Supplementary Materials

Diverse tumor susceptibility in Collaborative Cross mice: Identification of a new mouse model for human gastric tumorigenesis

Pin Wang[1,2], Yunshan Wang[2,3], Sasha A Langley[2], Yan-Xia Zhou[2,4], Kuang-Yu Jen[5], Qi Sun[6], Colin J Brislawn[7], Carolina M Rojas[8,9], Kimberly L Wahl[8,9], Ting Wang[6], Xiangshan Fan[6], Janet K Jansson[7], Susan E Celniker[2], Xiaoping Zou[1], David W Threadgill[8,9], Antoine M Snijders[2,10]*, Jian-Hua Mao[2,10]*

Supplementary Methods

Mouse husbandry

Mice were maintained on PicoLab Rodent Diet 20 (5053), housed in standard micro-isolator cages on corn cobb bedding with enrichment consisting of crinkle cut, naturalistic paper strands. Mice are negative for all of the following pathogens: MHV, Sendai, PVM, M. pulmonis, TMEV(GDVII), Reo-3, Parvo, EDIM, LCM and Ectromelia. Colon contents and pelage were inspected for pinworms and furmites respectively, and by PCR.

RNA library preparation and sequencing

At twelve weeks of age, stomach tissues were dissected from six CC mouse strains, three males and three females from each strain for a total of 36 samples. Stomach tissues were snap frozen in liquid nitrogen within 10 minutes of euthanasia. We isolated total RNA by homogenizing frozen tissue in Trizol reagent (Invitrogen) followed by chloroform phase separation. RNA was further purified using the RNeasy mini kit (Qiagen; 74104). DNA was removed using RNase-free DNase (Qiagen; 79254). For CC036 mice, RNA was prepared from individuals and samples were analyzed separately. For the other CC strains, RNA samples from male and female mice respectively, were pooled and analyzed jointly.

RNA was analyzed and quantified using the Bioanalyzer Chip (RNA Nano 6000) and a Qubit fluorometer. Sequencing libraries were prepared from 500 ng of total RNA using the NEBNext Poly(A) mRNA Magnetic Isolation Module (catalog number E7490, protocol revision 5.0), NEBNext Ultra Directional RNA Library Preparation Kit for Illumina (catalog number E7420, protocol revision 6.0), and NEBNext Multiplex Oligos for Illumina (catalog number E7600, protocol revision 2.0) following the manufacturer's instructions.

Individual libraries were normalized to 10nM and eight samples were pooled per lane. Sequencing was performed at UC Berkeley's QB3 Vincent J. Coates Genomics Sequencing Laboratory on an Illumina HiSeq4000 instrument, generating 150bp paired end reads. For each sample we produced between 9M and 59M total reads and 6M and 41M mapped reads. RNA-sequencing analysis was performed as described previously [1] [2]. RNA sequencing data has been submitted to the Sequence Read Archive under accession number SUB3887143.

Microbiome analyses

Genomic DNA was extracted from the homogenized fecal samples using the PowerSoil DNA Isolation Kit (http://www.mobio.com/) according to the manufacturer's instructions. PCR amplification of the V4 region of the 16S rRNA gene was performed using the protocol developed by the Earth Microbiome Project (http://press.igsb.anl.gov/earthmicrobiome/empstandard-protocols/16s/) and modern primers [3]. Amplicons were sequenced on an Illumina MiSeq using paired, 250 base-pair reads, according to the manufacturer's instructions and are available on OSF (https://osf.io/9tycg/). Reads were merged and filtered using VSEARCH v2.3.0 [4], before dereplication and removal of singletons and chimeras using UCHIME denovo. Remaining reads were clustered at 97% similarity and a feature-abundance table was constructed by mapping

labelled reads to chimaera-checked clusters. Taxonomy was assigned to the centroid of each cluster using the Greengenes database [5]. Graphing was performed in R, making use of the Phyloseq package [6].

IHC and lectin staining

Mouse CC036 stomach and tumor sections were embedded in paraffin, sectioned at 4μm, rehydrated and deparaffinized. Slides were incubated overnight at 4°C with primary antibodies. Parietal cells were identified by staining for the hydrogen potassium ATPase using a mouse monoclonal antibody (1/200; ab2866, Abcam Ltd, Cambridge, UK). Rabbit anti-Trefoil Factor 3 monoclonal antibody (1/200; ab202967, Abcam Ltd, Cambridge, UK) was used to identify the mucus epithelia of the gastrointestinal tract. Secondary antibodies were visualized with DAB staining. Biotinylated Ulex Europaeus Agglutinin I (UEA I) from Vector Laboratories (10 ug/ml) was incubated for 30 min at room temperature to identify surface mucous cells and mucous neck cells, respectively. The slides were counterstained with hematoxylin for 5 minutes. Ki-67 (Abcam; ab16667) and Triplex (CD45R: abcam ab64100; CD3: Spring M3070 and CD11b: Novus NB110-89474) IHC staining was performed at the UCSF Biorepository and Tissue Biomarker Technology Core.

Supplementary Figure Legends

Supplementary Figure 1. Variation in disease susceptibility across Collaborative Cross strains. A. Representative H&E images of CC040 mouse lung showing plant material surrounded by lymphocytic inflammation. **B.** Representative H&E image of normal lung tissue. **C.** Representative H&E image of CC032 mouse kidney showing polycystic kidney. **D.** Representative H&E image of CC013 mouse kidney showing renal disease. **E.** Representative H&E image of normal kidney tissue.

Supplementary Figure 2. Histology of a representative stomach tumor in CC036 mice. A. H&E image of stomach tumor. **B.** Ki-67 immunohistochemical staining of stomach tumor. **C.** Triplex immunohistochemical staining for CD3 (teal), CD45R (purple) and CD11b (yellow). Note that the majority of stained cells are of myeloid origin (CD11b positive). Sparse staining was observed for B-cells (CD45R positive) and T-cells (CD3 positive).

Supplementary Figure 3. Histology of another representative stomach tumor in CC036 mice. A. H&E image of stomach tumor. **B.** Ki-67 immunohistochemical staining of stomach tumor. **C.** Triplex immunohistochemical staining for CD3 (teal), CD45R (purple) and CD11b (yellow). Note that the majority of stained cells are of myeloid origin (CD11b positive). Sparse staining was observed for B-cells (CD45R positive) and T-cells (CD3 positive).

Supplementary Figure 4. Immunohistochemical and lectin staining to characterize CC036 stomach tumors. A-D. Representative staining for UEA1, HK-ATPase and TFF3. All tumors were positive for UEA1 (**A**). **B.** Nests of UEA1 positive cells were also found invading beyond the muscularis mucosa into the submucosa (arrows in image). Only 25% of tumors retained any

HK-ATPase positive cells (arrows in image) (**C**). Scattered TFF3 staining (<25%) was observed in tumors (arrows in image) (**D**).

Supplementary Figure 5. Helicobacter abundance level across Collaborative Cross mice. A. Relative abundance of Helicobacter in fecal samples across 18 CC strains. Blue bars indicate mice derived from animal facility 1 and orange bars indicate mice derived from animal facility 2. **B.** Relative abundance of Helicobacter in stomach tissues across 6 CC strains. Blue bars indicate mice derived from animal facility 1 and orange bars indicate mice derived from animal facility 2. Note that Helicobacter was not detected in three strains from mice derived from animal facility 2. The difference in Helicobacter was not significantly different between CC036 and CC061. **C.** Stomach microbiome composition across individual CC036 mice at the family level and colored at the order level.

Supplementary Figure 6. Founder haplotype contributions to gastric cancer susceptibility on candidate genetic loci.

Supplementary Figure 7. Histology of stomach mucosa in CC036 mice at 12 weeks of age. A. Representative H&E images of gastric antrum in two CC036 mice. **B.** Ki-67 immunohistochemical staining of gastric antrum in the same two CC036 mice. **C.** Triplex immunohistochemical staining for CD3 (teal), CD45R (purple) and CD11b (yellow) of gastric antrum in the same two CC036 mice. Note that the majority of stained cells are of myeloid origin (CD11b positive). Sparse staining was observed for B-cells (CD45R positive) and T-cells (CD3 positive).

Supplementary Figure 8. Sex differences in gene expression and immune phenotypes in CC036 mice. A. Heatmap of genes significantly differentially expressed in stomach tissues of

male and female CC036 mice. Genes expressed at higher and lower levels are indicated in red and green, respectively. **B.** Representative sex differences in immune parameters measured in blood of male and female CC036 mice. Details on additional parameters are in supplementary table 4. P-values were obtained by non-parametric Mann-Whitney test.

Supplementary Figure 9. Gene expression analysis of BMPR1A and SMAD4 in stomach tissues of CC036 and other CC strains. A. RNA-sequencing read alignment along the BMPR1A genomic sequence (left). Gene expression level of BMPR1A in individual CC strains. CC036 mice are indicated in red and other CC strains in blue. No statistically significant difference was observed between CC036 and other CC strains (by Mann-Whitney non-parametric test; adj. p = 1). B. RNA-sequencing read alignment along the SMAD4 genomic sequence (left). Gene expression level of SMAD4 in individual CC strains. CC036 mice are indicated in red and other CC strains in blue. No statistically significant difference was observed between CC036 and other CC strains (by Mann-Whitney non-parametric test; adj. p = 1).

Supplementary Table 1. Overall and tumor free survival in Collaborative Cross mice.

Supplementary Table 2. OTU table of stomach microbiome.

Supplementary Table 3. Genes differentially expressed in gastric tissues between CC036 and control CC strains.

Supplementary Table 4. Blood cell counts in male and female CC036 mice.

Supplementary Table 5. Human orthologs of mouse genes differentially expressed in stomach tissues between CC036 and control strains.

Supplementary Table 6. Significance in association of gene expression with its copy number.

Supplementary Table 7. Genes significantly associated with OS in human gastric cancer patients.

References

1. Snijders AM, Langley S, Mao JH, et al. An interferon signature identified by RNA-sequencing of mammary tissues varies across the estrous cycle and is predictive of metastasis-free survival. *Oncotarget* 2014;5(12):4011-25. doi: 10.18632/oncotarget.2148
2. Hang B, Wang Y, Huang Y, et al. Short-term early exposure to thirdhand cigarette smoke increases lung cancer incidence in mice. *Clin Sci (Lond)* 2018;132(4):475-88. doi: 10.1042/CS20171521
3. Walters W, Hyde ER, Berg-Lyons D, et al. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems* 2016;1(1) doi: 10.1128/mSystems.00009-15
4. Rognes T, Flouri T, Nichols B, et al. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 2016;4:e2584. doi: 10.7717/peerj.2584
5. McDonald D, Price MN, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 2012;6(3):610-8. doi: 10.1038/ismej.2011.139
6. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;8(4):e61217. doi: 10.1371/journal.pone.0061217