Journal Pre-proof

Germline Determinants of Esophageal Adenocarcinoma

Minyi Lee, B.S., George Eng, M.D., Ph.D., Anna Handte-Reinecker, B.S., MGH-MIT Gastrointestinal Cohorts Working Group, Vikram S. Deshpande, M.B.B.S., Omer H. Yilmaz, M.D., Ph.D., Manish K. Gala, M.D.

PII: S0016-5085(23)04789-3
DOI: https://doi.org/10.1053/j.gastro.2023.07.015
Reference: YGAST 65815

To appear in: *Gastroenterology*
Accepted Date: 20 July 2023

Please cite this article as: Lee M, Eng G, Handte-Reinecker A, MGH-MIT Gastrointestinal Cohorts Working Group, Deshpande VS, Yilmaz OH, Gala MK, Germline Determinants of Esophageal Adenocarcinoma, *Gastroenterology* (2023), doi: https://doi.org/10.1053/j.gastro.2023.07.015.

Title: Germline Determinants of Esophageal Adenocarcinoma

Short Title: Germline Determinants of Esophageal Adenocarcinoma

Minyi Lee, B.S.[1,2†], George Eng, M.D., Ph.D.[1,3,4†], Anna Handte-Reinecker, B.S.[1], MGH-MIT Gastrointestinal Cohorts Working Group, Vikram S. Deshpande, M.B.B.S.[5], Omer H. Yilmaz, M.D., Ph.D.[3,4,6], Manish K. Gala, M.D.[1,7*],

1. Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA
2. M.D.-Ph.D Program, Boston University School of Medicine, Boston, MA
3. Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA
4. David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA
5. Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA
6. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA
7. Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA

† Denotes shared first authorship

***Corresponding Author:** Dr. Manish K. Gala, MD
55 Fruit Street, GRJ 708
Boston, MA 02114
Email: mgala@mgh.harvard.edu
Phone: 617-726-7892
Fax: 978-882-6710

Conflict of Interest Statement: The authors have no relevant conflicts to disclose.

Grant Support: This work was supported by grants from the Elsa U. Pardee Foundation (Manish Gala), American College of Gastroenterology Junior Faculty Career Development Award (Manish Gala), and the NIDDK (K23DK103119)

Abbreviations: BE – Barrett's esophagus; HGD – high-grade dysplasia; HRD – homolgous recombination deficient; HRR – homologous recombination repair; MGH – Massachusetts General Hospital;

Acknowledgements:
Members of the MGH-MIT Gastrointestinal Cohorts Working Group are:
Alvin Jeon, M.D.[1], Eugeniu Stratulat, M.D.[1], Rachid Zagani, Ph.D.[1], Martin S. Taylor, M.D., Ph.D.[2], Steffen Rickelt, Ph.D.[3], Dmitriy Kedrin, M.D.,Ph.D.[1], Anupam Batra, M.D.[4,5,6], Ardeshir Hashmi, M.D.[4,7], Norman S. Nishioka, M.D.[1], Chin Hur, M.D., M.P.H.[1,8,9], Lipika Goyal, M.D.[10], Fateh Bazerbachi, M.D.[1,11], Lawrence Zukerberg, M.D.[2#]

1. Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA
2. Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA
3. David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA

4. Massachusetts General Hospital Division of Senior Health and Harvard Medical School, Boston, MA
5. Department of Oncology, Cambridge Memorial Hospital, Cambridge, Ontario, Canada
6. Department of Oncology, Grand River Regional Cancer Center, Kitchener, Ontario, Canada
7. Center for Geriatric Medicine, Cleveland Clinic, Cleveland, OH, USA
8. Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY
9. Division of Digestive and Liver Diseases, Columbia University Vagelos School of Medicine, New York, New York
10. Division of Oncology, Massachusetts General Hospital Cancer Center and Harvard Medical School, Boston, MA
11. Interventional Endoscopy Program, CentraCare and St. Cloud Hospital, St. Cloud, MN
12. Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA
13. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA
14. Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA

Author is recently deceased

Author Contributions: Minyi Lee: acquisition of data, analysis and interpretation of data, drafting of the manuscript, statistical analysis; **George Eng**: acquisition of data, analysis and interpretation of data, drafting of the manuscript; **Anna Handte-Reinecker**: acquisition of data, drafting of the manuscript; **Alvin Jeon:** acquisition of the data, analysis and interpretation of the data; **Eugeniu Stratulat:** acquisition of the data; **Rachid Zagani:** acquisition of the data; **Martin Taylor:** acquisition of the data, analysis and interpretation of the data, drafting of the manuscript; **Steffen Rickelt**: acquisition of the data; **Dmitriy Kedrin**: analysis and interpretation of the data; **Anupam Batra:** acquisition of the data; **Ardeshir Hashmi:** acquisition of the data, study supervision; **Norman Nishioka**: acquisition of the data; **Chin Hur**: acquisition of the data, drafting of the manuscript; **Lipika Goyal:** acquisition of the data; **Fateh Bazerbachi**: acquisition of the data, drafting of the manuscript; **Lawrence Zukerberg:** acquisition of the data, analysis and interpretation of the data; **Vikram Deshpande:** acquisition of the data, analysis and interpretation of the data; **Omer Yilmaz:** technical or material support, study supervision; **Manish Gala:** study concept and design, analysis and interpretation of data, statistical analysis, drafting of the manuscript, obtained funding, study supervision.

Data Transparency Statement: Genomic data from OCCAMS-UK is available from ICGC-ARGO (https://www.icgc-argo.org) and dbGAP (phs000178.v10, phs000598.v2 and phs001783.v1). Sequencing results from MGH Cohorts will be provided with reasonable requests and data protections.

The risk of progression for Barrett's esophagus (BE) is estimated to range from 0.12% to 0.5% per year.[1] Identification of clinical risk factors such as age, sex, obesity, smoking, presence of hiatal hernia, and length of BE, are insufficient to wholly account for the few individuals who progress from BE to adenocarcinoma.[2] To explain some of the unaccounted risk, we hypothesized that a significant fraction of individuals with BE who progress to adenocarcinoma harbor pathogenic germline mutations in cancer predisposing genes.

We examined the prevalence of monoallelic, pathogenic germline mutations associated with moderate to high risk of cancer in 640 study participants with esophageal adenocarcinoma (EAC) enrolled in publicly available genomic cohorts that performed either whole genome sequencing (ICGC-ARGO) or whole-exome sequencing (TCGA Pan-Cancer Cohort, Broad Institute Esophageal Adenocarcinoma Cohort, and Memorial Sloan Kettering Prospective Clinical Cohort, **Figure 1A**).[3-6] Pathogenic germline mutations were discovered in 59 out of 640 individuals (9.2%, **Figure 1B, Supplementary Table 1**). *ATM* was the most frequently mutated gene, occurring in 10 individuals (1.6%) followed by *CHEK2* (1.25%). Five individuals (0.8%) harbored germline mutations in *TP53*. Two individuals (0.3%) harbored distinct, splice-donor mutations in *CDH1* at intron 10. Despite this prevalence, somatic coding mutations that represent likely loss-of-heterozygosity events, were only present in 3/60 tumors (5.0%, 1 *BRCA2*- and 2 *TP53* mutation carriers).

As validation, we performed germline WES on prospective cohorts at Massachusetts General Hospital that encompass BE progressors who developed high-grade dysplasia or intramucosal carcinoma (102 individuals), BE without progression to dysplasia over 10+ years (75 individuals), and healthy nonagenarians without any prior known history of gastrointestinal neoplasia (100 individuals). Again, germline *ATM* mutations were the most frequent pathogenic alteration, occurring in 2% and 2.7% of progressors and non-progressors (short-segment BE),

respectively. Despite the lack of enrichment of *ATM* carriers among progressors, immunohistochemistry demonstrated loss of ATM staining among progressors and retained expression among non-progressors, implying epigenetic mechanisms for LOH (**Supplementary Figure 1A**).

Across all HGD/EAC cohorts, the prevalence of germline mutations in genes associated with monoallelic cancer predisposition within the Fanconi Anemia pathway (*BRCA2, PALB2, BRIP1, RAD51C, FANCA, FANCC, FANCM)* demonstrated enrichment over the carrier rate for all Fanconi Anemia genes in the general population (overall 2.3% vs. 0.6%). The age at diagnosis of those with high-grade dysplasia or adenocarcinoma did not differ between those with or without any germline mutations (**Figure 1C).**

Given this enrichment of pathogenic germline mutations in progressors, we examined if such germline alterations could influence the somatic mutanome. We examined the association of germline mutations with the development of pathogenic somatic *TP53* alterations, since such alterations have been associated with BE progression and genome doubling events (**Figure 1D**).[7] Pathogenic, somatic *TP53* mutations were detected among 75% of tumor exomes and 70% of tumor genomes. When stratified by somatic *TP53* mutant status, pathogenic germline mutations were present in 16.7% of cancer exomes with wild-type *TP53* versus 7.2% with*TP53* mutations (OR 2.6, 95% C.I. 0.9-6.8, $P = 0.04$, Fisher's exact test). Among cancer genomes, germline mutations were present in 15.6% of cases with wild-type *TP53* versus 6.1% of *TP53* mutants (OR 2.8, 95% C.I. 1.3-6.2, $P = 0.004$ Fisher's exact test).

To examine if the overall enrichment of germline mutations among *TP53* wildtype tumors is driven by select genes, we stratified somatic *TP53* mutant status by each cancer-predisposing gene (**Figure 1E**). *ATM* germline mutations demonstrated 100% mutual exclusivity with

pathogenic somatic *TP53* mutations (OR 0, 95% CI 0-0.2, $P = 2.9 \times 10^{-6}$, Fisher's exact test). We validated this mutual exclusivity with an independent cohort of 475 publicly available and non-redundant gastroesophageal adenocarcinomas previously sequenced on the MSK-IMPACT platform, with 7/7 *ATM* carriers harboring wild-type *TP53* (**Supplementary Figure 1B**). Exclusion of *ATM* carriers still demonstrated a persistent enrichment of germline mutations among *TP53* wildtype tumors, occurring in 10.1% and 12.1% of exomes and genomes, respectively.

Pathogenic germline *BRCA2* mutations also demonstrated a trend toward mutual exclusivity with *TP53* mutation (OR 0.2, 95% C.I. 0.2-1.4, $P = 0.06$, Fisher's exact test). Given the strong association of homologous recombination deficiency with somatic *TP53* mutations, we examined HRD status from tumor genomes utilizing the HRDetect algorithm. We observed HRD present in only 14/400 (3.5%) of tumor whole genomes, with only 1/4 *BRCA2* carriers demonstrating HRD (**Supplementary Figure 1C**). Among tumor exomes with either *BRCA2* or *PALB2* germline alterations, no samples demonstrated dominance by the single base substitution signature associated with HRD (Sig3, **Supplementary Figure 1D**).

Among 742 individuals with BE with HGD or EAC, we identified pathogenic germline mutations in monoallelic, cancer-predisposing genes among 9.0% of participants, compared to 2.7% of non-progressors. This overall enrichment suggests that these mutations facilitate the progression of Barrett's esophagus to adenocarcinoma. The ages of onset for those with germline mutations did not cluster among earlier-onset cases but occurred throughout the age spectrum, implying that these inherited mutations may require the development of BE and additional environmental factors as prerequisites to promote esophageal carcinogenesis.

Somatic *TP53* alterations have been identified as a key driver in the progression of nondysplastic Barrett's esophagus to dysplasia, functioning as a checkpoint for genome doubling events and chromosomal instability.[7] Validating its role as a key driver of progression, we did observe an overrepresentation of germline *TP53* mutations (0.7% among progressors). However, 25-30% of esophageal adenocarcinomas lacked somatic alterations in *TP53*. We discovered that such *TP53* wild-type tumors were significantly enriched for pathogenic germline mutations compared to *TP53*-mutant cancers (overall 15.9% vs. 6.6%, OR 2.7, 95% CI 1.5-4.8, $P = 4.2 \times 10^{-4}$, Fisher's exact test). This enrichment implies an early and causative role for even heterozygous germline mutations in BE progression since they can obviate the selection pressures for the acquisition of somatic *TP53* coding mutations. Multiple studies have demonstrated that heterozygosity of cancer predisposing genes can promote genomic instability.[8,9] Genome-wide association studies have quantified moderate effects associated with rare, heterozygous germline mutations.

Genetic testing has been recommended for all individuals diagnosed with pancreatic adenocarcinoma, where the prevalence of germline mutations is 7-10% and second hit mutations are uncommon.[10] Given the similar prevalence in EAC, universal genetic testing should be considered.

Figure Legends.

Figure 1: Germline Mutational Landscape Across Esophageal Adenocarcinoma. (A) Clinical characteristics of study participants from public genomic and MGH cohorts. ICGC-ARGO refers to International Cancer Genome Consortium Project Accelerating Research in Genomic Oncology; Broad/MSKCC Cohort refers to the pooled public exomes of esophageal adenocarcinoma available on dbGAP; TCGA refers to The Cancer Genome Atlas; Wellderly refers to healthy nonagenarians without history of gastrointestinal neoplasia. **(B)** Number of pathogenic mutations itemized by cancer-predisposing genes across multiple cohorts. Color-coding of entries demonstrates carrier-frequency in their respective cohorts. **(C)** Histogram showing the age at diagnosis of Barrett's esophagus with high-grade dysplasia or esophageal adenocarcinoma. Mutation carriers and non-carriers are color-coded by blue and gray, respectively. **(D)** Correlation of germline pathogenic mutations with somatic *TP53* status in tumors, segregated by exomes and genomes. **(E)** Correlations between individual genes mutated in the germline and somatic *TP53* status. *** designates $P < 0.001$ and * designates $P = 0.06$.

References.

1. Hvid-Jensen F, et al. N Engl J Med 2011;365:1375-83.

2. Parasa S, et al. Gastroenterology 2018;154:1282-1289 e2.

[dataset] 3. Fitzgerald R, et al. OCCAMS-UK. ICGC-ARGO. 2022. https://platform.icgc-argo.org/

[dataset] 4. The Cancer Genome Atlas Network. Esophageal Carcinoma. dbGAP. 2022. phs000178.v10.

[dataset] 5. Bass A, et al. Exome Sequencing of Esophageal Adenocarcinoma. dbGAP. 2021. phs000598.v2.

[dataset] 6. Solit D, et al. Exome recapture and sequencing of prospectively characterized clinical specimens from cancer patients. dbGAP. 2022. phs001783.v1

7. **Stachler MD, Taylor-Weiner A**, et al. Nat Genet 2015;47:1047-55.

8. Karaayvaz-Yildirim M, et al. Sci Adv 2020;6:eaay2611.

9. Oliveira C, et al. Gastroenterology 2009;136:2137-48.

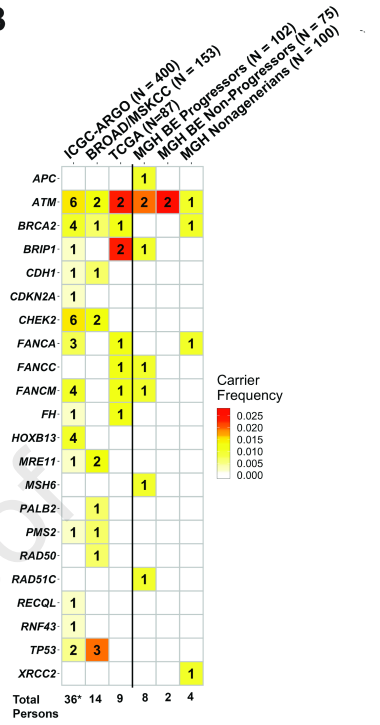
10. **Yurgelun MB, Chittenden AB, Morales-Oyarvide V**, et al. Genet Med 2019;21:213-223.

Author names in bold designate shared co-first authorship.

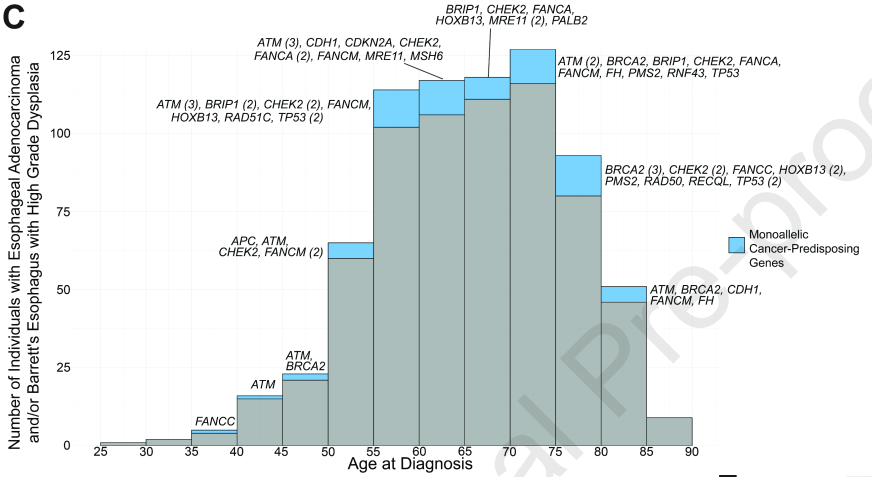
A

	Public Genomic Cohorts			MGH Cohorts		
	ICGC-ARGO (N = 400)	Broad/ MSKCC (dbGAP) (N=153)	TCGA (N = 87)	Barrett's Esophagus Progressors (N=102)	Barrett's Esophagus Non-Progressors (N=75)	Wellderly (N=100)
Sex						
Male	344 (86 %)	127 (83 %)	75 (86 %)	83 (81 %)	57 (76 %)	32 (32 %)
Female	56 (14 %)	26 (17 %)	12 (14 %)	19 (19 %)	18 (24 %)	68 (68 %)
Age at Diagnosis or Enrollment						
Mean (SD)	66 (± 10)	69 (± 11)	67 (± 12)	64 (± 10)	68 (± 10)	92 (± 2.5)
Unknown	NA	1 (0.7%)	NA	NA	NA	NA
Length of Barrett's Esophagus (BE)						
Short Segment BE	NA	NA	NA	42 (41 %)	48 (64 %)	NA
Long Segment BE	NA	NA	NA	47 (46 %)	25 (33 %)	NA
Very Long Segment BE	NA	NA	NA	8 (8 %)	2 (3 %)	NA
Unknown	400 (100%)	153 (100%)	87 (100 %)	5 (4.9%)	NA	NA

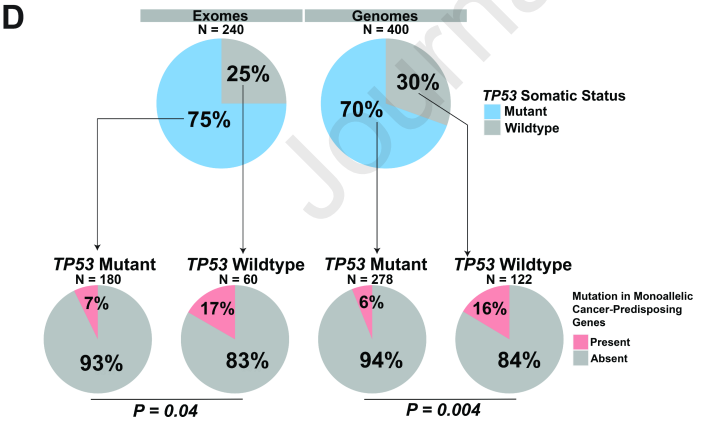
B



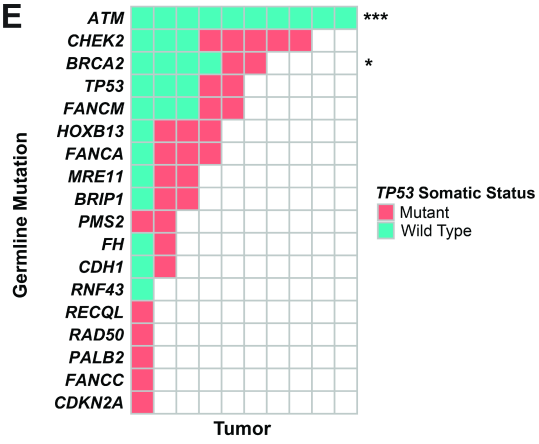
C



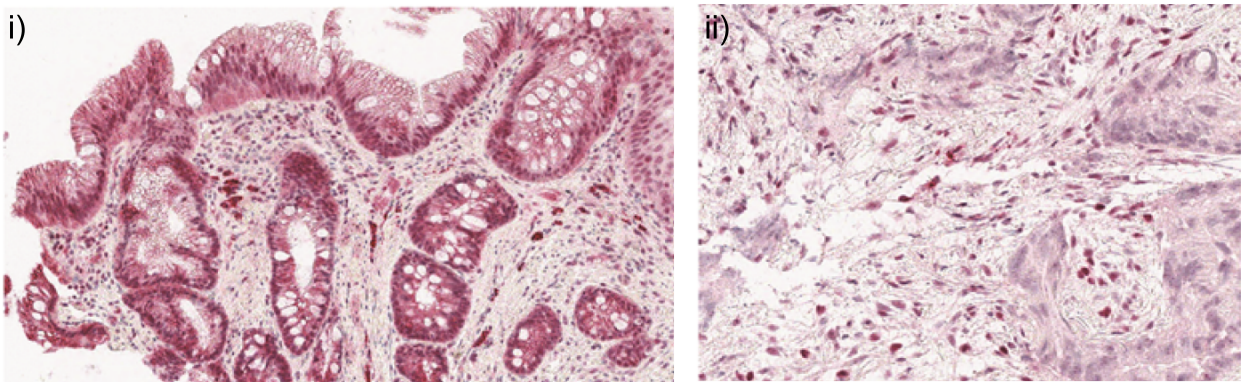
D



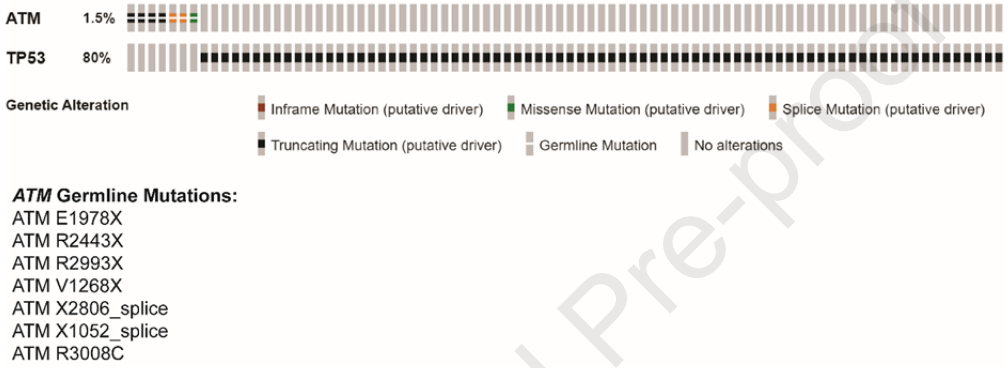
E



A

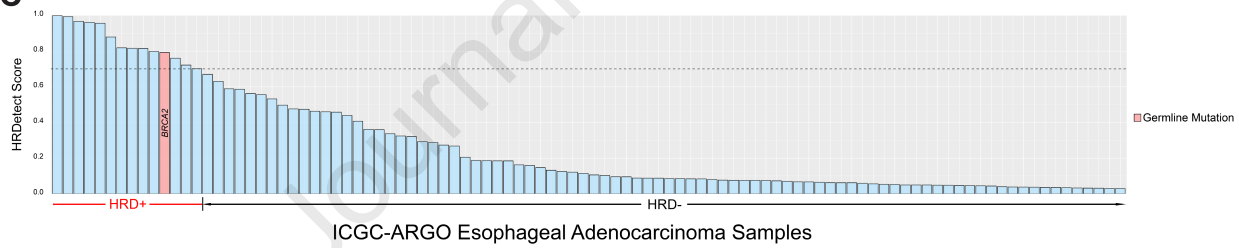


B

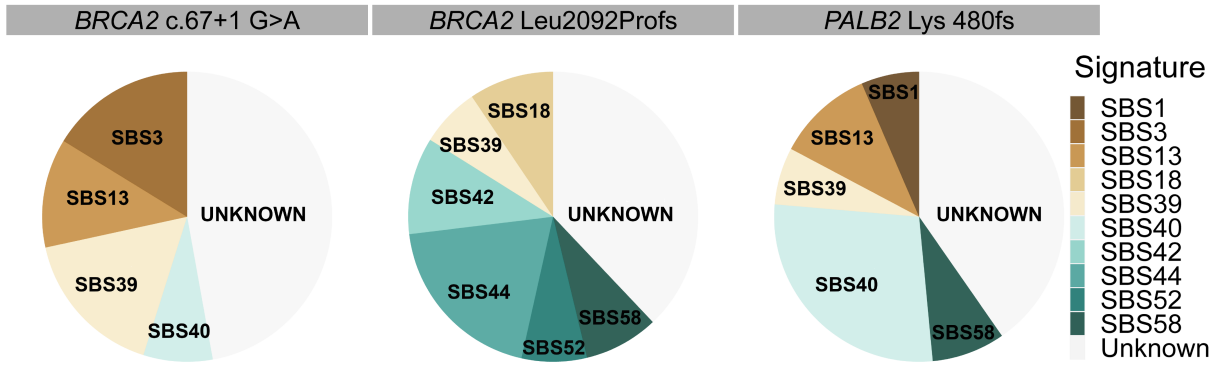


***ATM* Germline Mutations:**
ATM E1978X
ATM R2443X
ATM R2993X
ATM V1268X
ATM X2806_splice
ATM X1052_splice
ATM R3008C

C



D



Supplementary Figure 1.

(A) Immunohistochemistry staining of ATM with a red chromogen. **i)** Ample ATM staining persists in the epithelium of Barrett's esophagus without dysplasia. **ii)** ATM staining is lost in the esophageal adenocarcinoma of a carrier, but remains positive in the surrounding tissues. **(B)** Mutual exclusivity of pathogenic germline *ATM* mutations and *TP53* in MSK-IMPACT (Esophageal Adenocarcinoma). Pathogenic germline ATM mutations demonstrated 100% mutual exclusivity with somatic driver TP53 mutations in the cohort labeled as Esophageal/Stomach Cancer (MSK, 2020) from cBioPortal. **(C)** Homologous recombination deficiency (HRD) scores across all cancer genomes. The HRDetect algorithm was used with a prespecified score greater than 0.7 to designate HRD. 14/400 tumors demonstrated HRD, with only 1/4 *BRCA2* carriers with HRD. No carriers in other homologous recombination genes demonstrated HRD **(C)** Single base substitution signatures from tumor exomes of *BRCA2* and *PALB2* carriers. Tumors with dominant SBS3 (Sig3) are associated with HRD status.

Supplementary Methods.

Participants.

Integrated germline and somatic whole-genome or whole-exome analyses were performed across deposited data from 400/400 United Kingdom participants of the Oesophageal Cancer Clinical and Molecular Stratification (OCCAMS-GB) deposited in the International Cancer Genome Consortium Project Accelerating Research in Genomic Oncology (ICGC-ARGO), 87 participants in The Cancer Genome Atlas (TCGA) with esophageal adenocarcinoma (dbgap phs000178.v10), and 153 participants in dbGAP cohorts (phs000598.v2 and phs001783.v1) assembled by the Broad Institute (149/149 whole exome) and Memorial Sloan Kettering Cancer Center (4/4 whole exome), respectively. The Broad Institute genomic cohort of esophageal adenocarcinoma consisted of participants recruited from the University of Pittsburgh (Pittsburgh, PA), the University of Michigan (Ann Arbor, MI), and the Ontario Tissue Bank (Toronto, Canada). Written informed consent was obtained from all participants and local institutional IRB approval was obtained before deposition into data repositories.

Germline exome-wide analyses were performed on prospectively collected cohorts from the Massachusetts General Hospital (MGH) Barrett's Patient Registry (102 individuals with esophageal Adenocarcinoma or Barrett's esophagus with high-grade dysplasia, and 75 individuals with Barrett's esophagus without dysplasia who failed to progress over 10 years) and MGH Wellderly Cohort (100 nonagenarian participants without a known history of gastrointestinal neoplasia). An additional participant under the Barrett's Patient Registry was enrolled for organoid collection. Written informed consent was obtained from all participants and approved by the Massachusetts General Brigham IRB (protocols 2010P002224, 2016P000846, and 2015P000584).

Validation of specific germline-somatic correlations in esophageal adenocarcinoma was performed with 475 non-redundant participants (out of a total 478) enrolled by Memorial Sloan Kettering Cancer Center and sequenced on the MSK-IMPACT targeted sequencing platform. Germline and somatic mutational data were deposited in on the cBioPortal under the study Esophageal/Stomach Cancer (MSK, 2020) with participants with the histological subtypes of esophageal adenocarcinoma, adenocarcinoma of the gastroesophageal junction, esophagogastric adenocarcinoma, and esophageal poorly differentiated carcinoma.

Germline Sequencing.

Moderate to high-risk monoallelic cancer-predisposing genes were curated prospectively from the union set of commercially-available hereditary cancer panels (Ambry Genetics, Blueprint Genetics, Color Genomics, GeneDx, Invitae, and Fulgent). *AIP, ALK, AKT, APC, ATM, AXIN2, BAP1, BARD1, BMPR1A, BRCA1, BRCA2, BRIP1, CDC73, CDH1, CDK4, CDKN1B, CDKN1C, CDKN2A, CEBPA, CTNNA1, CHEK2, DICER1, DIS3L2, EPCAM, FANCA, FANCC, FANCM, FLCN, GATA2, GALNT12, GEN1, HOXB13, HRAS, KIT, MAX, MET, MLH1, MRE11, MSH2, MSH6, NBN, NF1, NF2, PALB2, PALLD, PIK3CA, PHOX2B, PMS2, POT1, PTEN, PRKAR1A, PTCH1, RAD50, RAD51C, RAD51D, RB1, RECQL, RET, RINT, RNF43, RPS20, RUNX1, SDHA, SDHAF2, SDHB, SDHC, SDHD, SMAD4, SMARCA4, SMARCB1, SMARCE1, STK11, SUFU, TSC1, TSC2, TERC, TERT, TMEM127, TP53, WT1,* and *VHL* were selected.

Downloaded from their respective repositories, aligned BAM files of inferred germline DNA (derived from either blood or normal tissue) sequenced on Illumina HiSeq or NovaSeq platforms were analyzed using the Genome Analysis Toolkit (GATK) with hard filtration recommendations, annotated with Annovar, and subsequently inspected in the Integrated Genomics Viewer (IGV).[1-4] For validation cohorts from the MGH Barrett's Patient Registry and MGH Wellderly cohorts,

whole-exome sequencing was performed on germline DNA on Illumina HiSeq2500 instruments with 100-bp paired-end reads.

Pathogenic variants in cancer-predisposing genes were identified as established ClinVar annotations (Pathogenic or Likely Pathogenic), or loss-of-function mutation (nonsense, frameshift, and essential splice-site) in those predisposition genes that serve as tumor suppressor genes. Loss-of-function mutations with minor allele frequency less than 1% that affected all transcript isoforms (RefSeq) or occurring in isoform-specific exons harboring already known ClinVar pathogenic or likely pathogenic annotations were included.

Somatic Mutational Analyses.

For datasets obtained from ICGC-ARGO and TCGA, somatic mutational profiles were obtained and analyzed in their respective repositories. For ICGC-ARGO samples, filtered variant calls (single nucleotide variant and insertion/deletions) by the GATK Mutect2 algorithm were utilized. Copy number variation analyzed by ASCAT, and structural variation from BRASS were utilized for homologous recombination deficiency analyses.[5, 6] For TCGA samples, GATK Mutect2 filtered variant calls were utilized. Copy number variation analyses were performed with the Sequenza package in R.[7] Cancer exomes from the Broad Institute and Memorial Sloan Kettering Cohorts were analyzed with the identical pipeline to those samples analyzed from the TCGA Pan-cancer cohort.[8]

Homologous recombination deficiency was inferred from somatic sequencing by two methods. For tumors sequenced by whole genome sequencing the HRDetect computational algorithm was employed.[9] HRDetect scores were calculated using a logistic regression classifier. The classifier requires six features: proportion of short deletions with microhomology at the breakpoint junction, number of mutations attributed to COSMIC single base substitution

signatures 3 and 8 and to rearrangement signatures 3 and 5, and HRD-LOH index. Features were calculated using the R package signature.tools.lib. HRDetect scores were computed both as point estimates and as a distribution obtained from 1000 bootstrapped scores. HRDetect scores ≥ 0.7 were deemed homologous recombination deficient (HRD+), as previously performed.[9-11] For tumors, sequenced by whole exome-sequencing, COSMIC single base substitution signature 3 (Sig3) ratios were calculated from passed Mutect2 calls using the R package deconstructSigs.[12] A detectable Sig3 component was noted as a Sig3+ tumor.

Immunohistochemistry.

Immunohistochemistry for ATM [mouse monoclonal, ab78, 2C1 (1A1), dilution 1:2000, Abcam] was performed on paraffin embedded whole slides tissues sections using the automated LabVision Autostainer 360 (ThermoScientific). After primary antibody incubation and several washes, the secondary ImmPRESS polymer detection system (MP-5402, Vector Laboratories) was used according to the manufacturers protocols. The Vulcan Fast Red Chromogen Kit 2 (red staining; Biocare Medical) was applied as substrate. Image documentation was performed using the Leica Aperio AT2 slide scanner system.

Statistical Analyses.

Fisher exact tests were performed to compare the number of study participants with or without pathogenic germline mutations in the context of somatic *TP53* mutational status. Fisher exact tests were also used to determine the mutual exclusivity of cancer-predisposing genes with somatic *TP53* mutations.

Supplementary References

1. McKenna A, et al. Genome Res 2010;20:1297-303.
2. Robinson JT, et al. Nat Biotechnol 2011;29:24-6.
3. DePristo MA, et al. Nat Genet 2011;43:491-8.
4. Wang K, et al. Nucleic Acids Res 2010;38:e164.
5. **Van Loo P, Nordgard SH**, et al. Proc Natl Acad Sci U S A 2010;107:16910-5.
6. Ahdesmaki MJ, et al. PeerJ 2017;5:e3166.
7. Favero F, et al. Ann Oncol 2015;26:64-70.
8. **Liu Y, Sethi NS, Hinoue T, Schneider BG**, et al. Cancer Cell 2018;33:721-735 e8.
9. **Davies H, Glodzik D**, et al. Nat Med 2017;23:517-525.
10. **Golan T, O'Kane GM, Denroche RE**, et al. Gastroenterology 2021;160:2119-2132 e9.
11. Chopra N, et al. Nat Commun 2020;11:2662.
12. Rosenthal R, et al. Genome Biol 2016;17:31.

Author names in bold designate shared co-first authorship

Supplementary Table 1: Germline Mutations Across Cohorts

Progressors

Gene	Chromosome	Position (GRCh37)	Reference Allele	Alternate Allele	HGVS	Clinvar ID	Clinvar Significance
APC	5	112151261	C	T	NM_000038.6:c.904C>T NP_000029.2:p.Arg302Ter	798	Pathogenic
ATM	11	108202611	CTCTAGAATT	C	NM_000051.4:c.7638_7646delTAGAATTTC NP_000042.3:p.Arg2547_Ser2549del	3019	Pathogenic
ATM	11	108172425	C	T	NM_000051.4:c.5228C>T NP_000042.3:p.Thr1743Ile	127403	Pathogenic/ Likely Pathogenic
ATM	11	108121752	CAG	C	NM_000051.4:c.1564_1565delGA NP_000042.3:p.Glu522fs	127340	Pathogenic
ATM	11	108202605	G	A	NM_000051.4:c.7630-1G>A	969935	Likely Pathogenic
ATM	11	108106541	TATCTC	T	NM_000051.4:c.478_482delTCTCA NP_000042.3:p.Ser160fs	185501	Pathogenic
ATM	11	108199877	T	C	NM_000051.4:c.7219T>C NP_000042.3:p.Ser2407Pro	628940	Likely Pathogenic
ATM	11	108129749	C	T	NM_000051.4:c.2413C>T NP_000042.3:p.Arg805Ter	216021	Pathogenic
ATM	11	108160410	A	T	NM_000051.4:c.4318A>T NP_000042.3:p.Lys1440Ter	407482	Pathogenic
ATM	11	108198392	T	TA	NM_000051.4:c.6997dup NP_000042.3:p.Thr2333fs	140818	Pathogenic
ATM	11	108141874	G	A	NM_000051.4:c.2921+1G>A	141182	Pathogenic
ATM	11	108158439	C	A	NM_000051.4:c.4106C>A NP_000042.3:p.Ser1369Ter	379550	Pathogenic
BLM	15	91333877	A	T	NM_000057.4:c.2824-2A>T	371621	Likely Pathogenic
BRCA2	13	32911247	G	T	NM_000059.4:c.2755G>T NP_000050.3:p.Glu919Ter	1418898	Pathogenic
BRCA2	13	32906712	T	G	NM_000059.4:c.1097T>G	266609	Pathogenic

					NP_000050.3:p.Leu366Ter		
BRCA2	13	32900634	A	G	NM_000059.4:c.517-2A>G	51801	Pathogenic/Likely Pathogenic
BRCA2	13	32968863	C	G	NM_000059.4:c.9294C>G NP_000050.3:p.Tyr3098Ter	38229	Pathogenic
BRCA2	13	32914766	CTT	C	NM_000059.4:c.6275_6276delTT NP_000050.3:p.Leu2092fs	9318	Pathogenic/Likely Pathogenic
BRCA2	13	32890665	G	A	NM_000059.4:c.67+1G>A	52160	Pathogenic
BRIP1	17	59793412	G	A	NM_032043.3:c.2392C>T NP_114432.2:p.Arg798Ter	4738	Pathogenic
BRIP1	17	59770857	TTC	T	NM_032043.3:c.2507_2508delTC NP_114432.2:p.Arg836fs	856022	Pathogenic
BRIP1	17	59934523	G	C	NM_032043.3:c.275C>G NP_114432.2:p.Ser92Ter	821772	Pathogenic
BRIP1	17	59938807	C	A	NM_032043.3:c.93+1G>T	141838	Likely Pathogenic
CDH1	16	68849663	G	GT	NM_004360.5:c.1565+2dup	406624	Pathogenic/Likely Pathogenic
CDH1	16	68849664	T	G	NM_004360.5:c.1565+2T>G		Loss of Function
CDKN2A	9	21994233	T	TC	NM_058195.4:c.97dup NP_478102.2:p.Glu33fs	571028	Loss of Function Uncertain Significance
CHEK2	22	29121087	A	G	NM_007194.4:c.470T>C NP_009125.1:p.Ile157Thr	5591	Pathogenic: 3 Likely Pathogenic: 13 Pathogenic, Low Penetrance: 1 Risk Allele:1 Uncertain Significance: 8
CHEK2	22	29091231	C	T	NM_007194.4:c.1260-1G>A	185068	Likely Pathogenic: 1 VUS: 2 Loss of Function
CHEK2	22	29091856	AG	A	NM_007194.4:c.1100delC NP_009125.1:p.Thr367fs	128042	Pathogenic
CHEK2	22	29092945	C	T	NM_007194.4:c.1039G>A NP_009125.1:p.Asp347Asn	182432	Likely Pathogenic: 2 Uncertain Significance: 3
CHEK2	22	29091856	AG	A	NM_007194.4:c.1100delC NP_009125.1:p.Thr367fs	128042	Pathogenic

CHEK2	22	29121087	A	G	NM_007194.4:c.470T>C NP_009125.1:p.Ile157Thr	5591	Pathogenic: 3 Likely Pathogenic: 13 Pathogenic, Low Penetrance: 1 Risk Allele:1 Uncertain Significance: 8
CHEK2	22	29121087	A	G	NM_007194.4:c.470T>C NP_009125.1:p.Ile157Thr	5591	Pathogenic: 3 Likely Pathogenic: 13 Pathogenic, Low Penetrance: 1 Risk Allele:1 Uncertain Significance: 8
CHEK2	22	29130625	G	A	NM_007194.4:c.85C>T NP_009125.1:p.Gln29Ter	187694	Pathogenic
FANCA	16	89818545	C	T	NM_000135.4:c.3066+1G>A	974251	Likely Pathogenic
FANCA	16	89858441	GCCAA	G	NM_000135.4:c.1115_1118delTTGG NP_000126.2:p.Val372fs	3440	Pathogenic
FANCA	16	89828369	G	GA	NM_000135.4:c.2839dup NP_000126.2:p.Ser947fs	188383	Pathogenic
FANCA	16	89846347	G	A	NM_000135.4:c.1645C>T NP_000126.2:p.Gln549Ter	936622	Pathogenic
FANCC	9	98011506	TC	T	NM_000136.3:c.67delG NP_000127.2:p.Asp23fs	12049	Pathogenic
FANCC	9	98011506	TC	T	NM_000136.3:c.67delG NP_000127.2:p.Asp23fs	12049	Pathogenic
FANCM	14	45667921	C	T	NM_020937.4:c.5791C>T NP_065988.1:p.Arg1931Ter	526381	Pathogenic
FANCM	14	45667921	C	T	NM_020937.4:c.5791C>T NP_065988.1:p.Arg1931Ter	526381	Pathogenic
FANCM	14	45667921	C	T	NM_020937.4:c.5791C>T NP_065988.1:p.Arg1931Ter	526381	Pathogenic
FANCM	14	45667921	C	T	NM_020937.4:c.5791C>T NP_065988.1:p.Arg1931Ter	526381	Pathogenic
FANCM	14	45667921	C	T	NM_020937.4:c.5791C>T	526381	Pathogenic

					NP_065988.1:p.Arg1931Ter		
FANCM	14	45667921	C	T	NM_020937.4:c.5791C>T NP_065988.1:p.Arg1931Ter	526381	Pathogenic
FANCM	14	45667921	C	T	NM_020937.4:c.5791C>T NP_065988.1:p.Arg1931Ter	526381	Pathogenic
FH	1	241661227	A	ATTT	NM_000143.4:c.1431_1433dupAAA NP_000134.2:p.Lys477dup	42095	Pathogenic: 7 Likely Pathogenic:4 Uncertain: 5 Likely Benign: 1
FH	1	241667527	G	C	NM_000143.4:c.923C>G NP_000134.2:p.Ala308Gly	392178	Likely Pathogenic
HOXB13	17	46805705	C	T	NM_006361.6:c.251G>A NP_006352.2:p.Gly84Glu	128031	Likely Pathogenic: 1 Uncertain Significance: 1
HOXB13	17	46805705	C	T	NM_006361.6:c.251G>A NP_006352.2:p.Gly84Glu	128031	Likely Pathogenic: 1 Uncertain Significance: 1
HOXB13	17	46805705	C	T	NM_006361.6:c.251G>A NP_006352.2:p.Gly84Glu	128031	Likely Pathogenic: 1 Uncertain Significance: 1
HOXB13	17	46805705	C	T	NM_006361.6:c.251G>A NP_006352.2:p.Gly84Glu	128031	Likely Pathogenic: 1 Uncertain Significance: 1
MRE11	11	94180454	G	A	NM_005591.4:c.1714C>T NP_005582.1:p.Arg572Ter	8784	Pathogenic
MRE11A	11	94180442	G	A	NM_005591.4:c.1726C>T NP_005582.1:p.Arg576Ter	184445	Pathogenic
MRE11A	11	94189489	C	A	NM_005591.4:c.1516G>T NP_005582.1:p.Glu506Ter	140941	Pathogenic: 3 Likely Pathogenic: 1 Uncertain Significance: 1
MSH6	2	48030612	C	T	NM_000179.3:c.3226C>T NP_000170.1:p.Arg1076Cys	89357	Pathogenic/Likely Pathogenic
PALB2	16	23646431	TGA	T	NM_024675.4:c.1436_1437delCT	None	Loss of Function
PMS2	7	6038830	T	G	NM_000535.7:c.614A>C NP_000526.2:p.Gln205Pro	91361	Likely Pathogenic
PMS2	7	6045549	C	A	NM_000535.7:c.137G>T NP_000526.2:p.Ser46Ile	9245	Pathogenic

RAD50	5	131953874	C	T	NM_005732.4:c.3277C>T NP_005723.2:p.Arg1093Ter	5872	Pathogenic
RAD51C	17	56801451	C	T	NM_058216.3:c.955C>T NP_478123.1:p.Arg319Ter	140799	Pathogenic
RECQL	12	21628609	C	T	NM_002907.4:c.1098+1G>A	967538	Loss of Function
RNF43	17	56435519	C	CT	NM_017763.6:c.1617dup		Loss-of-Function
TP53	17	7578555	C	T	NM_000546.6:c.376-1G>A	481003	Pathogenic
TP53	17	7578382	G	T	NM_000546.6:c.548C>T NP_00537.3:pSer183Leu		Loss-of-Function (PMID: 11313981 PMID: 20407015)
TP53	17	7578212	G	A	NM_000546.6:c.637C>T NP_000537.3:p.Arg213Ter	43590	Pathogenic
TP53	17	7577112	C	CACAA	NM_000546.6:c.825_826InsTGTT		Loss-of-Function
TP53	17	7574003	G	A	NM_000546.6:c.1024C>T NP_000537.3:p.Arg342Ter	182970	Pathogenic

MGH Non-Progressors

Gene	Chromosome	Position (GRCh37)	Reference Allele	Alternate Allele	HGVS	Clinvar ID	Clinvar Significance
ATM	11	108186599	A	G	NM_000051.4:c.6056A>G NP_000042.3:p.Tyr2019Cys	230152	Likely Pathogenic: 1 Uncertain Significance: 3
ATM	11	108206686	A	T	NM_000051.4:c.8266A>T NP_000042.3:p.Lys2756Ter	135780	Pathogenic

MGH Wellderly

Gene	Chromosome	Position (GRCh37)	Reference Allele	Alternate Allele	HGVS	Clinvar ID	Clinvar Significance
ATM	11	108179837	A	G	NM_000051.4:c.5763-1050A>G	3021	Pathogenic/Likely Pathogenic
BRCA2	13	32930687	C	T	NM_000059.4:c.7558C>T NP_000050.3:p.Arg2520Ter	52353	Pathogenic
FANCA	16	89813298	T	C	NM_000135.4:c.3349A>G NP_000126.2:p.Arg1117Gly	219752	Pathogenic/Likely Pathogenic
XRCC2	7	152345917	TCA	T	NM_005431.2:c.651_652del NP_005422.1:p.Cys217_Asp218delinsTer	420029	Likely Pathogenic