Nugen Technologies) due to limitation of

547   RNA availability. In the event any of the samples for a specific target does not amplify, the

548   relative expression values of all the samples for that target are increased by one to allow

549   visualization of the values on a log scale. Sequences of qPCR primers are collated in

550   Supplementary Table 4.

551   **Transcriptome profiling and analysis**

552   *Single cell RNAseq CEL-Seq and RaceID*

553   Single Lgr5-EGFP^{Hi} pyloric epithelial cells from Lgr5-DTR-EGFP mice[5] were isolated by FACS

554   (as described earlier) and collected in each well of 96-well plates. Total RNA extracted from

555   each cell was used to generate single-cell RNA expression libraries as described[48]. A total of

556   285 Lgr5-EGFP^{Hi} cells from three mice were sequenced on Illumina HighSeq 2500 instrument

557   using 101 base-pair paired end sequencing. *K*-means clustering in RaceID was used to

558   delineate clusters of subpopulations as previously described[20].

559   *Microarray and analysis*

560   Labelling, hybridization and washing protocols for microarrays were performed according to

561   Origene instructions. RNA quality was first determined by assessing the integrity of the 28s

562   and 18s ribosomal RNA bands on Agilent RNA 60000 Pico LabChips in an Agilent 2100

563   Bioanalyser (Agilent Technologies). A minimum of 2ng of RNA was used to generate SPIA-

564   amplified cDNA using the Ovation Pico WTA system (Nugen Technologies). Five micrograms

565   of SPIA-amplified purified cDNA was then fragmented and biotin-labelled using the Nugen

566   Encore Biotin module (Nugen Technologies). Microarray was performed using the Affymetrix

567   Mouse ST v2.0 GeneChips (Affymetrix), which consists of more than 28,000 probes for

568   previously annotated genes. The individual microarrays were washed and stained in an

569   Affymetrix Fluidics Station 450, and hybridized probe fluorescence was detected using the

570   Affymetrix G3000 GeneArray Scanner. Image analysis was carried out on the Affymetrix

571   GeneChip Command Console v2.0 using the MAS5 algorithm. CEL files were generated for

572   each array and used for gene expression analysis. The CEL files were then processed in R

573   (v3.2.3) with the Bioconductor (v3.2) libraries 'oligo' (v1.34.2), 'pd.mogene.2.0.st' (v3.14.1)

574   and 'limma' (v3.26.8). We used Robust Multi-array Average (RMA) to perform background

575   correction and normalization with the 'rma' function implemented in the 'oligo' package ('target'

576   parameter was set to 'core' to obtain expression values at the gene level). The experimental

577   design was stored as a single factor with individual levels for each combination of Lgr5-GFP

578   level (high, low, negative) or Aqp5 status (positive, negative). Linear models were fitted to the

579   expression data with the function 'lmFit' (default parameters). The relevant contrasts were

580   fitted with 'contrasts.fit' (default parameters); differential expression was tested with 'eBayes'

581   (default parameters). Differential gene expression was analyzed using Partek Genomics Suite

582   software (Partek). Relative gene expressions are depicted as single values as given by Partek

583   Analysis Software. Gene Set Enrichment Analysis was performed using the GSEA v6.1[46,47].

584   *RNASEQ*

585   AQP5+ and AQP5- cells were collected directly into RLT Plus buffer by FACS sorting. Total

586   RNA was isolated using Qiagen RNeasy Micro Kit (QIAGEN). RNA quality was first

587   determined by assessing the integrity of the 28s and 18s ribosomal RNA bands on Agilent

588   RNA 60000 Pico LabChips in an Agilent 2100 Bioanalyser (Agilent Technologies). Amplified

589   cDNA library was prepared according to manufacturer's instructions with SMARTer Stranded

590   Total RNA-Seq Kit v2 - Pico Input Mammalian (Takara) using 10ng of input Total RNA.

591   Indexed 150bp paired end sequencing was performed on HiSeq 2500 (Illumina) and Illumina

592   Real-Time Analysis (RTA) software was used for base-calling to generate Fastq files. The

593 reads were mapped to Genome Reference Consortium Human Build 38 patch release 12

594 (GRCh38.p12) with STAR software version 2.5.3a with the following options issued: --

595 outFilterType BySJout, --outFilterMultimapNmax 10, --alignSJoverhangMin 15, --

596 alignSJDBoverhangMin 1, --outFilterMismatchNmax 12, --outFilterMatchNminOverLread 0.4,

597 --alignIntronMin 20, --alignIntronMax 2000000, --outSAMattrIHstart 0, --outSAMmapqUnique

598 244, --outMultimapperOrder Random, --outReadsUnmapped None, --outFilterIntronMotifs

599 None, --outSAMmode Full, --outSAMattributes All --quantMode GeneCounts, --

600 clip3pAdapterSeq

601 AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT.

602 Counts per sample were subsequently concatenated in a statistical software, R version 3.2.3,

603 and reads were normalized with Trimmed mean of M values (TMM) normalisation as

604 implemented in edgeR version 3.12.1 (with limma_3.26.9). Differential expression testing was

605 performed with the edgeR function "glmQLFit" using a design matrix that took sample batches

606 and AQP5 status into account. Differentially expressed genes (DEGs) were those with more

607 than 2-fold change between AQP5+ and AQP5- samples, with FDR<0.05. Due to likely

608 inclusion of immune cells in the profile, immune-related genes[43] were omitted, resulting in a

609 final list of >500 DEGs. Gene Ontology, Overrepresentation Analysis (ORA) and PANTHER

610 Pathway analysis of the DEGs were performed on the PANTHER Classification System[44]

611 using default parameters.

612 *Transcriptomic pathway signature analysis of human gastric cancer*

613 Level 3 TCGA RNA-seq normalized matrix for 415 GC and 35 normal gastric samples and

614 their corresponding clinical information were downloaded from the Broad Institute TCGA

615 Genome Data Analysis Center (GDAC) Firehose (https://gdac.broadinstitute.org/). Gene

616 expression data of 200 GC and 100 matched normal gastric samples were generated using

617 Affymetrix Human Genome U133 Plus 2.0 Array (GSE15459) and processed as described

618 previously[16]. All normal samples and only tumours of antral or pyloric origin were included for

619 analysis. To determine activity of PI3K, Wnt and KRAS pathways in primary tumours, we

620   utilized published pathway signatures by several groups: KRAS signature based on differential

621   gene expression analysis between high KRAS mutation and low KRAS mutation/wildtype CRC

622   tumours[18]; PI3K signature composed of genes modulated *in vitro* by PI3K inhibitors, according

623   to the CMap signature[17] and lastly, intestinal Wnt signature defined by profiling CRC cell lines

624   carrying an inducible block of the Wnt pathway and differential gene expression analysis of

625   human colon adenoma and adenocarcinomas versus normal colonic epithelium[19]. For each

626   pathway signature, only up-regulated genes in pathway activation were selected for

627   downstream analysis.

628   To quantify the relative activation level in a specific oncogenic pathway, we derived a "μ score"

629   for each sample profile. In brief, the transcriptomic "μ score" was defined as the average of

630   standardized expression values of those up-regulated genes in a specific oncogenic pathway

631   (after the log-transformed values centred to the standard deviation from the median across

632   the samples included in the analysis). For each oncogenic pathway, "μ scores" were

633   calculated for all normal samples, and the "μ score" at the 90% percentile of the normal

634   samples was used as the cut-off to define a pathway as "hyperactivated". The "μ score" for

635   each tumour was determined and a "μ score" higher than the cut-off was considered to be

636   "hyperactivated" for that particular pathway. For each combination of pathways (Wnt,

637   Wnt/PI3K or Wnt/PI3K/Kras), the concurrence rate was given by the frequency of tumours that

638   were "hyperactivated" in all the pathways in question.

639   **Histology**

640   *Immunohistochemistry (IHC) and immunofluorescence (IF).*

641   IHC and IF were performed according to standard protocols. In summary, tissues were fixed

642   in 4% paraformaldehyde/PBS (w/v) overnight at 4◦C, and processed into paraffin blocks. 8 μm

643   sections from the paraffin blocks and tissue microarray slides were deparaffinated and

644   rehydrated, followed by antigen retrieval via heating to 121◦C in a pressure cooker in standard

645   10mM citric acid pH6 buffer, a commercial citrate pH 6.1 buffer (S1699, DAKO) or Tris/EDTA

646 buffer, pH 9.0 (S2367, DAKO). Primary antibodies used were chicken anti-EGFP (1:2,000,

647 Abcam, ab290), rabbit anti-EGFP (1:200; Cell Signalling, 2956S), rabbit anti-Ki67 (1:200;

648 Thermofisher, MA5-14520), rabbit anti-GIF (1:10,000; provided by D. H. Alpers, Washington

649 University School of Medicine, USA), rabbit anti-RFP (1:200; Rockland, 600-401-379), rabbit

650 anti-aquaporin5 (1:200, Santa Cruz, SC-28628 and 1:500, LSBio, LS-C756566), rabbit anti-

651 Slc9a3 (1:200; Santa Cruz, SC-16103-R), rabbit anti-mucin6 (1:200; LsBio, LS-C312108),

652 rabbit anti-A4gnt (1:500, Novus Biologicals, NBP1-89129), rabbit anti-Gastrin (1:200, Leica

653 Biosystems, NCL-GASp), mouse anti-MUC5AC (1:200; Leica Biosystems, NCL-HGM-45-M1),

654 rabbit anti-vimentin (1:500; Abcam, ab92547), mouse anti-E-cadherin (1:200; BD

655 Transduction Laboratories, 610181), mouse anti-β-catenin (1:200; BD Transduction

656 Laboratories, 610154), mouse anti-RFP (1:200; Abcam, 129244), mouse anti-ChgA (1:200;

657 Abcam, 15160), rabbit anti-Phospho-MAPK (1:200; Cell Signalling, 4370S), mouse anti-H-K-

658 ATPase (1:1,000; MBL International Corporation, D032-3), rabbit anti-Phospho-Akt (1:200;

659 Cell Signalling, 3787L). Detailed information about clone number, and antibody validation is

660 found in Supplementary Table 6. The peroxidase-conjugated secondary antibodies used were

661 mouse/rabbit EnVision+ (DAKO) for HRP IHC or anti-chicken/rabbit/mouse Alexa

662 488/568/647 IgG (1:500; Invitrogen) for IF. GSII-Lectin-AF568 (1:500, Thermofisher) was

663 incubated on the slides for 1 hour at RT together with secondary antibodies. IHC sections

664 were dehydrated, cleared, and mounted with DPX (Sigma) while IF sections were mounted in

665 Hydromount (National Diagnostics) with Hoechst for nuclear staining. Immunostainings and

666 imaging were performed on a minimum of three biological replicates and representative

667 images of the replicates were included in the manuscript.

668 *Haematoxylin and eosin (H&E).*

669 H&E staining was performed on FFPE sections according to standard laboratory protocols.

670 *Whole mount analysis and vibratome sectioning*

671   Tissues were fixed in 4% paraformaldehyde/PBS (w/v) overnight at 4◦C. Whole mount tissues

672   were permeabilized in 2% TritonX-100/PBS (v/v) overnight at 4◦C, while 500 µm vibratome

673   sections were generated by sectioning tissues embedded in 4% low-melting point agarose

674   with a vibrating microtome (Leica) and permeabilized in 2% TritonX-100/PBS (v/v) overnight

675   at 4◦C. Rapiclear (Sunjin lab) was used to clear whole mount tissues and vibratome sections

676   according to manufacturer's instructions. Hoescht was used as nuclear counterstain.

677   *In situ hybridization (ISH)*

678   ISH and co-ISH was performed using RNAscope[45] 2.5 High Definition Brown Assay and 2.5

679   High Definition Duplex Reagent Assay (Advanced Cell Diagnostics) respectively, according to

680   manufacturer's instructions. DapB was used as negative control for all the RNAscope

681   experiments. *In situ* hybridization and imaging were performed on a minimum of three

682   biological replicates and representative images of the replicates were included in the

683   manuscript.

684   *Analysis and scoring of stainings on mouse and human FFPE sections*

685   Overlap of Slc9a3-GFP and Aqp5-antibody stains was determined by counting stained cells

686   contacting the lumen at the base of glands. The entire height of the gland base surrounding

687   the gland base lumen had to be visible to avoid over-or underrepresentation of localized

688   populations.

689   Samples of H&E sections of mouse gastric tumours were evaluated by qualified veterinary

690   and clinical histopathologists. Scoring of AQP5 staining on human gastric cancer specimens

691   was performed by qualified clinical histopathologists. The tumour in the tissue section was

692   considered positive for AQP5 if staining was observed in more than 5% of the malignant cells.

693   Subcellular localization of AQP5, relative staining intensities and stained features were all

694   determined by qualified pathologists.

695   **Microscopy imaging**

696   *Image acquisition*

697   IHC and H&E slides were imaged with Zeiss AxioImager Z1 Upright microscope. RNAscope

698   slides and large area images were captured with Nikon Ni-E microscope/DS-Ri2 camera. IF

699   slides were imaged using Olympus FV1000 and FV3000 confocal microscopes. Cultured

700   organoids were imaged with Olympus DP-27 camera on Olympus IX53 inverted microscope.

701   *Image processing*

702   RNAscope and images with large areas were processed with NIS-Elements AR software

703   (Nikon) with EDF and stitching features respectively. IF images were processed using ImageJ

704   (NIH) and whole mount organoid images were processed using Imaris 8.0 (Bitplane).

705   **Statistics and reproducibility**

706   Gene expression data were quantified and depicted as the mean±SEM. Statistical analyses

707   were performed using GraphPad Prism. Data were tested for statistical significance by paired

708   two-tailed *t*-test unless otherwise stated in figure legends. Statistical significance of overlap

709   between the two gene sets in Fig. 1b was determined by hypergeometric distribution

710   (http://nemates.org/MA/progs/overlap_stats.html*). Precise *P* values of statistical significance

711   are shown in the respective figures. Representative images of all histological experiments and

712   FACS sorting strategies were performed at least thrice independently with similar results,

713

714   **Supplementary References**

715   38.   Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization
716   system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).

717   39.   Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional
718   expression of oncogenic K-ras. *Genes Dev.* (2001). doi:10.1101/gad.943001

719   40.   Shibata, H. *et al.* Rapid colorectal adenoma formation initiated by conditional targeting
720   of the APC gene. *Science (80-.).* (1997). doi:10.1126/science.278.5335.120

721 41. Suzuki, A. *et al.* T cell-specific loss of Pten leads to defects in central and peripheral
722   tolerance. *Immunity* (2001). doi:10.1016/S1074-7613(01)00134-0

723 42. Leushacke, M. *et al.* Lgr5-expressing chief cells drive epithelial regeneration and
724   cancer in the oxyntic stomach. *Nat Cell Biol* **19**, 774–786 (2017).

725 43. Monaco, G. *et al.* RNA-Seq Signatures Normalized by mRNA Abundance Allow
726   Absolute Deconvolution of Human Immune Cell Types. *Cell Rep.* (2019).
727   doi:10.1016/j.celrep.2019.01.041

728 44. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene
729   function analysis with the panther classification system. *Nat. Protoc.* (2013).
730   doi:10.1038/nprot.2013.092

731 45. Wang, F. *et al.* RNAscope: A novel in situ RNA analysis platform for formalin-fixed,
732   paraffin-embedded tissues. *J. Mol. Diagnostics* (2012).
733   doi:10.1016/j.jmoldx.2011.08.00

734 46. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach
735   for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A.
736   (2005). doi:10.1073/pnas.0506580102

737 47. Mootha, V. K. et al. PGC-1α-responsive genes involved in oxidative phosphorylation
738   are coordinately downregulated in human diabetes. Nat. Genet. (2003).
739   doi:10.1038/ng1180

740

Author Information

All correspondence and requests for materials should be addressed to Nick Barker (Nicholas.barker@imb.a-star.edu.sg).

Ethics Declaration

The authors declare the following competing interest:

N.B. and S.H.T. are co-inventors of the provisional patent application 10201911742W titled "A method for functional classification and diagnosis of cancers". This patent covers the analysis of human cancers using their signalling pathway statuses.

768   All the other authors declare no competing interest.

769

770   **Author contributions**

771   S.H.T and Y.S. contributed to all aspects of the study – they designed, performed all empirical

772   experiments, collected, analyzed data and wrote the manuscript. Y.S. designed, performed

773   experiments, collected and analyzed data for profiling and validation studies for candidate

774   marker identification. J.G. performed immunostaining, CEL-Seq experiments, mouse

775   husbandry. R.S. provided advice and technical help with human and mouse cancer, analysed

776   data and wrote the manuscript. K.M. performed FACS and immunostaining for human AQP5-

777   FACS experiments. P.P. performed immunostaining and mouse husbandry. L.T.T. performed

778   mouse husbandry. E.W. generated the trangsenic mouse lines. T.S., S.W.T.H., analysed

779   human cancer data in pathway analysis. S.L.I.J.D. analysed microarray, CEL-Seq and

780   RNASEQ data. S.M. performed FACS experiments. A.F. provided advice and technical help

781   with human experiments and mouse cancer models. M.O. T.T., H.G., S.S., T.M., K.G.Y., J.S.

782   and A.S. provided patient samples. H.G, S.S. and T.M. analysed and scored stained patient

783   samples. P.T. designed and supervised cancer frequency analysis. N.B. supervised the

784   project, analyzed the data and wrote the manuscript. All authors discussed results and edited

785   the manuscript.

786

787   **Data and code availability**

788   Microarray data that support the findings of this study have been deposited in the Gene

789   Expression Omnibus (GEO) under accession code GSE121803. RNASEQ data of AQP5+

790   and AQP5- human samples have also been deposited in the Gene Expression Omnibus

791   (GEO) under accession code GSE133036. Source data for Figs. 1-4 and Extended Data Figs.

792   1, 2, 4-6, 9 are provided with the paper. Any other relevant data supporting the findings of this

793   study are available from the corresponding author on reasonable request.

794   31

Extended Data Figure Legends

796 Extended Data Fig. 1

797 (a-c) FACS sorting strategy sorting GFP^hi and GFP^lo cells from Lgr5-EGFP-IRES-CreERT2

798 pylorus (a), small intestine (b) and colon (c).

799 (d) *Lgr5* expression (qPCR) in sorted populations and unsorted tissues of the gastrointestinal

800 tract data represented as means± s.e.m. (n=4 biological replicates); one-way ANOVA test.

801 (e) Aqp5 protein expression in the mouse stomach through to the duodenum by

802 immunostaining (n=3 biological replicates).

803 (f-i) A*qp5* mRNA expression in the corpus (f), Brunner's glands (g), small intestine (h) and

804 colon (i) by ISH (n=3 biological replicates).

805 (j-n) *A4gnt* (j), *GIF* (k), *Muc6* (l), *Slc9a3* (m) and *Spp1* (n) expression in the corpus, Brunner's

806 glands, small intestine and colon by qPCR, ISH, and co-ISH with Lgr5. For histology

807 experiments, n=3; For qPCR, n=4 biological replicates for GIF, Muc6 and Slc9a3 qPCRs

808 where data is represented as means± s.e.m., n=2 technical replicates from a pooled sample

809 of 8 for A4gnt and Spp1 qPCRs.

810 Scale bars represent 500 μm in (e), and 20 μm in (f-n).

811

812 Extended Data Fig. 2

813 (a) Exon1-knockin gene strategy to generate EGFP-CreERT2 reporters of *Aqp5*, *A4gnt*,

814 *Slc9a3* and *Spp1* expression.

815 (b) 3'UTR-knockin gene strategy to generate 2A-EGFP, 2A-CreERT2 or 2A-DTR reporters of

816 *Aqp5*, *Slc9a3* and *Lgr5* expression.

817 (c-f) GFP signal in the pylorus and small intestine of Aqp5-2A-EGFP (c-d) and Slc9a3-2A-

818 EGFP mice (e-f).

819 (g-h) Quantification of the overlap between Slc9a3-2A-EGFP+ cells and Aqp5+ cells (g)

820 (n=102 glands from three mice) and a representative image of the immunostaining (h). Results

821 are presented as means±s.e.m.

822   (i-j) Quantification of overlap between Lgr5-EGFP+ cells with Aqp5+ cells (i) (n=117 glands

823   from four mice) and a representative image of the immunostaining (j). Results are presented

824   as means±s.e.m.

825   (k-m) Co-localization of Lgr5-EGFP in the pylorus with GIF (k), Gastrin (l) and Chromogranin

826   A (m) (n=3 biological replicates).

827   (n) t-SNE map of single Lgr5^{Hi} cells from the pylorus (n=285 cells from three mice).

828   (o-s) t-SNE maps showing enrichment of candidate markers in major (o-q) and minor (r-s)

829   subpopulations of Lgr5^{Hi} pyloric cells (n=285 cells from 3 mice).

830   (t) Frequency of 10 published proliferation markers (*Bcl2, Ccnd1, Ckap2, Foxm1, Ki67, Mcm2,*

831   *Mybl2, Plk1, Rrm2, Top2a*) in major versus minor subpopulations, compared by two-tailed

832   Mann-Whitney test (n=285 cells from 3 mice, 248 cells in major, 8 cells in minor 1 and 1 cells

833   in minor 2 populations).

834   (u-v) *Aqp5* (u) and *Lgr5* (v) expression in cells sorted from Aqp5-EGFP-IRES-CreERT2 pylori

835   (means±s.e.m.; n= 4 biological replicates).

836   (w-x) Co-immunostaining for EGFP driven by Aqp5-EGFP-IRES-CreERT2 and endogenous

837   Aqp5 (w) and Ki67 (x) (n=3 mice).

838   Scale bars represent 50 µm in (g),(h) and 25 µm in (k),(m),(w),(x).

839

840   <u>Extended Data Fig. 3</u>

841   (a-p) Lineage tracing in A4gnt-, Aqp5-, Spp1- and Slc9a3-EGFP-IRES-CreERT2 mice crossed

842   with tdTomato^{LSL} reporter mice after short trace (20-48 hours) in the pylorus (a-d) and small

843   intestine (i-l) and long trace (>3 months) in the pylorus (e-h) and small intestine (m-p) (n=3 per

844   genotype).

845   (q-x) Lineage tracing in pylorus (q-t) and small intestine (u-*x*) of the Aqp5-2A-CreERT2 and

846   Slc9a3-2A-CreERT2 mice after short trace (q-r, u-v) and long trace (s-t, w-x).

847   (y-b') Whole-mount imaging of pylorus from induced Aqp5-EGFP-IRES-CreERT2;tdTomato^{LSL}

848   mice 20 hours (y) and 6 months (z) after induction. Whole-mount imaging of pylorus from

849   uninduced 8 week-old (a') and 8 month-old (b') Aqp5-EGFP-IRES-CreERT2;tdTomato^{LSL} mice

850   (n=3 per condition). dTom signal through the entire height of the pyloric epithelium is shown

851   and DAPI from the upper parts of pyloric glands is depicted for clarity.

852   (c'-e') dTom signal in clusters of glands in 1 year-traced pylorus (c') and small intestine (d'-e').

853   (f'-h') dTom expression in gastric corpus (f'), colon (g') and brunner's glands (h') 20 h and 6

854   months after induction.

855   N=3 biological replicates. Scale bars represent 50 μm.

856

857   <u>Extended Data Fig. 4</u>

858   (a) Aqp5+ gating strategy with cells from WT pylorus stained only with propidium iodide.

859   (b) Heatmap of transcriptomes from Aqp5+ and Aqp5- cells (n=4 biological replicates).

860   (c) GSEA analysis comparing degree of overlap between transcriptomes of Aqp5+ cells and

861   Lgr5+ cells from the pylorus using Kolmogorov-Smirnov statistic (n=4 biological replicates

862   each).

863   (d-e) Relative *Aqp5* (d) and *Lgr5* (e) expression (from microarray) in Aqp5+ and Aqp5- cells

864   (n=4 biological replicates) by one-way ANOVA in Partek suite.

865   (f-h) Relative expression of various pyloric markers (f), other published pyloric stem cell

866   markers (g) and lineage and proliferation markers (h) in Aqp5+ population versus Aqp5-

867   population in microarray (n=4 biological replicates). Data represented as means, as derived

868   from Partek Analysis Software by one-way ANOVA.

869   (i-j) Aqp5 staining in a whole-mount organoid (i) and Aqp5 co-localization with Ki67 (j) in an

870   organoid section. Organoids were derived from single Aqp5+ cells (n=3 biological replicates).

871   (k-r) Pylori of DT-treated WT (k-n) and Aqp5-2A-DTR (o-r) mice stained for H&E (k, o), E-

872   Cadherin (l, p), Gif (m, q) and Gastrin (n, r) (n=3 biological replicates).

873   (s-u) Outgrowth efficiency of Aqp5+, Aqp5-, Lgr5-EGFP+, Lgr5-EGFP- cells (s) (n=5 biological

874   replicates for Aqp5+/- cells, n=3 biological replicates for Lgr5-EGFP+/- cells). Representative

875   images of organoids derived from Lgr5-EGFP+ (t) and Lgr5-EGFP- (u) cells. Paired two-sided

876   t-test.

877   Scale bars represent 25 μm in (i-j), 50 μm in (k) to (r), and 500 μm in (t) and (u).

878

Extended Data Fig. 5

(a-f') *MUC6* (a-b'), *A4GNT* (c-d') and *SLC9A3* (e-f') expression (co-ISH with *LGR5* and

immunostaining) in normal human pylorus (n=3 biological replicates).

(g-j) Co-ISH to co-localize *AQP5* with *PEPC* (g), *MUC6* (h), *GIF* (i) and *MUC5AC* (j) pyloric

lineage markers (n=3 biological replicates).

(k-l) AQP5 labelling in whole mount human organoids (k) and AQP5 colocalization with KI67

(l) in organoid sections (n=3 biological replicates).

(m-o) Relative *AXIN2* (m), *TFF2* (n) and *AQP5* (o) expression in AQP5+ cell-derived organoids

3 days after WNT3A, Noggin and FGF10 withdrawal by qPCR (n=3 biological replicates).

Scale bars represent 100 μm in all panels except (a'), (b'), (c'), (d'), (e'), (f'), (k) where they

represent 25 μm.

Extended Data Fig. 6

(a-e) qPCR validation (green, n=5 biological replicates) and RNASEQ values (blue, n=8

biological replicates) of murine stem cell markers (a), membrane components (b), chemokine

signaling components (c), extracellular matrix components (d) and other genes (e). Two-sided

Mann-Whitney test was used to determine statistical significance of qPCR results differences

for all genes except *AQP5*, which was determined by two-tail paired t-test. qPCR and

RNASEQ results are presented as means.

(f-f') ISH of SMOC2 (brown) on normal human pylorus. (f') is magnified inset of surface

mucosa, while (f") is magnified inset of gland base (n=4 biological replicates). Scale bars

represent 100 μm in (f), and 10 μm in (f') and (f").

(g) Top 10 Panther pathways enriched with most candidate genes.

Extended Data Fig. 7

(a-b) Co-hyperactivation status of the Wnt, PI3K and Kras pathways in human distal gastric

cancer samples from TCGA[3] (a) (n=155) and GSE15459[16] (b) (n=42) datasets. Heatmaps

906   show distribution of pathway hyperactivation statuses across samples. Graphs depict

907   distribution of μ-scores (degree of signalling activity) of normal and tumour samples for each

908   of the pathways examined.

909

910   <u>Extended Data Fig. 8</u>

911   (a) Sample sizes, tumour and invasion incidences observed in various permutations of

912   CreERT2 drivers and oncogenic alleles.

913   (b-g) Whole mount and H&E images of entire pyloric regions for each CreERT2-oncogenic

914   allele combination.

915   (h-l) H&E images of pylori from multiple pyloric marker-CreERT2;Kras$^{LSL-G12D}$ models (h-i),

916   APK-only model (without CreERT2 driver) (j), and small intestine (k) and colon (l) from Aqp5-

917   IRES-CreERT2;APK gastric cancer mouse model.

918   Scale bars represent 1 cm in whole mount insets in (b-g), 200 μm in H&E images in (b-l).

919

920   <u>Extended Data Fig. 9</u>

921   (a-g) Immunostaining of various markers in Aqp5-IRES-CreERT2;APK pyloric tumour.

922   (h-h") Co-ISH of Aqp5 and Lgr5 in tumour region, magnified in (h'). Dual ISH of Aqp5 and Lgr5

923   in an adjacent normal pyloric region from the same mouse (h").

924   (i-i') Representative H&E stain of a salivary gland tumour from Aqp5-IRES-CreERT2;APK

925   mouse.

926   (j-o) Immunostaining of various markers in Slc9a3-2A-CreERT2;AP pyloric tumour.

927   (p-w) Immunostaining of various markers in the APK-only (no CreERT2) control pyloric

928   stomach.

929   (x-x') Colocalization of Aqp5-EGFP and endogenous Aqp5 protein in pyloric tumour from an

930   Aqp5-IRES-CreERT2;APK mouse.

931   (y-z) Control FACS gating for GFP+ cells using normal Aqp5-EGFP-IRES-CreERT2 (y) and

932   WT (z) pylori (n=3 biological replicates).

933 (y-z") Organoid assay for stemness of Aqp5-GFP+ tumour cells (n=3 biological replicates). (y)

934 Experimental timeline. FACS gating strategy to isolate GFP+ tumour cells from Aqp5-

935 CreERT2;APK pyloric tumour (z), GFP+ cells from normal Aqp5-EGFP-IRES-CreERT2

936 pylorus (z'), and control GFP gating with WT pylorus (z").

937 Scale bars represent 100 µm in (a-g and i-x'), 100 µm in (h'), 20 µm in (h").

938 Extended Data Fig. 10

939 (a) Summary of AQP5 expression in a tumour microarray panel of 145 cores of human distal

940 gastric cancer. AQP5 expression is scored as positive if observed in > 5% malignant cells.

941 (b-e) Examples of AQP5+ cores with intestinal (b-c) and diffuse (d-e) subtypes, often with

942 cytoplasmic and/or membranous staining.

943 (f) Summary of AQP5 expression from 54 full sections of distal human gastric cancer.

944 (g-n) AQP5 expression in intestinal (g),(h),(k),(l) and diffuse (l),(j),(m),(n) subtypes. Yellow

945 arrowheads indicate cells co-expressing AQP5 and KI67.

946 (o) Summary of other observations of AQP5 expression in full gastric tumour sections.

947 (p-s) AQP5 expression in the invasive edge of the tumour (p), intestinal metaplasia (IM) (q,

948 dotted lines denote IM region negative for AQP5), Signet ring cells (r, black arrows denote

949 cells with weak AQP5 expression) and tumour cells in lymph node (LN) metastasis (s).

950 *Mixed refers to AQP5 localization in cytoplasm and nucleus, or cytoplasm and membrane.

951 Scale bars represent 20 µm in (g-n), (q-r), 50 µm in (b-e), (p), (s).

952 **Supplementary information** is available for this paper.

953 Peer review information.

954 Reprints and permissions information is available at www.nature.com/reprints.

Fig. 1

a

Log2 -4.64   0.00   4.64

- Pylorus GFP^hi (green)
- All SI arrays (orange)
- Pylorus GFP^lo (magenta)
- All colon arrays (blue)
- Non GFP^hi stomach arrays (yellow)

b

$p = $ 1E-119

Lgr5-EGFP-IRES-CreERT2 array: 22 | 67 | Lgr5-DTR-EGFP array: 113

A4gnt, Gif, Aqp5, Muc6, Slc9a3, Spp1

c

Gene	Fold change Pyl-GFP^Hi vs other GI groups (Log2)
Slc9a3	3.334
A4gnt	2.771
Gif	2.733
Aqp5	2.603
Spp1	2.402
Muc6	2.202

d

Relative Quantification of Aqp5 Expression

Pylorus Lgr5Hi, Pylorus Lgr5Lo, Corpus, Unsorted pylorus, SI Lgr5Hi, SI Lgr5Lo, Unsorted SI, Colon Lgr5Hi, Colon Lgr5Lo, Unsorted Colon

e A4gnt **f** Aqp5 **g** Gif

h Muc6 **i** Slc9a3 **j** Spp1

k Aqp5 / DAPI **l** Aqp5 / Lgr5 **m** Lgr5-GFP / Aqp5 / DAPI

Fig. 2

a *Aqp5* ATG EGFP-IRES-CreERT2

b Aqp5-GFP Gif

c Aqp5-GFP Gast

d Aqp5-GFP Muc5ac

e Aqp5-GFP ChgA

f Aqp5-GFP dTOM DAPI — 20hours

g 5 days

h 1 month

i 6 months

j Chga dTOM DAPI

k Gast dTOM DAPI

l GIF dTOM DAPI

m Muc5ac dTOM DAPI

n Aqp5- / Aqp5+ 3.8% (±0.3) — SSC / Aqp5-AF647

o Relative Aqp5 expression levels — $p=0.0005$ — Aqp5+ / Aqp5-

p Relative Lgr5 expression levels — $p=0.047$ — Aqp5+ / Aqp5-

q p-value / Fold change (log2) — Muc6 Slc9a3 Aqp5 A4gnt Gif Lgr5 Spp1

r Outgrowth efficiency (%) — $p=0.0030$ — Aqp5+ / Aqp5-

s

t Highest passage number — $p=0.0005$ — Max # weeks in culture

u Day 0 Plate organoids / 1 Withdraw WNT3A, NOGGIN, FGF10 / 4 Harvest organoids

v Relative Aqp5 expression levels — $p=0.047$ — Ctrl / Diff

w Relative Lgr5 expression levels — $p=0.0031$ — Ctrl / Diff

x Relative Muc5ac expression levels — $p=0.0028$ — Ctrl / Diff

Fig. 3

a AQP5

b AQP5 LGR5

c

d AQP5 LGR5

e DAPI AQP5 KI67

f

g AQP5- AQP5+ 3.7% (±0.9)
SSC
AQP5-AF647

h *p*=0.049
Relative AQP5 expression level
AQP5+ AQP5-

i *p*=0.14
Outgrowth efficiency (%)
AQP5+ AQP5-

j

k *p*=0.024
Relative LGR5 expression levels
Ctrl Diff

l *p*=0.0091
Relative MUC5AC expression levels
Ctrl Diff

m
-1 0 1 2 3
AQP5+ AQP5-

n
GO Cellular components
Membrane
Extracellular Space
Endomembrane system
Vesicle
Protein-containing complex
% of total hits (579 genes)

GO Molecular function
Protein binding
Ion binding
Hydrolase activity
Molecular transducer activity
Transporter activity
% of total hits (579 genes)

ORA Pathway
Notch signaling pathway *p*=0.050
Integrin signalling pathway *p*=0.0002
Alzheimer disease -presenilin pathway *p*=0.0024
CCKR signaling map *p*=0.0036
Fold enrichment

Fig. 4

Fig. 4

a	Total sample size	Tumour incidence	Invasion incidence	Average survival time
Exon1-IRES-CreERT2	23	19 (82.6%)	3 (15.8%)	6.7 months
3' UTR-2A-CreERT2	12	12 (100%)	8 (66.7%)	3.6 months

Aqp5-EGFP-IRES-CreERT2; APK (intramucosal) Slc9a3-2A-CreERT2; AP (early invasive)

Extended Data Fig. 1

Extended Data Fig. 2

a Targeting strategy for EGFP-IRES-CreERT2 construct

Target gene locus (*Aqp5, A4gnt, Spp1, Slc9a3*)

Targeting construct

Recombined allele

Final knockin allele

b Targeting strategy for 2A-CreERT2, 2A-EGFP and 2A-DTR constructs

Target gene locus (*Aqp5, Slc9a3, Lgr5*)

Targeting construct

Recombined allele

Final knockin allele

*Or 2A-EGFP or 2A-DTR

3' UTR-knockin (2A-EGFP)

Aqp5
- **c** Pylorus — EGFP / DAPI
- **d** Small intestine

Slc9a3
- **e** Pylorus
- **f** Small intestine

Colocalization

Aqp5 and Slc9a3
- **g** Slc9a3-EGFP/Aqp5 ab — 2.4%±0.3, 0.2%±0.1, 97.4%±0.3 — Slc9a3+Aqp5, Aqp5 only, Slc9a3 only
- **h** Slc9a3-2A-EGFP / Aqp5 / DAPI

Aqp5 and Lgr5
- **i** Lgr5-EGFP/Aqp5 ab — 5.1%±0.6, 0.7%±0.3, 94.1%±0.8 — Lgr5+Aqp5, Aqp5 only, Lgr5 only
- **j** Lgr5-EGFP / Aqp5 / DAPI

- **k** Lgr5-GFP Gif E-Cadherin DAPI
- **l** Lgr5-GFP Gastrin E-Cadherin DAPI
- **m** Lgr5-GFP Chromagranin A E-Cadherin DAPI
- **n** Minor subpopulation 1 *Gast Chga*; Minor subpopulation 2 *Krt8 Krt18*; Major subpopulation *Aqp5 Gif Muc6*

- **o** Aqp5
- **p** Gif
- **q** Muc6
- **r** Krt8

- **s** Gast
- **t** % of population expressing proliferation markers — p=0.021, p=0.0010 — Major, Minor 1, Minor 2
- **u** Relative Aqp5 expression levels — p=0.0045 — Aqp5 GFP+, Aqp5 GFP-
- **v** Relative Lgr5 expression levels — p=0.036 — Aqp5 GFP+, Aqp5 GFP-
- **w** Aqp5-GFP / Aqp5
- **x** Aqp5-GFP / Ki67

Extended Data Fig. 3

Extended Data Fig. 4

a

Aqp5- / Aqp5+
0.01%
(±0.005)
SSC
Aqp5-AF647

b

Aqp5+
Aqp5-
Log2 -2.23   0.00   2.23

c

Enrichment plot: LGR5
NES =2.60
FDR p-value = 0.000

d *p*=3.8E-5

Relative **Aqp5** expression
Aqp5+   Aqp5-

e *p*=0.0002

Relative **Lgr5** expression
Aqp5+   Aqp5-

f Relative expression of other Lgr5+ pyloric markers

Relative **A4gnt** expression — *p*=5.2E-5
Relative **Muc6** expression — *p*=4.0E-6
Relative **Spp1** expression — *p*=0.0004
Relative **Slc9a3** expression — *p*=5.2E-6
Aqp5+   Aqp5-

g Relative expression of other pyloric stem cell markers in Aqp5+ and Aqp5- cells

Relative **Lrig1** expression — *p*=0.008
Relative **Runx1** expression — *p*=0.0003
Relative **Axin2** expression — *p*=0.19
Relative **Cckbr** expression — *p*=0.72
Relative **Sox2** expression — *p*=0.35
Aqp5+   Aqp5-

h Relative expression of lineage markers

Relative **Gif** expression — *p*=0.0002
Relative **Muc5ac** expression — *p*=0.003
Relative **Gast** expression — *p*=0.14
Relative **Chga** expression — *p*=0.18
Relative **Ki67** expression — *p*=0.0098
Aqp5+   Aqp5-

i

Aqp5
Phalloidin
DAPI

j

Aqp5
Ki67
DAPI

	H&E	E-Cadherin	Gif	Gastrin
WT	**k**	**l**	**m**	**n**
Aqp5-2A-DTR	**o**	**p**	**q**	**r**

Comparison of organoid initiation efficiencies of Aqp5+ and Lgr5+ cells

s

p=0.47
p=0.0030
p=0.049
Outgrowth efficiency (%)
Aqp5+   Aqp5-   Lgr5+   Lgr5-

t

u

Extended Data Fig. 5

Extended Data Fig. 6

a Murine stem cell markers

AQP5 — *p*=0.016 / *p*=5.62E-14
A4GNT — *p*=0.37 / *p*=1.52E-03
MUC6 — *p*=0.095 / *p*=2.38E-10
SMOC2 — *p*=0.42 / *p*=1.12E-05

b Membrane markers

CCR1 — *p*=0.008 / *p*=1.70E-53
MARCH1 — *p*=0.008 / *p*=7.00E-41
MRC1 — *p*=0.008 / *p*=2.25E-62
SLC37A2 — *p*=0.008 / *p*=6.21E-53

c Chemokine signaling

CCL2 — *p*=0.056 / *p*=7.55E-31
CCL18 — *p*=0.10 / *p*=2.64E-32
CLEC4G — *p*=0.048 / *p*=9.10E-04
CXCL9 — *p*=0.008 / *p*=1.04E-28

d Extracellular components

PGA5 — *p*=0.056 / *p*=3.12E-49
MMP12 — *p*=0.008 / *p*=9.25E-45
MMP19 — *p*=0.016 / *p*=8.78E-16
TGFBI — *p*=0.008 / *p*=1.47E-66

e Other genes

C11orf96 — *p*=0.0079 / *p*=2.68E-32
C6orf58 — *p*=0.048 / *p*=9.10E-04

f SMOC2

f'

f''

g Panther Pathway analysis

% gene hits/total pathway hits

- Integrin signalling pathway
- Inflammation mediated by chemokine and cytokine signaling pathway
- CCKR signaling map
- Angiogenesis
- Wnt signaling pathway
- Gonadotropin-releasing hormone receptor pathway
- Alzheimer disease-presenilin pathway
- Cadherin signaling pathway
- Huntington disease
- FGF signaling pathway

Extended Data Fig. 7

a   **Cancer Genome Atlas Research Network (2014)**

Hyperactivated pathways	Frequency
Wnt	82.6%
Wnt; PI3K	61.3%
Wnt; PI3K; Kras	59.4%

b   **Muratani et al (2014) GSE 15459**

Hyperactivated pathways	Frequency
Wnt	81.0%
Wnt; PI3K	64.3%
Wnt; PI3K; Kras	57.1%

Extended Data Fig. 8

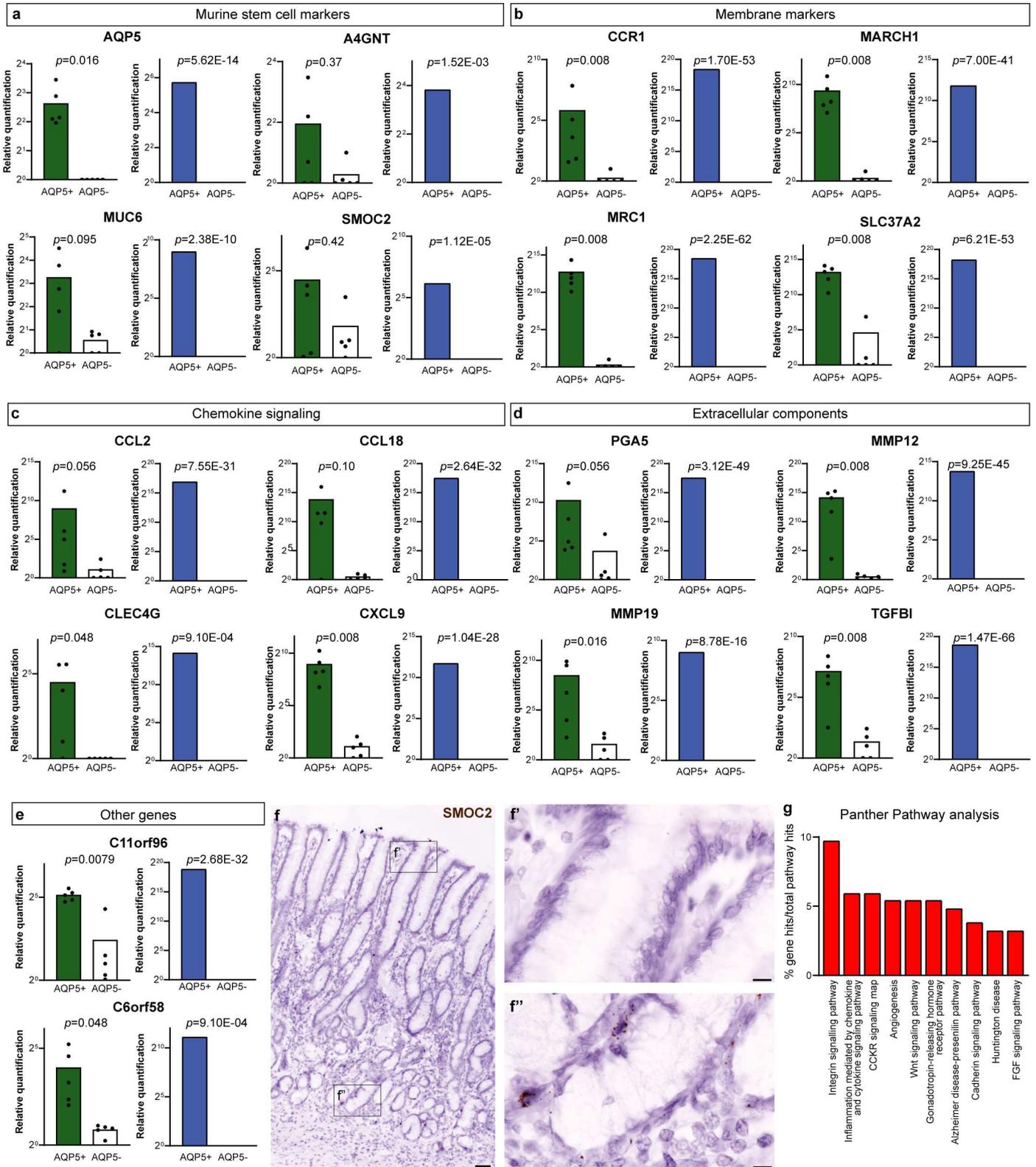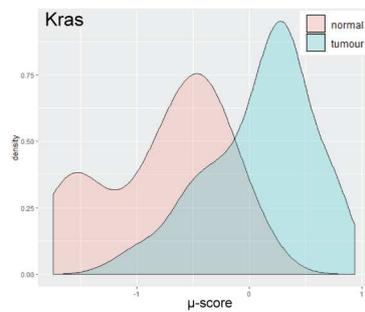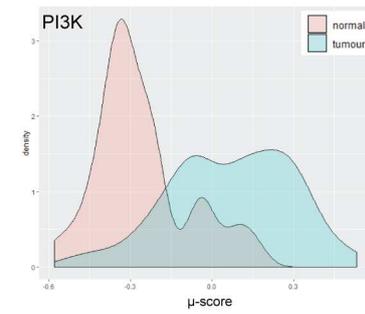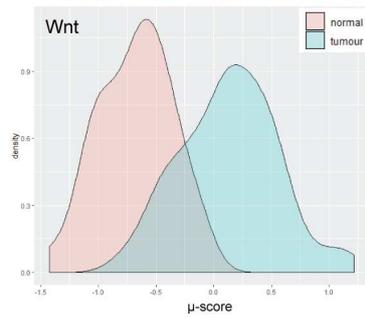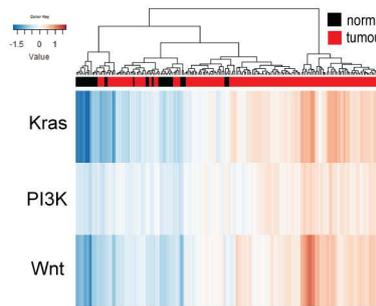|	APC (A)	APC;PTEN(AP)	APC;PTEN;Kras(APK)
IRES-CreERT2 (Aqp5, A4gnt)	n =4/5, 2 invasive 7 -11 mths	n = 5/7 4.5-6 mths	n=10/11, 1 invasive 2-4.5 mths
2A-CreERT2 (Aqp5, Slc9a3)	n=3/3, 3 invasive 3.5-8 mths	n=6/6, 5 invasive 2-3.5 mths	n=3/3 1-1.5 mths

IRES-CreERT2   **2A-CreERT2**

Controls
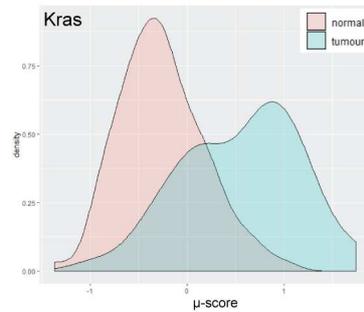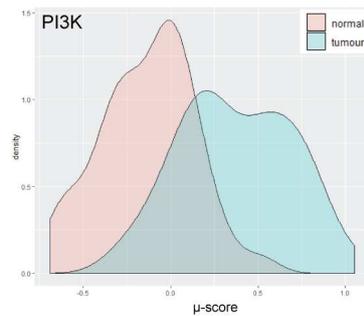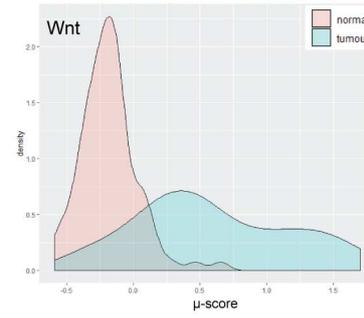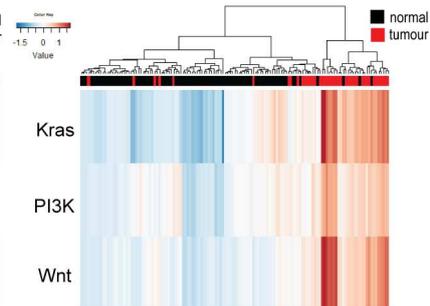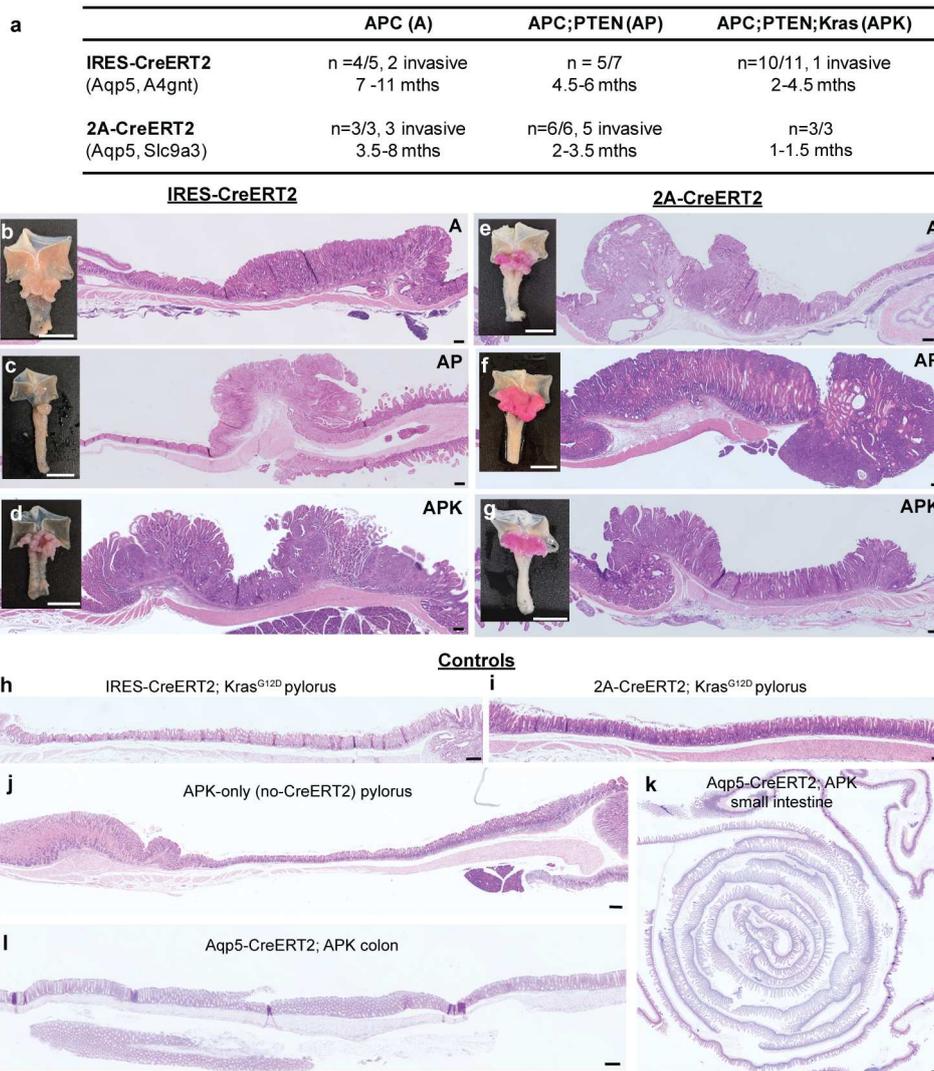
Extended Data Fig. 9

Extended Data Fig. 10

TMA	Total cores #	AQP5+ cores		Cellular localization (% of AQP5+ cores)			
		#	%	Cytoplasm	Membrane	Nucleus	Mixed*
Intestinal	77	56	72.7	96.4	42.9	1.8	39.3
Diffuse	44	31	70.5	93.5	29.0	6.5	32.3
Mixed	24	16	66.7	100.0	37.5	6.3	43.8
Overall	**145**	**103**	**71.0**	**96.1**	**37.9**	**3.9**	**37.9**

a (table above)

Intestinal	Diffuse

b AQP5 **c** **d** **e**

Full sections	Total samples #	AQP5+ samples		Cellular localization (% of AQP5+ samples)			
		#	%	Cytoplasm	Membrane	Nucleus	Mixed*
Intestinal	29	21	72.4	21	15	1	13
Diffuse	25	20	80.0	20	7	3	9
Overall	**54**	**41**	**75.9**	**100.0**	**53.7**	**9.8**	**63.7**

f (table above)

Intestinal		Diffuse	
Mucosa	Submucosa	Mucosa	Submucosa

g AQP5 **h** AQP5 **i** AQP5 **j** AQP5

k AQP5 CDH1 DAPI **l** AQP5 KI67 DAPI **m** AQP5 CDH1 DAPI **n** AQP5 KI67 DAPI

Full sections	AQP5 expression				
	Stronger in submucosa than mucosa	In poorly cohesive cells	In intestinal metaplasia	In signet ring cells	In lymph node metastases
Positive samples	21 out of 41	28 out of 40	12 out of 26	10 out of 15	10 out of 12
	51.2%	**70.0%**	**46.2%**	**66.7%**	**83.3%**

o (table above)

p AQP5 *Luminal side* *Invasive end* **q** IM *Normal glands* **r** *Signet ring cells* **s** *Lymph node metastasis*