nature

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41586-018-0811-x

In the format provided by the authors and unedited.

Age-related remodelling of oesophageal epithelia by mutated cancer drivers

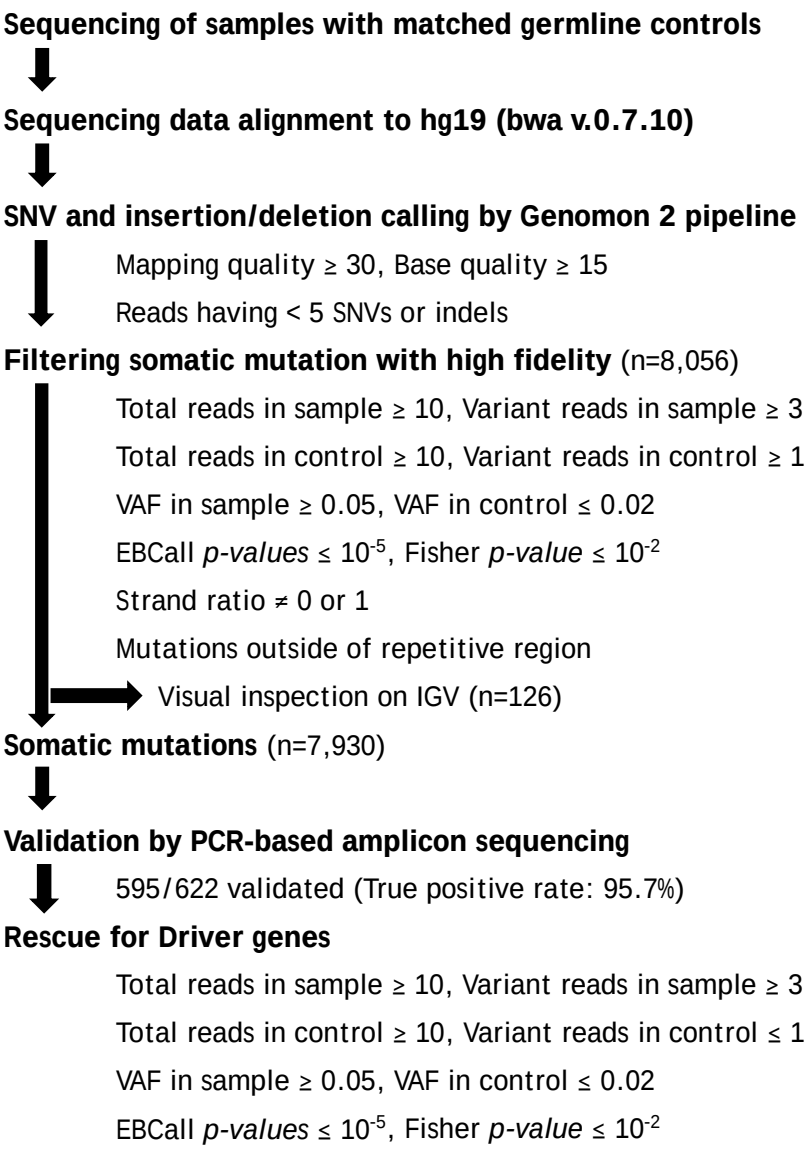
Akira Yokoyama[1,2,14], Nobuyuki Kakiuchi[1,3,14], Tetsuichi Yoshizato[1,14], Yasuhito Nannya[1], Hiromichi Suzuki[1], Yasuhide Takeuchi[1,4], Yusuke Shiozawa[1], Yusuke Sato[1], Kosuke Aoki[1], Soo Ki Kim[1], Yoichi Fujii[1], Kenichi Yoshida[1], Keisuke Kataoka[1], Masahiro M. Nakagawa[1], Yoshikage Inoue[1,5], Tomonori Hirano[1,3], Yuichi Shiraishi[6], Kenichi Chiba[6], Hiroko Tanaka[7], Masashi Sanada[8], Yoshitaka Nishikawa[2], Yusuke Amanuma[2], Shinya Ohashi[2], Ikuo Aoyama[2], Takahiro Horimatsu[2], Shin'ichi Miyamoto[3], Shigeru Tsunoda[5], Yoshiharu Sakai[5], Maiko Narahara[9], J. B. Brown[10], Yoshitaka Sato[11], Genta Sawada[12], Koshi Mimori[12], Sachiko Minamiguchi[4], Hironori Haga[4], Hiroshi Seno[3], Satoru Miyano[6,7], Hideki Makishima[1], Manabu Muto[2,15] & Seishi Ogawa[1,13,15]*

[1]Department of Pathology and Tumor Biology, Kyoto University, Kyoto, Japan. [2]Department of Clinical Oncology, Kyoto University Hospital, Kyoto, Japan. [3]Department of Gastroenterology and Hepatology, Kyoto University, Kyoto, Japan. [4]Department of Diagnostic Pathology, Kyoto University, Kyoto, Japan. [5]Department of Surgery, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [6]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [7]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [8]Department of Advanced Diagnosis, Clinical Research Center, Nagoya Medical Center, Nagoya, Japan. [9]Department of Human Genetics, McGill University, Montreal, Quebec, Canada. [10]Laboratory of Molecular Biosciences, Life Science Informatics Research Unit, Kyoto University, Kyoto, Japan. [11]Sato Clinic, Okayama, Japan. [12]Department of Surgery, Kyushu University Beppu Hospital, Beppu, Japan. [13]Department of Medicine, Center for Hematology and Regenerative Medicine, Karolinska Institute, Stockholm, Sweden. [14]These authors contributed equally: Akira Yokoyama, Nobuyuki Kakiuchi, Tetsuichi Yoshizato. [15]These authors jointly supervised this work: Manabu Muto, Seishi Ogawa. *e-mail: sogawa-tky@umin.ac.jp

Supplementary Information

Contents

Supplementary Method 1: **Flow chart of mutation calling**



hg19; Homo sapiens (human) genome assembly GRCh37

bwa; Burrows-Wheeler Aligner

SNV; single nucleotide variant

VAF; variant allele fraction

EBCall; Empirical Bayesian mutation Calling

IGV; Integrative Genomics Viewer

Supplementary Method 2; a sample script for a random effects model meta-analysis

```
#############################################################################
# load library
library(metafor)
#############################################################################
# function for calculating the coordinate of intersection point of perpendicular line from (x,y) to line with "a" slope
and "b" intercept
calc_dis <- function(a, b, x1, y1){
  x <- (x1+a*y1-b*a)/(a^2+1)
  return(c(x=x, y=a*x+b))
}
#############################################################################
# data read
res <- read.table("Path to data_for_script.txt")

res_nd <- subset(res, Path=="nondysplasia")
res_nd_4 <- subset(res_nd, Size==4)
res_nd_0.8 <- subset(res_nd, Size==0.8)
res_nd_0.2 <- subset(res_nd, Size==0.2)
res_nd_4_l <- subset(res_nd_4, ESCC_risk==0)
res_nd_4_h <- subset(res_nd_4, ESCC_risk==1)
res_nd_0.8_l <- subset(res_nd_0.8, ESCC_risk==0)
res_nd_0.8_h <- subset(res_nd_0.8, ESCC_risk==1)
res_nd_0.2_l <- subset(res_nd_0.2, ESCC_risk==0)
res_nd_0.2_h <- subset(res_nd_0.2, ESCC_risk==1)
#############################################################################
# This is an example of the code to calculate effect of ESCC risks on "mutation number".
# Modify this code for other metrices.
#############################################################################
# For 4mm2 sample
# Set up the regression line of samples without risk
out_4_l <- lm(res_nd_4_l$mut_num~res_nd_4_l$age-1);out_4_l
out2_4_l <- summary(out_4_l)

# Calculate the distance from the regression line to each risk (-) sample
```

```
dis_info_4_l <- NULL
for(i in 1:nrow(res_nd_4_l)){
  tmp_age <- res_nd_4_l[i,"age"]
  tmp_num <- res_nd_4_l[i,"mut_num"]
  tmp_res <- calc_dis(out_4_l$coefficients[1], 0,tmp_age, tmp_num)
  dis <- sqrt((tmp_age- tmp_res[1])^2+(tmp_num- tmp_res[2])^2)
  if(out_4_l$coefficients[1]*tmp_age -tmp_num<0){
    judge <- 1 #above
    dis <- dis
  }else{
    judge <- 0 # below
    dis <- -dis
  }
  tmp_res2 <- c(dis,judge)
  dis_info_4_l <- rbind(dis_info_4_l,tmp_res2)
}
colnames(dis_info_4_l) <- c("dis","judge")
rownames(dis_info_4_l) <- 1:nrow(dis_info_4_l)
dis_info_4_l <- as.data.frame(dis_info_4_l)

# Set up the regression line of samples with risk
out_4_h <- lm(res_nd_4_h$mut_num~res_nd_4_h$age-1);out_4_h
out2_4_h <- summary(out_4_h)

# Calculate the distance from the regression line to each risk (+) sample
dis_info_4_h <- NULL
for(i in 1:nrow(res_nd_4_h)){
  tmp_age <- res_nd_4_h[i,"age"]
  tmp_num <- res_nd_4_h[i,"mut_num"]
  tmp_res <- calc_dis(out_4_l$coefficients[1], 0,tmp_age, tmp_num)
  dis <- sqrt((tmp_age- tmp_res[1])^2+(tmp_num- tmp_res[2])^2)
  if(out_4_l$coefficients[1]*tmp_age -tmp_num<0){
    judge <- 1 #above
    dis <- dis
  }else{
    judge <- 0 # below
```

```
    dis <- -dis
  }
  tmp_res2 <- c(dis,judge)
  dis_info_4_h <- rbind(dis_info_4_h,tmp_res2)
}
colnames(dis_info_4_h) <- c("dis","judge")
rownames(dis_info_4_h) <- 1:nrow(dis_info_4_h)
dis_info_4_h <- as.data.frame(dis_info_4_h)

# Mann-Whitney U test
w.out <- wilcox.test(dis_info_4_h$dis,dis_info_4_l$dis, alternative="greater")

#############################################################################
# For 0.8mm2 sample
# Set up the regression line of samples without risk
out_0.8_l <- lm(res_nd_0.8_l$mut_num~res_nd_0.8_l$age-1);out_0.8_l
out2_0.8_l <- summary(out_0.8_l)

# Calculate the distance from the regression line to each risk (-) sample
dis_info_0.8_l <- NULL
for(i in 1:nrow(res_nd_0.8_l)){
  tmp_age <- res_nd_0.8_l[i,"age"]
  tmp_num <- res_nd_0.8_l[i,"mut_num"]
  tmp_res <- calc_dis(out_0.8_l$coefficients[1], 0,tmp_age, tmp_num)
  dis <- sqrt((tmp_age- tmp_res[1])^2+(tmp_num- tmp_res[2])^2)
  if(out_0.8_l$coefficients[1]*tmp_age -tmp_num<0){
    judge <- 1 #above
    dis <- dis
  }else{
    judge <- 0 # below
    dis <- -dis
  }
  tmp_res2 <- c(dis,judge)
  dis_info_0.8_l <- rbind(dis_info_0.8_l,tmp_res2)
}
colnames(dis_info_0.8_l) <- c("dis","judge")
```

```
rownames(dis_info_0.8_l) <- 1:nrow(dis_info_0.8_l)
dis_info_0.8_l <- as.data.frame(dis_info_0.8_l)

# Set up the regression line of samples with risk
out_0.8_h <- lm(res_nd_0.8_h$mut_num~res_nd_0.8_h$age-1);out_0.8_h
out2_0.8_h <- summary(out_0.8_h)

# Calculate the distance from the regression line to each risk (+) sample
dis_info_0.8_h <- NULL
for(i in 1:nrow(res_nd_0.8_h)){
  tmp_age <- res_nd_0.8_h[i,"age"]
  tmp_num <- res_nd_0.8_h[i,"mut_num"]
  tmp_res <- calc_dis(out_0.8_l$coefficients[1], 0,tmp_age, tmp_num)
  dis <- sqrt((tmp_age- tmp_res[1])^2+(tmp_num- tmp_res[2])^2)
  if(out_0.8_l$coefficients[1]*tmp_age -tmp_num<0){
    judge <- 1 #above
    dis <- dis
  }else{
    judge <- 0 # below
    dis <- -dis
  }
  tmp_res2 <- c(dis,judge)
  dis_info_0.8_h <- rbind(dis_info_0.8_h,tmp_res2)
}
colnames(dis_info_0.8_h) <- c("dis","judge")
rownames(dis_info_0.8_h) <- 1:nrow(dis_info_0.8_h)
dis_info_0.8_h <- as.data.frame(dis_info_0.8_h)

# Mann-Whitney U test
w.out <- wilcox.test(dis_info_0.8_h$dis,dis_info_0.8_l$dis, alternative="greater")

#############################################################################
# For 0.2mm2 sample
# Set up the regression line of samples without risk
out_0.2_l <- lm(res_nd_0.2_l$mut_num~res_nd_0.2_l$age-1);out_0.2_l
out2_0.2_l <- summary(out_0.2_l)
```

```
# Calculate the distance from the regression line to each risk (-) sample
dis_info_0.2_l <- NULL
for(i in 1:nrow(res_nd_0.2_l)){
  tmp_age <- res_nd_0.2_l[i,"age"]
  tmp_num <- res_nd_0.2_l[i,"mut_num"]
  tmp_res <- calc_dis(out_0.2_l$coefficients[1], 0,tmp_age, tmp_num)
  dis <- sqrt((tmp_age- tmp_res[1])^2+(tmp_num- tmp_res[2])^2)
  if(out_0.2_l$coefficients[1]*tmp_age -tmp_num<0){
    judge <- 1 #above
    dis <- dis
  }else{
    judge <- 0 # below
    dis <- -dis
  }
  tmp_res2 <- c(dis,judge)
  dis_info_0.2_l <- rbind(dis_info_0.2_l,tmp_res2)
}
colnames(dis_info_0.2_l) <- c("dis","judge")
rownames(dis_info_0.2_l) <- 1:nrow(dis_info_0.2_l)
dis_info_0.2_l <- as.data.frame(dis_info_0.2_l)

# Set up the regression line of samples with risk
out_0.2_h <- lm(res_nd_0.2_h$mut_num~res_nd_0.2_h$age-1);out_0.2_h
out2_0.2_h <- summary(out_0.2_h)

# Calculate the distance from the regression line to each risk (+) sample
dis_info_0.2_h <- NULL
for(i in 1:nrow(res_nd_0.2_h)){
  tmp_age <- res_nd_0.2_h[i,"age"]
  tmp_num <- res_nd_0.2_h[i,"mut_num"]
  tmp_res <- calc_dis(out_0.2_l$coefficients[1], 0,tmp_age, tmp_num)
  dis <- sqrt((tmp_age- tmp_res[1])^2+(tmp_num- tmp_res[2])^2)
  if(out_0.2_l$coefficients[1]*tmp_age -tmp_num<0){
    judge <- 1 #above
    dis <- dis
```

```
  }else{
    judge <- 0 # below
    dis <- -dis
  }
  tmp_res2 <- c(dis,judge)
  dis_info_0.2_h <- rbind(dis_info_0.2_h,tmp_res2)
}
colnames(dis_info_0.2_h) <- c("dis","judge")
rownames(dis_info_0.2_h) <- 1:nrow(dis_info_0.2_h)
dis_info_0.2_h <- as.data.frame(dis_info_0.2_h)

# Mann-Whitney U test
w.out <- wilcox.test(dis_info_0.2_h$dis,dis_info_0.2_l$dis, alternative="greater")

# Apply the same calculation to 0.8mm2 and 0.2 mm2 samples

# Meta analysis for ESCC risk
out.m <- rbind(c(mean(dis_info_0.2_h$dis),sd(dis_info_0.2_h$dis),length(dis_info_0.2_h$dis),
                 mean(dis_info_0.2_l$dis),sd(dis_info_0.2_l$dis),length(dis_info_0.2_l$dis)),
               c(mean(dis_info_0.8_h$dis),sd(dis_info_0.8_h$dis),length(dis_info_0.8_h$dis),
                 mean(dis_info_0.8_l$dis),sd(dis_info_0.8_l$dis),length(dis_info_0.8_l$dis)),
               c(mean(dis_info_4_h$dis),sd(dis_info_4_h$dis),length(dis_info_4_h$dis),
                 mean(dis_info_4_l$dis),sd(dis_info_4_l$dis),length(dis_info_4_l$dis)))
colnames(out.m) <- c("mean_h","sd_h","num_h","mean_l","sd_l","num_l")
out.m <- as.data.frame(out.m)

# Calculate effect sizes and outcome measures
# "SMDH" for the standardized mean difference with heteroscedastic population variances in the two groups (Bonett,
2008, 2009).
out.m2 <- escalc(measure="SMDH", m1i=mean_h,m2i=mean_l,
                 sd1i=sd_h,sd2i=sd_l,n1i=num_h,n2i=num_l, data=out.m)

# Random-effects model
out.m3=rma(yi, vi, data=out.m2)

# Plot forest plot
```

```
forest(out.m3,showweights=T,alim=c(-4,4),at=(-4:4),
       slab=c("0.2 mm2","0.8 mm2", "4 mm2"),cex=1.2,cex.lab=1,cex.axis=1,
       main="Number of mutations and ESCC risk")
text(1,5,paste("P=",signif(out.m3$pval,3)))

#########################################################################
# Meta analysis for age
out.m <- rbind(c(sqrt(out2_0.2_h$r.squared),length(dis_info_0.2_h$dis),
                 sqrt(out2_0.2_l$r.squared),length(dis_info_0.2_l$dis)),
               c(sqrt(out2_0.8_h$r.squared),length(dis_info_0.8_h$dis),
                 sqrt(out2_0.8_l$r.squared),length(dis_info_0.8_l$dis)),
               c(sqrt(out2_4_h$r.squared),length(dis_info_4_h$dis),
                 sqrt(out2_4_l$r.squared),length(dis_info_4_l$dis)))
colnames(out.m) <- c("coef_h","num_h","coef_l","num_l")
out.m <- as.data.frame(out.m)

# Calculate effect sizes and outcome measures
# "ZCOR" for the Fisher's r-to-z transformed correlation coefficient (Fisher, 1921).
out.m2 <- escalc(measure="ZCOR", ri=coef_l,ni=num_l,data=out.m)

# Random-effects model
out.m3=rma(yi, vi, data=out.m2)

# Plot forest plot
forest(out.m3,showweights=T,alim=c(-2,2),at=(0:20-10)/10,
       slab=c("0.2 mm2","0.8 mm2", "4 mm2"),cex=1.2,cex.lab=1,cex.axis=1,
       main="Number of mutations and age in risk (-)")
text(1,5,paste("P=",signif(out.m3$pval,3)))
```

Supplementary Note 1. Acknowledgements

This work was supported by Grants-in-Aid from the Ministry of Health, Labor and Welfare of Japan (Health and Labour Sciences Research Expenses for Commission and Applied Research for Innovative Treatment of Cancer), the Project for Development of Innovative Research on Cancer Therapeutics from the Japan Agency for Medical Research and Development, AMED (JP15cm0106056h0005 and JP16cm0106501h0001) [Kyoto; S. Ogawa], the practical research for innovative cancer control from AMED (JP18ck0106296) [Kyoto; M.M.], the Ministry of Education, Culture, Sports, Science and Technology of Japan, the High Performance Computing Infrastructure System Research Project (hp150232; This research used computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project) [Kyoto; S. Ogawa, Tokyo; S.M.], Scientific Research on Innovative Areas (15H05909) [Kyoto; S. Ogawa, Tokyo; S.M.], "Stem Cell Aging and Disease" (14430052) [Nagoya; M.S.], Takeda Science Foundation [Kyoto; S. Ogawa, H.M., T.Y.], Uehara Memorial Foundation [Kyoto; H.S.], and Foundation for Promotion of Cancer Research [Kyoto; H.S.]. S. Ogawa and J.B.B. are recipients of JSPS Core-to-Core Program, A. Advanced Research Networks. We thank Yukiko Nakai, Tomomi Ida, Maki Nakamura, Takeshi Shirahari, and Ai Takatsu for technical assistance and Ravi Velaga for English editing. Cryo-preserved oesophageal samples were provided from the Kyoto University Hospital Cancer Biobank. We thank the Centre for Anatomical, Pathological and Forensic Medical Research, Kyoto University Graduate School of Medicine and Masahiro Hirata, for preparing microscope slides. We also thank all staffs at the Endoscopy Unit, Kyoto University Hospital for endoscopic sampling.