Accepted Manuscript

Germline Mutations in *PALB2*, *BRCA1*, and *RAD51C*, Which Regulate DNA Recombination Repair, in Patients with Gastric Cancer

Ruta Sahasrabudhe, Paul Lott, Mabel Bohorquez, Ted Toal, Ana P. Estrada, John J. Suarez, Alejandro Brea-Fernández, José Cameselle-Teijeiro, Carla Pinto, Irma Ramos, Alejandra Mantilla, Rodrigo Prieto, Alejandro Corvalan, Enrique Norero, Carolina Alvarez, Teresa Tapia, Pilar Carvallo, Luz M. Gonzalez, Alicia Cock-Rada, Angela Solano, Florencia Neffa, Adriana Della Valle, Chris Yau, Gabriela Soares, Alexander Borowsky, Nan Hu, Li-Ji He, Xiao-You Han, Philip R. Taylor, Alisa M. Goldstein, Javier Torres, Magdalena Echeverry, Clara Ruiz-Ponte, Manuel R. Teixeira, Luis G. Carvajal Carmona

PII: S0016-5085(16)35521-4
DOI: 10.1053/j.gastro.2016.12.010
Reference: YGAST 60870

To appear in: *Gastroenterology*
Accepted Date: 15 December 2016

Please cite this article as: Sahasrabudhe R, Lott P, Bohorquez M, Toal T, Estrada AP, Suarez JJ, Brea-Fernández A, Cameselle-Teijeiro J, Pinto C, Ramos I, Mantilla A, Prieto R, Corvalan A, Norero E, Alvarez C, Tapia T, Carvallo P, Gonzalez LM, Cock-Rada A, Solano A, Neffa F, Della Valle A, Yau C, Soares G, Borowsky A, Hu N, He L-J, Han X-Y, Latin American Gastric Cancer Genetics Collaborative Group, Taylor PR, Goldstein AM, Torres J, Echeverry M, Ruiz-Ponte C, Teixeira MR, Carvajal Carmona LG, Germline Mutations in *PALB2*, *BRCA1*, and *RAD51C*, Which Regulate DNA Recombination Repair, in Patients with Gastric Cancer, *Gastroenterology* (2017), doi: 10.1053/j.gastro.2016.12.010.

Germline Mutations in *PALB2, BRCA1,* and *RAD51C,* Which Regulate DNA Recombination Repair, in Patients with Gastric Cancer

Short Title - HR gene mutations in gastric cancer

Ruta Sahasrabudhe[1*], Paul Lott[1*], Mabel Bohorquez[2], Ted Toal[1], Ana P. Estrada[2], John J. Suarez[2], Alejandro Brea-Fernández[3], José Cameselle-Teijeiro[4], Carla Pinto[5], Irma Ramos[6], Alejandra Mantilla[7], Rodrigo Prieto[2a], Alejandro Corvalan[8], Enrique Norero[8], Carolina Alvarez[9], Teresa Tapia[9], Pilar Carvallo[9], Luz M. Gonzalez[10], Alicia Cock-Rada[10], Angela Solano[11], Florencia Neffa[12], Adriana Della Valle[12] , Chris Yau[13], Gabriela Soares[14], Alexander Borowsky[15], Nan Hu[16], Li-Ji He[17] , Xiao-You Han[18], Latin American Gastric Cancer Genetics Collaborative Group[19], Philip R. Taylor[16], Alisa M. Goldstein[16], Javier Torres[5], Magdalena Echeverry[2], Clara Ruiz-Ponte[3], Manuel R. Teixeira[5,20], Luis G. Carvajal Carmona[1, 21]

[*] These authors contributed equally to the manuscript

[1]Genome Center and Department of Biochemistry and Molecular Medicine, School of Medicine, University of California, Davis, USA

[2]Grupo de Citogenética, Filogenia y Evolución de Poblaciones, Facultades de Ciencias y Facultad de Ciencias de la Salud, Universidad del Tolima, Ibagué, Colombia

[3]Fundación Pública Galega de Medicina Xenómica (FPGMX), Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Genomics Medicine Group, Hospital Clínico, 15706 Santiago de Compostela, University of Santiago de Compostela, Galicia, Spain

[4] Servicio de Anatomía Patológica, Hospital Clínico, Santiago de Compostela, Spain

[5]Department of Genetics, Portuguese Oncology Institute of Porto (IPO-Porto), Porto, Portugal

[6]Unidad de Investigación en Enfermedades Infecciosas, UMAE Pediatria, IMSS, México City, México

[7]Departamento de Patologia, UMAE Oncologia, IMSS, Mexico City, Mexico

[8]Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile

[9]Departamento de Biología Celular y Molecular, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile

[10] Universidad de Antioquia, Medellin, Colombia

[11]INBIOMED, UBA/CONICET y CEMIC, CABA, Argentina

[12]Laboratorio GENIA, Montevideo, Uruguay

[13]Wellcome Trust Centre for Human Genetics, University of Oxford, UK

[14]Centro de Genética Médica Dr. Jacinto Magalhães (CGMJM), Centro Hospitalar do Porto, Porto, Portugal

[15]Department of Pathology, School of Medicine, University of California, Davis, USA

[16]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

[17]Yangcheng Cancer Hospital, Yangcheng, Shanxi, PR China

[18]Shanxi Cancer Hospital, Taiyuan, Shanxi, PR China

[19] Full details listed in the acknowledgements

[20]Institute of Biomedical Sciences (ICBAS), University of Porto, Porto, Portugal

[21]Fundación de Genómica y Genética Molecular, Colombia

[a] In memoriam

Grant support – No Stomach For Cancer Inc and the University of California, Davis provided the principal funding for this study**.** AC received funding from Conicyt-Fondap (15130011), CR-P received funding from Investigacion Sanitaria/FEDER (14/00230, 14/00173) and from the European Commission H2020-PHC (GA nº635290). ME and LGC-C received funding from Fondo de Investigaciones, Universidad del Tolima (Projects 30113, 160120516, 450110, 160114) and from COLCIENCIAS (Project 470115). JT was supported by CONACYT-Fronteras de la Ciencia (clave 773), and Instituto Mexicano del Seguro Social (FIS/IMSS/PROT/PRIO/13/027). This work was also supported in part by

the Intramural Research Program of the National Cancer Institute, the National Institutes of Health, the Division of Cancer Epidemiology and Genetics. LGC-C received funding from the V Foundation from Cancer Research and from the National Cancer Institute (K12CA138464 and R21CA199631) of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Abbreviations **–** GC=gastric cancer, HDGC=hereditary diffuse gastric cancer, WES=whole exome sequencing, LOH=loss of heterozygosity, BC=breast cancer, R=homologous recombination

Corresponding Author

Luis G. Carvajal-Carmona, PhD

Genome Center and Department of Biochemistry and Molecular Medicine

School of Medicine, University of California, Davis

UC Davis Genome Center, 451 Health Sciences Drive

Davis, California 95616, USA

Phone: +1 530-752-9654, Fax: +1 530-754-9658

Email: lgcarvajal@ucdavis.edu

Disclosure: The authors have no conflict of interest to disclose.

Author contribution: Study concept and design = RS, MT, LGCC; Acquisition of data = RS, PL, MB, APE, JJS, ABF, CP, IR, AM, RP, JS, AC, EN, CA, TT, PC, LMG, ACR, AS, FN, ADV, CY, GS, AB, NH, LJH, XYH, PRT, AMG, JT, ME, CRP, MT, LGCC; Analysis and interpretation of data = PL, RS, TT, LGCC; Drafting of the manuscript = RS, LGCC; Critical revision of the manuscript for important intellectual content = RS, PL, TT, AG, JT, ME, JCT, AB, CRP, MT, LGCC; Statistical analysis = RS, PL,

TT, LGCC; Obtained funding= LGCC; Administrative, technical, or material support = RS, PL, MT, JT, CRP, ME, LGCC; Study supervision = RS, PL, MT, JT, CRP, ME, LGCC

Abstract: Up to 10% of cases of gastric cancer are familial, but so far, only mutations in *CDH1* have been associated with gastric cancer risk. To identify genetic variants that affect risk for gastric cancer, we collected blood samples from 28 patients with hereditary diffuse gastric cancer (HDGC) not associated with mutations in *CDH1* and performed whole-exome sequence analysis. We then analyzed sequences of candidate genes in 333 independent HDGC and non-HDGC cases. We identified 11 cases with mutations in *PALB2*, *BRCA1*, or *RAD51C* genes, which regulate homologous DNA recombination. We found these mutations in 2 of 31 patients with HDGC (6.5%) and 9 of 331 patients with sporadic gastric cancer (2.8%). Most of these mutations had been previously associated with other types of tumors and partially co-segregated with gastric cancer in our study. Tumors that developed in patients with these mutations had a mutation signature associated with somatic homologous recombination deficiency. Our findings indicate that defects in homologous recombination increase risk for gastric cancer.

KEY WORDS: stomach, tumor, WES, interaction

Worldwide, gastric cancer (GC) is the fifth most commonly diagnosed malignancy and the third cause of cancer-related deaths [1]. Up to 10% of cases show familial clustering, suggesting a genetic basis [2]. *CDH1* mutations are a known cause of hereditary diffuse gastric cancer (HDGC), explaining ~ 40% of cases [3,4], but the genetics of non-HDGC remain largely unknown. To identify novel GC genes, we analyzed *CDH1* mutation-negative HDGC cases using whole exome sequencing (WES) followed by candidate gene targeted analyses in independent HDGC and non-HDGC cases.

WES of 28 *CDH1*-negative European HDGC cases identified three with candidate causal variants (Table 1): nonsense (p.Arg414Ter) and splice site (c.3201+1G>T) *PALB2* mutations, and a nonsense *RAD51C* (p.Arg237Ter) mutation. No deleterious mutations were seen in other known cancer genes (Supplementary methods). *PALB2* and *RAD51C* are both critical in homologous recombination (HR), a major DNA repair pathway [5]. Both of the above *PALB2* mutations have been previously reported as pathogenic in breast cancer (BC) families [6] and *RAD51C* p.Arg237Ter is reported as pathogenic in ClinVar[7].

We then performed targeted sequencing of *PALB2* and *RAD51C,* their interaction partners *BRCA1/2* and *CDH1* in 173 additional Latin American GC cases. Based upon enrichment of HR mutations in our discovery cohort and a recent report showing multiple intestinal, diffuse and mixed histology gastric tumors with a somatic HR deficiency signature [8] , our validation cohort included both HDGC and non-HDGC cases of diffuse and non-diffuse histology (Supplementary methods). Targeted sequencing identified four additional mutation carriers: two sharing a known Hispanic *BRCA1* founder mutation (p.Gln1111Asnfs) [9] and two with novel *PALB2* mutations (p.Pro918Gln and p.Lys628_Cys630del) with predicted deleterious effects. Residue Pro918 falls in the PALB2 WD40 domain, which mediates interactions with BRCA2, RAD51 and RAD51C, whereas Lys628-Cys630 resides in the binding domain of MRG15, a transcription regulator and whose PALB2 interaction is required for homology directed DNA double-strand break repair indicating potential pathogenicity of these two novel mutations [10, 11].

In a third phase of the study, we genotyped all six *PALB2, RAD51C* and *BRCA1* mutations described above plus four known Hispanic *BRCA1/2* founder mutations (Supplementary methods) in 160 independent Latin American non-HDGC cases and found three additional mutation carriers, one with a *BRCA1* mutation (p.Gly559Valfs) and two with *PALB2* mutations (p.Lys628_Cys630del and p.Arg414Ter, Table 1). Interestingly, during the preparation of this manuscript, our clinic-based Portuguese collaborator (MT), identified one additional GC case (GM037589) with *PALB2* p.Arg414Ter. None of the seven *PALB2, RAD51C* and *BRCA1* mutations, detected in 11 unrelated Caucasian and Latin American cases, was detected in 1,170 population-matched controls (see mutation details in Supplementary Table 1).

Clinical details of our mutation carriers are given in Table 1. Most of them had diffuse histology, two had HDGC syndrome (CG-05 and GM022584) and one reported history of hereditary breast and ovarian cancer (HBOC, case CG-36, not shown). These mutation carriers were predominantly non-smokers and/or negative for *Helicobacter pylori* infection (Table 1), which suggest that GC risk in most of these cases was not driven by these two known environmental risk factors[12].

To obtain additional evidence of the causality of our HR gene mutations, we carried out loss of heterozygosity (LOH), mutational signature and co-segregation analyses in available samples from tumors and relatives. For LOH and mutational signatures, we performed WES in four available tumor samples from three *PALB2* (CG-12/p.Arg414Ter, CG-028/p.Lys628_Cys630del and 3CG-103/p.Pro918Gln) and *RAD51C* mutation carriers (Table 1). We found no LOH or compound heterozygosity in these tumor samples (not shown). Interestingly, when we analyzed the somatic WES data for mutational signatures, we found that all four tumors were enriched for a signature indicative of HR defects [13, 14], providing evidence for the causality of these mutations (Supplementary methods, Supplementary Figures 1 and 2).

Figure 1 shows available pedigrees from mutation carriers. Case 3CG-103 and her daughter were both diagnosed with GC and carried the *PALB2* p.Pro918Gln mutation (Figure 1A). GM037589, a *PALB2* p.Arg414Ter carrier, developed GC and BC and had a sister diagnosed with ovarian and endometrial

cancer who also carried *PALB2* p.Arg414Ter (Figure 1B). The *RAD51C* p.Arg237Ter carrier's son died of colon cancer but did not carry the mutation (Figure 1C). We found that GC was the predominantly diagnosed malignancy among unavailable relatives of these carriers (Figures 1A-1D). Although we did not have access to samples from relatives of the *PALB2* p.Lys628_Cys630del carriers, our local collaborators found this mutation co-segregating in an unrelated breast cancer family (unpublished). Albeit limited, our co-segregation data partially support GC causality of *PALB2* mutations. The *RAD51C* co-segregation data is however inconclusive but the presence of a strong HR signature in the gastric tumor (see above) of this mutation carrier warrants further studies on *RAD51C* as a candidate GC gene.

In summary, our study identified eleven cases with mutations in *PALB2*, *BRCA1* and *RAD51C*, three closely-related HR genes. Some of these mutations are known to be pathogenic in other cancer types. Out of 362 cases analyzed, 6.45% of the HDGC cases (2 out of 31) and 2.7% (9 out of 331) of non-HDGC cases had *PALB2*, *BRCA1* or *RAD51C* mutations, suggesting that HR genes play a role in GC risk. Our data also provide evidence of a germline basis for the recently reported HR mutational signature in gastric tumors and strengthens the evidence for a causal role of these genes, specifically *PALB2,* in GC, as previously observed [4, 15]. Future larger studies are needed to definitively assign causality and understand the penetrance and prevalence of HR gene mutations in GC and to further understand if and why some individuals from HBOC families with HR gene mutations develop GC. Further characterizations of the GC histology in HR gene mutation carriers are also needed as we found instances where the same mutation was found in cases with different histologies (CG-12 and CG-008 with *PALB2* p.Arg414ter and CG-039 and CG-028 with PALB2 p.Lys628_Cys630del, Table 1). *CDH1* mutation negative families might benefit from HR gene testing and increased endoscopic surveillance and targeted therapies, such as PARP inhibitors [8].

ACKNOWLEDGMENTS

We thank all patients and their families for participating in the study and to multiple clinicians and research nurses who helped with patient recruitment. We are thankful to No Stomach For Cancer Inc. and the University of California Davis for providing principal financial support to the study. We are grateful to former Carvajal-Carmona laboratory members John Williamson, Julian Halmai, Anna Marie Tuazon and Cathy Wang for their support during the study. We are also grateful to Drs. Yuichi Shiraishi and Matthew Stephens for their support and stimulating discussions about mutational signature analyses.

Members of the Latin American Gastric Cancer Genetics Collaborative Group: Magdalena Echeverry[1], Mabel Bohorquez[1], Rodrigo Prieto[1], John Suarez[1], Gilbert Mateus[2], Maria Mercedes Bravo[3], Fernando Bolaños[4], Alejandro Vélez[5], Alejandro Corvalan[6], Pilar Carvallo[7], Javier Torres[8], Luis Carvajal-Carmona[9].

[1] Universidad del Tolima, Ibagué, Colombia; [2] Hospital Federico Lleras Acosta, Ibagué, Colombia; [3] Instituto Nacional de Cancerología, Bogotá, D.C, Colombia; [4]Hospital Hernando Moncaleano Perdomo, Neiva, Colombia; [5] Hospital Pablo Tobón Uribe, Medellín, Colombia; [6]Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile; [7]Departamento de Biología Celular y Molecular, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile, [8] Unidad de Investigación en Enfermedades Infecciosas, UMAE Pediatria, IMSS, México City, México; [8] University of California, Davis.

REFERENCES

1. Ferlay J. Int J Cancer 2015;136:E359-86.
2. La Vecchia C. Cancer 1992;70:50-5.
3. Guilford P. Nature 1998;392:402-5.
4. Hansford S. JAMA Oncol 2015;1:23-32.
5. Prakash R. Cold Spring Harb Perspect Biol 2015;7:a016600.
6. Antoniou AC. N Engl J Med 2014;371:1651-2.

7. Landrum MJ. Nucleic Acids Res 2016;44:D862-8.

8. Alexandrov LB. Nat Commun 2015;6:8683.

9. Torres D. Breast Cancer Res Treat 2007;103:225-32.

10. Park JY. Oncogene 2014;33:4803-12.

11. Sy SM. J Biol Chem 2009;284:21127-31.

12. Cover TL. Gut Microbes 2013;4:482-93.

13. Alexandrov LB. Nature 2013;500:415-21.

14. Shiraishi Y. PLoS Genet 2015;11:e1005657.

15. **Lu C**, **Xie M**, **Wendl MC**, et al. Nat Commun 2015;6:10086.

Author names in bold designate shared co-first authorship

FIGURE LEGENDS

Figure 1: Available pedigrees of mutation carriers.

Table 1: Details of clinical information of the mutation carriers

Mutation details	ID	Age of onset	Sex	Histology	Satisfied HDGC criteria ?	*Helicobacter Pylori* infection	History of smoking
PALB2 c.1240C>T, p.Arg414Ter	CG-12[a*]	69	M	Intestinal	No	NA	NA
	CG-008[c]	48	F	Diffuse	NA	NA	Yes
	GM037589	46	F	NA	No	Negative	No
PALB2 c.3201+1G>T	CG-05[a]	50	M	Diffuse	Yes	Negative	No
PALB2 c. 1882_1890delGCAGGACTT, p.Lys628_Cys630del	CG-039[b]	47	F	Diffuse	NA	Negative	No
	CG-028[c*]	81	M	Intestinal	NA	Negative	Yes
PALB2 c.2753C>A, p.Pro918Gln	3CG-103[b*]	79	F	Mixed	No	Negative	Yes
BRCA1 c.3331_3334delCAAG, p.Gln1111Asnfs	CG-036[b]	67	F	Diffuse	No	NA	No
	CG-059[b]	54	M	Diffuse	No	NA	No
BRCA1 c.1674delA, p.Gly559Valfs	CG-001[c]	65	M	NA	No	Positive	Yes
RAD51C c.709 C>T, p.Arg237Ter	GM022584[a*]	73	M	Diffuse	Yes	Negative	No

Identified by: [a] WES, [b] targeted sequencing or [c] genotyping. *:LOH and mutational signature analyzed. NA: Not available

SUPPLEMENTARY MATERIALS AND METHODS

Phase I - Variant discovery by whole-exome sequencing (WES)

Patient recruitment: For whole exome sequencing (WES) analysis, we included twenty-eight GC cases (and six relatives from four different families) with Hereditary Diffuse Gastric Cancer (HDGC was defined according to the published guidelines [1]) recruited in the Portuguese Oncology Institute (University of Porto, Portugal) and in the Genomic Medicine group (Santiago de Compostela, Spain). Sample collection was undertaken with informed consent and ethical review board approval of the corresponding institution, in accordance with the tenets of the Declaration of Helsinki. All of these 28 index HDGC cases tested negative for *CDH1* mutations at clinical laboratories in these two Portuguese and Spanish institutions. The average age of HDGC index cases was 48.2 years (standard deviation=13.2 years). 15 of these cases were males and 13 were females. Interestingly, one of these patients (CG-12), who was initially included as an HDGC case, on histological re-examination by two independent surgical pathologists (JC-T and AB) was re-classified as having intestinal histology. This case was therefore reclassified as a non-HDGC in our study.

WES library preparation: Samples were prepared for WES using Agilent SureSelect XT2 protocol. Briefly, up to 1µg of DNA was sheared using Covaris E220 sonicator. Fragments were end-repaired, A-tailed and Illumina-compatible adaptors were ligated at the ends. The fragments were then enriched using PCR. Eight multiplexed samples were hybridized to the bait set, washed and captured fragments were amplified by PCR. Samples were then sequenced on an Illumina HiSeq2000 sequencer with 100PE sequencing.

WES data analysis: For data analysis, publically available tools as well as custom shell scripts were used. Raw data was trimmed for adaptors and sequence quality and then aligned to the human reference genome GRCh37 with decoy sequences using BWA-mem v0.7.12 [2-4]. For WES, duplicates were removed with Picard v1.129 (http://picard.sourceforge.net). BAM files were locally realigned using GATK IndelRealigner v3.3 and recalibration of the quality scores was performed using GATK BaseRecalibrator v3.3 [5]. Multiple callers were used to call variants: GATK HaplotypeCaller non-joint v3.3 [6] Freebayes v0.9.14-17 [7], SNVER [8], Varscan v2.3.7 [9], Samtools mpileup v1.2.[10] Calls were filtered based upon: coverage >=10, number of reads supporting variant >=5, minimum variant frequency >=0.20, minimum frequency of variant reads present on opposite strand >0.10, minimum average read quality >=22. Variants were annotated using Annovar [11]. In addition, SNP and INDEL calling was performed using GATK HaplotypeCaller joint genotyping. Calling, variant filtering, and variant score recalibration were performed using GATK v3.3 Best Practices [6, 12]. Variants called by at least two different callers were considered for further analysis. To select the most informative SNVs, filtering of the initial data was performed to exclude all synonymous SNVs, SNVs that map to pseudo-genes, repeated regions, segmental duplications and "dispensable" genes. The remaining protein sequence-altering variants were subjected to frequency filtering using data from publicly available datasets such as the Exome Variant Server, the UK10K study, dbSNP and the 1000 Genomes Project to exclude variants with >1% MAF. Of the remaining 7781 variants, SNVs in known cancer predisposition genes [13] were identified (N=45). Of those, 2 SNVs were protein- truncating (*PALB2*: p.Arg414Ter and *RAD51C*: p.Arg237Ter) with predicted deleterious amino acid substitutions (based on Polyphen, SIFT, MutationAssessor and MutationTaster) and one variant resulted in disruption of a splice site. For the above three candidate causal variants, pileups were visually inspected in IGV [14]. No truncating, deleterious mutations were seen in any other cancer genes.

Phase II- Candidate gene validation by targeted sequencing

Patient recruitment: For WES replication by targeted sequencing; we included 14 Chilean GC cases recruited in a local cancer clinic, four of which satisfied HDGC criteria. Thus, our study included a total of 31 HDGC index cases in the discovery (n=27) and validation (n=4) phases. Out of the remaining 10 Chilean non-HDGC cases, five had intestinal GC and five were of unknown histology. For targeted sequencing, we also included additional GC cases from Colombia (N=90) and Mexico (N=69) out of which 104 cases had diffuse histology, 42 had mixed histology, one had intestinal histology, and in 12 cases histology was unknown. Together, 53 cases had early onset GC (<50 years). Chilean cases were recruited in Dr. Sótero del Río Hospital, and Clinical Hospital Pontificia Universidad Católica (both in Santiago, Chile). The Ethics Committees of Dr. Sótero del Río Hospital and Clinical Hospital Pontificia Universidad Católica de Chile approved the recruitment protocols. Colombian cases for validation phases II and III (see below) were recruited from a multi-center study in Colombia and in the Instituto Mexicano de Seguro Social (IMSS) following protocols approved by University of Tolima (Ibague, Colombia) and IMSS National Council for Research on Health (Mexico City, Mexico).

Targeted sequencing library preparation and data analysis: ~350bp PCR amplicons covering the entire coding regions of *BRCA1, BRCA2, CDH1, PALB2* and *RAD51C* were amplified from 50ng of genomic DNA using Fluidigm Access array system and libraries were sequenced on a MiSeq platform with 250PE reads. Sequence data analysis was performed with a bioinformatics pipeline similar to the one described for WES above.

Phase III - Mutation validation by genotyping

Patient recruitment and genotyping: For genotyping, we included 160 non-HDGC cases from Colombia (N=93) and Mexico (N=67) that included 24 cases with diffuse histology, 117 with intestinal histology, 8 with mixed histology and 11 with unknown of histology. All six sequence-

identified *PALB2, RAD51C* and *BRCA1* mutations in Phase I and II (see above and body of the manuscript), as well as four additional known Hispanic *BRCA1/2* founder mutations (c.5123C>T /p.Ala1708Val and c.1674delA/p.Gly559Valfs in *BRCA1* and c.2808_2811delTAAA/p.Ala938Profs and c.4889C>G/p.Ser1630Ter in *BRCA2*) were included in Phase III of genotyping. Genotyping of these 10 mutations was performed using competitive allele-specific PCR using KASP assays (LGC genomics), following manufacturer's guidelines.

Sanger sequencing: All mutations identified using WES, targeted sequencing and genotyping in Phases I, II and III were verified using Sanger sequencing. Details of the sequencing primers are as follows: *PALB*_p.Arg414Ter - Forward: TGAACTTGGTTGTCCTGTGC, Reverse: TGACACTCTTGATGGCAGGA. *PALB2*_c.3201+1G, Forward: TTTGCCCTCAGGTCCTACAG, Reverse: TGGTTTGTTGGAAGAATGTGA, *PALB2*_p.Lys628_Cys630del, Forward: CCTCCATTTCTGTATCCATGC, Reverse: AAGAGGATTCCCTTTCTTGGA, *PALB2*_p.Pro918Gln – Forward : CCAGCTGACAGAGACAAAGATG, Reverse: TCTGAGCCTTCAAATGATGAAA, *BRCA1*_p.Gln1111Asnf – Forward: GGGTGAAAGGGCTAGGACTC, Reverse: CAGAGGGCCAAAATTGAATG, *BRCA1*_p.Gly559Valfs – Forward: ACCAAACGGAGCAGAATGGT, Reverse: GCAATTCAGTACAATTAGGTGGGC, *RAD51C*_p.Arg237Ter - Forward: GGTCCCTGCTCTCTTGGAGA, Reverse: ACCAACCAAACGTAACTTTACTCAA.

WES of tumor DNA for loss of heterozygosity (LOH) and mutational signature analysis

DNA was extracted, using a Qiagen tissue kit, from formalin fixed paraffin embedded (FFPE) tumor tissue samples from four cases: CG-12 (*PALB2* nonsense mutation carrier), 3CG-103 (*PALB2* missense mutation carrier), CG-028 (*PALB2* in-frame deletion carrier) and GM022584 (*RAD51C* nonsense mutation carrier). WES was performed using KAPA and Agilent SureSelect

XT kits following manufacturer's guidelines. Samples were sequenced on a HiSeq4000 using PE150 sequencing. Sequence data analysis was performed using GATK best practices as described above and somatic variants were called with GATK MuTect2 [15].

Mutational signature analysis: Mutational signature analysis in somatic tissue is a recent field that is undergoing active development, improvement and statistical grounding. The first general signature model for mutation signature analysis was developed by Alexandrov *et al* [16] and was used to analyze the TCGA dataset, leading to the first defined mutational signature resulting from defects in homologous recombinational DNA repair (HR), annotated as 'Signature 3' [17]. A conceptually different theoretical model of mutation signatures was developed by Shiraishi *et al* [18] with an accompanying computational framework called pmsignature. This model pools all mutations from all the samples and seeks signatures that occur relatively frequently in the mutation pool. The output from the analysis is a matrix of estimated signature parameters defining the signatures, and a membership weight matrix that estimates the relative contribution of each signature to the mutations in each sample. The number of signatures that is found, K, is a parameter that must be specified *a priori*. The Shiraishi signature model differs from the earlier model in that it assumes independence of the adjacent bases, so the number of parameters with a single surrounding base is far fewer than with the Alexandrov model, leading to more statistically stable parameter estimates. We combined the mutations of our four tumor samples with 40 TCGA GC whole exomes to increase the power to detect common GC signatures and to provide positive and negative HR signature controls. Of the 40 samples, 20 were selected from the 27 samples with non-zero value for 'Signature 3' and 20 were selected from the remaining samples with a zero value [19]. We configured the Shiraishi framework to use five bases of total context (the mutated base plus two bases upstream and two bases downstream) and to include the transcription strand as a mutation feature. The mutation signature analysis was done using the R language [20]. In order to detect an HR signature, we first determined which of the 27 Shiraishi signatures was

most similar to the Alexandrov *et al* 'signature 3' by using both Frobenius and cosine similarity measures. Heatmaps depicting the Frobenius and cosine similarity of each of the 27 Shiraishi cancer signatures to each of the 30 Alexandrov (COSMIC) cancer signatures are shown in Supplementary Figure 1A and 1B respectively. For Frobenius similarity, Shiraishi signatures 16, 23, 24, and 25 all have similarity >= 0.7 to COSMIC signature 3. For cosine similarity, Shiraishi signatures 16, 23, and 25 all have similarity > 0.7 to COSMIC signature 3. We have designated Shiraishi signatures 16 and 23-25 as HR signatures on heatmaps that show Shiraishi signatures. Knowing which Shiraishi signatures correspond to an HR signature, we proceeded to determine which signature, if any, of K signatures produced by our analysis, are similar to one of those Shiraishi HR signatures. We used Frobenius similarity in that case, since both signatures being compared are Shiraishi signatures, and the comparison is more reliable than the Alexandrov-Shiraishi comparison. Frobenius similarity showed that, at K=3, signature #1 [noted as 1(HR)] was most similar to the Shiraishi HR signatures 16, 23, and 25 (full analysis, Supplementary Figure 1C). Tumor DNA from our study samples was derived from FFPE tissue, and was thus expected to have a higher percentage of C: G>T: A mutations. Therefore we analyzed mutational signatures after removing C:G>T:A from our study samples as well as from control samples (restricted analysis). Similar to the full analysis, we first identified signatures with high Frobenius similarity to Shiraishi HR signatures, using K=3 (Supplementary Figure 1D). After optimizing the method, we proceeded to determine whether an HR signature was demonstrated by the four study samples where somatic WES data was available (see above). As shown in Supplementary Figure 2, our study samples as well as the TCGA positive controls, at K=3, in full and restricted analysis have a significantly higher relative contribution or membership weight for the HR signature compared to the negative controls. Interestingly, another hallmark of somatic HR deficiency is a high frequency of large indels [16, 19]. Consistently, similar to TCGA HR-positive controls, the mean deletion length found in the tumors from our four *PALB2/RAD51C* mutation carriers was higher than in TCGA nonHR GC cases (31.6 bp vs. 15.4 bp, $P= 3 \times 10^{-7}$).

REFERENCES

1. van der Post RS. J Med Genet 2015;52:361-374.
2. Buffalo V. (Version 0.994Beta) [Software]. Available at https://github.com/vsbuffalo/scythe. 2011.
3. Joshi N. (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle. 2011.
4. Li H. Preprint at arXiv:1303.3997v2 [q-bio.GN] 2013.
5. McKenna A. Genome Res 2010;20:1297-303.
6. DePristo MA. Nat Genet 2011;43:491-8.
7. Garrison EM. Preprint at arXiv:1207.3907v2 [q-bio.GN] 2012.
8. Wei Z. Nucleic Acids Res 2011;39:e132.
9. Koboldt DC. Genome Res 2012;22:568-76.
10. **Li H**, **Handsaker B**. Bioinformatics 2009;25:2078-9.
11. Wang K. Nucleic Acids Res 2010;38:e164.
12. Van der Auwera GA. Curr Protoc Bioinformatics 2013;11:11 10 1-11 10 33.
13. Rahman N. Nature 2014;505:302-8.
14. Robinson JT. Nat Biotechnol 2011;29:24-6.
15. Cibulskis K. Nat Biotechnol 2013;31:213-9.
16. Alexandrov LB. Cell Rep 2013;3:246-59.
17. Alexandrov LB. Nature 2013;500:415-21.
18. Shiraishi Y. PLoS Genet 2015;11:e1005657.
19. Alexandrov LB. Nat Commun 2015;6:8683.
20. Team RC. URL http://www.R-project.org/. 2016.

Author names in bold designate shared co-first authorship

SUPPLEMENTARY MATERIALS AND METHODS

Phase I - Variant discovery by whole-exome sequencing (WES)

Patient recruitment: For whole exome sequencing (WES) analysis, we included twenty-eight GC cases (and six relatives from four different families) with Hereditary Diffuse Gastric Cancer (HDGC was defined according to the published guidelines [1]) recruited in the Portuguese Oncology Institute (University of Porto, Portugal) and in the Genomic Medicine group (Santiago de Compostela, Spain). Sample collection was undertaken with informed consent and ethical review board approval of the corresponding institution, in accordance with the tenets of the Declaration of Helsinki. All of these 28 index HDGC cases tested negative for *CDH1* mutations at clinical laboratories in these two Portuguese and Spanish institutions. The average age of HDGC index cases was 48.2 years (standard deviation=13.2 years). 15 of these cases were males and 13 were females. Interestingly, one of these patients (CG-12), who was initially included as an HDGC case, on histological re-examination by two independent surgical pathologists (JC-T and AB) was re-classified as having intestinal histology. This case was therefore reclassified as a non-HDGC in our study.

WES library preparation: Samples were prepared for WES using Agilent SureSelect XT2 protocol. Briefly, up to 1µg of DNA was sheared using Covaris E220 sonicator. Fragments were end-repaired, A-tailed and Illumina-compatible adaptors were ligated at the ends. The fragments were then enriched using PCR. Eight multiplexed samples were hybridized to the bait set, washed and captured fragments were amplified by PCR. Samples were then sequenced on an Illumina HiSeq2000 sequencer with 100PE sequencing.

WES data analysis: For data analysis, publically available tools as well as custom shell scripts were used. Raw data was trimmed for adaptors and sequence quality and then aligned to the human reference genome GRCh37 with decoy sequences using BWA-mem v0.7.12 [2-4]. For WES, duplicates were removed with Picard v1.129 (http://picard.sourceforge.net). BAM files were locally realigned using GATK IndelRealigner v3.3 and recalibration of the quality scores was performed using GATK BaseRecalibrator v3.3 [5]. Multiple callers were used to call variants: GATK HaplotypeCaller non-joint v3.3 [6] Freebayes v0.9.14-17 [7], SNVER [8], Varscan v2.3.7 [9], Samtools mpileup v1.2.[10] Calls were filtered based upon: coverage >=10, number of reads supporting variant >=5, minimum variant frequency >=0.20, minimum frequency of variant reads present on opposite strand >0.10, minimum average read quality >=22. Variants were annotated using Annovar [11]. In addition, SNP and INDEL calling was performed using GATK HaplotypeCaller joint genotyping. Calling, variant filtering, and variant score recalibration were performed using GATK v3.3 Best Practices [6, 12]. Variants called by at least two different callers were considered for further analysis. To select the most informative SNVs, filtering of the initial data was performed to exclude all synonymous SNVs, SNVs that map to pseudo-genes, repeated regions, segmental duplications and "dispensable" genes. The remaining protein sequence-altering variants were subjected to frequency filtering using data from publicly available datasets such as the Exome Variant Server, the UK10K study, dbSNP and the 1000 Genomes Project to exclude variants with >1% MAF. Of the remaining 7781 variants, SNVs in known cancer predisposition genes [13] were identified (N=45). Of those, 2 SNVs were protein- truncating (*PALB2*: p.Arg414Ter and *RAD51C*: p.Arg237Ter) with predicted deleterious amino acid substitutions (based on Polyphen, SIFT, MutationAssessor and MutationTaster) and one variant resulted in disruption of a splice site. For the above three candidate causal variants, pileups were visually inspected in IGV [14]. No truncating, deleterious mutations were seen in any other cancer genes.

Phase II- Candidate gene validation by targeted sequencing

Patient recruitment: For WES replication by targeted sequencing; we included 14 Chilean GC cases recruited in a local cancer clinic, four of which satisfied HDGC criteria. Thus, our study included a total of 31 HDGC index cases in the discovery (n=27) and validation (n=4) phases. Out of the remaining 10 Chilean non-HDGC cases, five had intestinal GC and five were of unknown histology. For targeted sequencing, we also included additional GC cases from Colombia (N=90) and Mexico (N=69) out of which 104 cases had diffuse histology, 42 had mixed histology, one had intestinal histology, and in 12 cases histology was unknown. Together, 53 cases had early onset GC (<50 years). Chilean cases were recruited in Dr. Sótero del Río Hospital, and Clinical Hospital Pontificia Universidad Católica (both in Santiago, Chile). The Ethics Committees of Dr. Sótero del Río Hospital and Clinical Hospital Pontificia Universidad Católica de Chile approved the recruitment protocols. Colombian cases for validation phases II and III (see below) were recruited from a multi-center study in Colombia and in the Instituto Mexicano de Seguro Social (IMSS) following protocols approved by University of Tolima (Ibague, Colombia) and IMSS National Council for Research on Health (Mexico City, Mexico).

Targeted sequencing library preparation and data analysis: ~350bp PCR amplicons covering the entire coding regions of *BRCA1, BRCA2, CDH1, PALB2* and *RAD51C* were amplified from 50ng of genomic DNA using Fluidigm Access array system and libraries were sequenced on a MiSeq platform with 250PE reads. Sequence data analysis was performed with a bioinformatics pipeline similar to the one described for WES above.

Phase III - Mutation validation by genotyping

Patient recruitment and genotyping: For genotyping, we included 160 non-HDGC cases from Colombia (N=93) and Mexico (N=67) that included 24 cases with diffuse histology, 117 with intestinal histology, 8 with mixed histology and 11 with unknown of histology. All six sequence-

identified *PALB2, RAD51C* and *BRCA1* mutations in Phase I and II (see above and body of the manuscript), as well as four additional known Hispanic *BRCA1/2* founder mutations (c.5123C>T /p.Ala1708Val and c.1674delA/p.Gly559Valfs in *BRCA1* and c.2808_2811delTAAA/p.Ala938Profs and c.4889C>G/p.Ser1630Ter in *BRCA2*) were included in Phase III of genotyping. Genotyping of these 10 mutations was performed using competitive allele-specific PCR using KASP assays (LGC genomics), following manufacturer's guidelines.

Sanger sequencing: All mutations identified using WES, targeted sequencing and genotyping in Phases I, II and III were verified using Sanger sequencing. Details of the sequencing primers are as follows: *PALB*_p.Arg414Ter - Forward: TGAACTTGGTTGTCCTGTGC, Reverse: TGACACTCTTGATGGCAGGA. *PALB2*_c.3201+1G, Forward: TTTGCCCTCAGGTCCTACAG, Reverse: TGGTTTGTTGGAAGAATGTGA, *PALB2*_p.Lys628_Cys630del, Forward: CCTCCATTTCTGTATCCATGC, Reverse: AAGAGGATTCCCTTTCTTGGA, *PALB2*_p.Pro918Gln – Forward : CCAGCTGACAGAGACAAAGATG, Reverse: TCTGAGCCTTCAAATGATGAAA, *BRCA1*_p.Gln1111Asnf – Forward: GGGTGAAAGGGCTAGGACTC, Reverse: CAGAGGGCCAAAATTGAATG, *BRCA1*_p.Gly559Valfs – Forward: ACCAAACGGAGCAGAATGGT, Reverse: GCAATTCAGTACAATTAGGTGGGC, *RAD51C*_p.Arg237Ter - Forward: GGTCCCTGCTCTCTTGGAGA, Reverse: ACCAACCAAACGTAACTTTACTCAA.

WES of tumor DNA for loss of heterozygosity (LOH) and mutational signature analysis

DNA was extracted, using a Qiagen tissue kit, from formalin fixed paraffin embedded (FFPE) tumor tissue samples from four cases: CG-12 (*PALB2* nonsense mutation carrier), 3CG-103 (*PALB2* missense mutation carrier), CG-028 (*PALB2* in-frame deletion carrier) and GM022584 (*RAD51C* nonsense mutation carrier). WES was performed using KAPA and Agilent SureSelect

XT kits following manufacturer's guidelines. Samples were sequenced on a HiSeq4000 using PE150 sequencing. Sequence data analysis was performed using GATK best practices as described above and somatic variants were called with GATK MuTect2 [15].

Mutational signature analysis: Mutational signature analysis in somatic tissue is a recent field that is undergoing active development, improvement and statistical grounding. The first general signature model for mutation signature analysis was developed by Alexandrov *et al* [16] and was used to analyze the TCGA dataset, leading to the first defined mutational signature resulting from defects in homologous recombinational DNA repair (HR), annotated as 'Signature 3' [17]. A conceptually different theoretical model of mutation signatures was developed by Shiraishi *et al* [18] with an accompanying computational framework called pmsignature. This model pools all mutations from all the samples and seeks signatures that occur relatively frequently in the mutation pool. The output from the analysis is a matrix of estimated signature parameters defining the signatures, and a membership weight matrix that estimates the relative contribution of each signature to the mutations in each sample. The number of signatures that is found, K, is a parameter that must be specified *a priori*. The Shiraishi signature model differs from the earlier model in that it assumes independence of the adjacent bases, so the number of parameters with a single surrounding base is far fewer than with the Alexandrov model, leading to more statistically stable parameter estimates. We combined the mutations of our four tumor samples with 40 TCGA GC whole exomes to increase the power to detect common GC signatures and to provide positive and negative HR signature controls. Of the 40 samples, 20 were selected from the 27 samples with non-zero value for 'Signature 3' and 20 were selected from the remaining samples with a zero value [19]. We configured the Shiraishi framework to use five bases of total context (the mutated base plus two bases upstream and two bases downstream) and to include the transcription strand as a mutation feature. The mutation signature analysis was done using the R language [20]. In order to detect an HR signature, we first determined which of the 27 Shiraishi signatures was

most similar to the Alexandrov *et al* 'signature 3' by using both Frobenius and cosine similarity measures. Heatmaps depicting the Frobenius and cosine similarity of each of the 27 Shiraishi cancer signatures to each of the 30 Alexandrov (COSMIC) cancer signatures are shown in Supplementary Figure 1A and 1B respectively. For Frobenius similarity, Shiraishi signatures 16, 23, 24, and 25 all have similarity >= 0.7 to COSMIC signature 3. For cosine similarity, Shiraishi signatures 16, 23, and 25 all have similarity > 0.7 to COSMIC signature 3. We have designated Shiraishi signatures 16 and 23-25 as HR signatures on heatmaps that show Shiraishi signatures. Knowing which Shiraishi signatures correspond to an HR signature, we proceeded to determine which signature, if any, of K signatures produced by our analysis, are similar to one of those Shiraishi HR signatures. We used Frobenius similarity in that case, since both signatures being compared are Shiraishi signatures, and the comparison is more reliable than the Alexandrov-Shiraishi comparison. Frobenius similarity showed that, at K=3, signature #1 [noted as 1(HR)] was most similar to the Shiraishi HR signatures 16, 23, and 25 (full analysis, Supplementary Figure 1C). Tumor DNA from our study samples was derived from FFPE tissue, and was thus expected to have a higher percentage of C: G>T: A mutations. Therefore we analyzed mutational signatures after removing C:G>T:A from our study samples as well as from control samples (restricted analysis). Similar to the full analysis, we first identified signatures with high Frobenius similarity to Shiraishi HR signatures, using K=3 (Supplementary Figure 1D). After optimizing the method, we proceeded to determine whether an HR signature was demonstrated by the four study samples where somatic WES data was available (see above). As shown in Supplementary Figure 2, our study samples as well as the TCGA positive controls, at K=3, in full and restricted analysis have a significantly higher relative contribution or membership weight for the HR signature compared to the negative controls. Interestingly, another hallmark of somatic HR deficiency is a high frequency of large indels [16, 19]. Consistently, similar to TCGA HR-positive controls, the mean deletion length found in the tumors from our four *PALB2/RAD51C* mutation carriers was higher than in TCGA nonHR GC cases (31.6 bp vs. 15.4 bp, $P= 3 \times 10^{-7}$).

REFERENCES

1. van der Post RS. J Med Genet 2015;52:361-374.
2. Buffalo V. (Version 0.994Beta) [Software]. Available at https://github.com/vsbuffalo/scythe. 2011.
3. Joshi N. (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle. 2011.
4. Li H. Preprint at arXiv:1303.3997v2 [q-bio.GN] 2013.
5. McKenna A. Genome Res 2010;20:1297-303.
6. DePristo MA. Nat Genet 2011;43:491-8.
7. Garrison EM. Preprint at arXiv:1207.3907v2 [q-bio.GN] 2012.
8. Wei Z. Nucleic Acids Res 2011;39:e132.
9. Koboldt DC. Genome Res 2012;22:568-76.
10. **Li H**, **Handsaker B**. Bioinformatics 2009;25:2078-9.
11. Wang K. Nucleic Acids Res 2010;38:e164.
12. Van der Auwera GA. Curr Protoc Bioinformatics 2013;11:11 10 1-11 10 33.
13. Rahman N. Nature 2014;505:302-8.
14. Robinson JT. Nat Biotechnol 2011;29:24-6.
15. Cibulskis K. Nat Biotechnol 2013;31:213-9.
16. Alexandrov LB. Cell Rep 2013;3:246-59.
17. Alexandrov LB. Nature 2013;500:415-21.
18. Shiraishi Y. PLoS Genet 2015;11:e1005657.
19. Alexandrov LB. Nat Commun 2015;6:8683.
20. Team RC. URL http://www.R-project.org/. 2016.

Author names in bold designate shared co-first authorship

Supplementary Table 1: Details of mutations identified in the study

Chr position (Genome assembly = GRCh37/hg19)	Ref	Alt	Gene name	Trasncript ID	cDNA change	Protein change and effect	Pathogenicity prediction	Type, effect on protein	ExAC frequency
16: 23646627	G	A	*PALB2*	NM_024675.3	c.1240C>T	p.Arg414Ter,	Reported Pathogenic in ClinVar	Nonsense, truncates protein	NA
16: 23625324	C	A	*PALB2*	NM_024675.3	c.3201+1G>T		Reported Pathogenic in ClinVar	Splice-donor variant	NA
16: 23641585-23641593	GCAGGACTT	-	*PALB2*	NM_024675.3	c. 1882_1890 delGCAGGACTT	p.Lys628_Cys630del ,	Reported as VUS* in ClinVar,	In-frame deletion, possible effect on recruitment to DNA damage site (see text)	$3.31X10^{-5}$
16: 23635411	G	T	*PALB2*	NM_024675.3	c. 2753C>A	p.Pro918Gln	Reported as VUS* in ClinVar, predicted deleterious in SIFT, PolyPhen, LRT and MutationTaster	Missense, possible effect on protein – protein interaction	$1.742X10^{-5}$
17: 41244214-41244217	CAAG	-	*BRCA1*	NM_007294.3	c. 3331_3334del CAAG	p.Gln1111Asnfs	Pathogenic	Frameshift deletion, truncates protein	NA
17: 41245874	A	-	*BRCA1*	NM_007294.3	c.1674delA	p.Gly559Valfs	Reported Pathogenic in ClinVar	Frameshift deletion, truncates protein	NA
17: 56787223	C	T	*RAD51C*	NM_058216.2	c.709C>T	p.Arg237Ter	Reported Pathogenic in ClinVar	Nonsense, truncates protein	$8.23X10^{-6}$

Supplementary Table 5: *VUS = Variant of uncertain significance, ExAC = Exome aggregation consortium

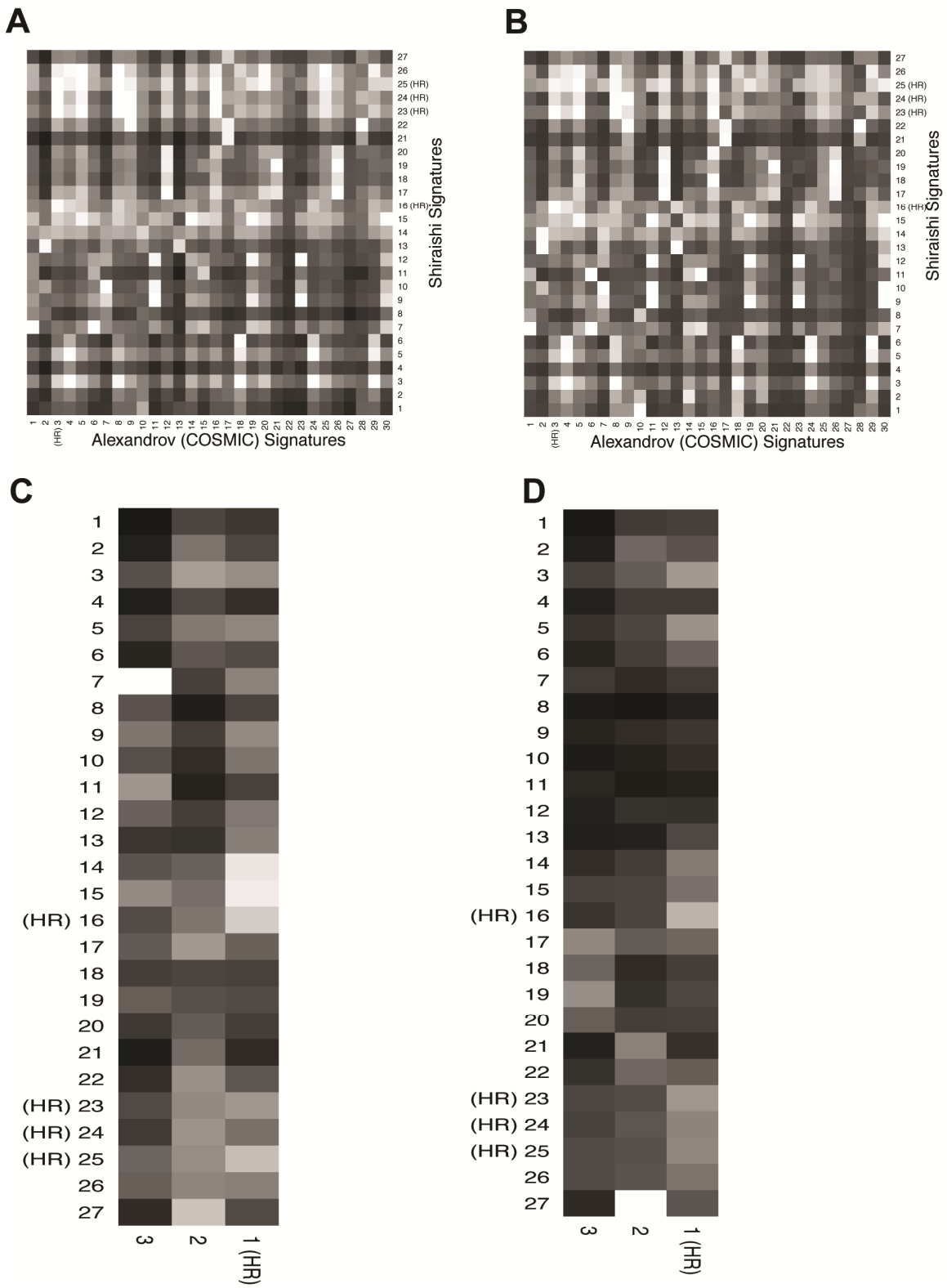
SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1: Mutational signature analysis methods. Similarity between mutation signatures identified by Alexandrov *el al* (COSMIC) and Shiraishi *el al.* using **A =** Frobenius similarity measures and **B** = Cosine similarity measures. For Frobenius similarity, Shiraishi signatures 16, 23, 24, and 25 all have similarity >= 0.7 to COSMIC signature #3, and for cosine similarity, Shiraishi signatures 16, 23, and 25 all have similarity > 0.7 to COSMIC signature #3. Frobenius similarity was used to determine, at K=3, which signature showed most similarity to the Shiraishi HR signatures. Considering that our study samples were derived from FFPE tumor DNA, this analysis was performed on the full set of SNV mutations (full analysis) as well as after removal of C:G>T:A changes (restricted analysis), a known artifact of FFPE tissue processing. Signatures with high Frobenius similarity to the Shiraishi HR signatures were identified for K=3 for **C** = full analysis and **D** = restricted analysis. As shown in C and D, signature #1 (noted on axis as HR) is most similar to Shiraishi HR signature.

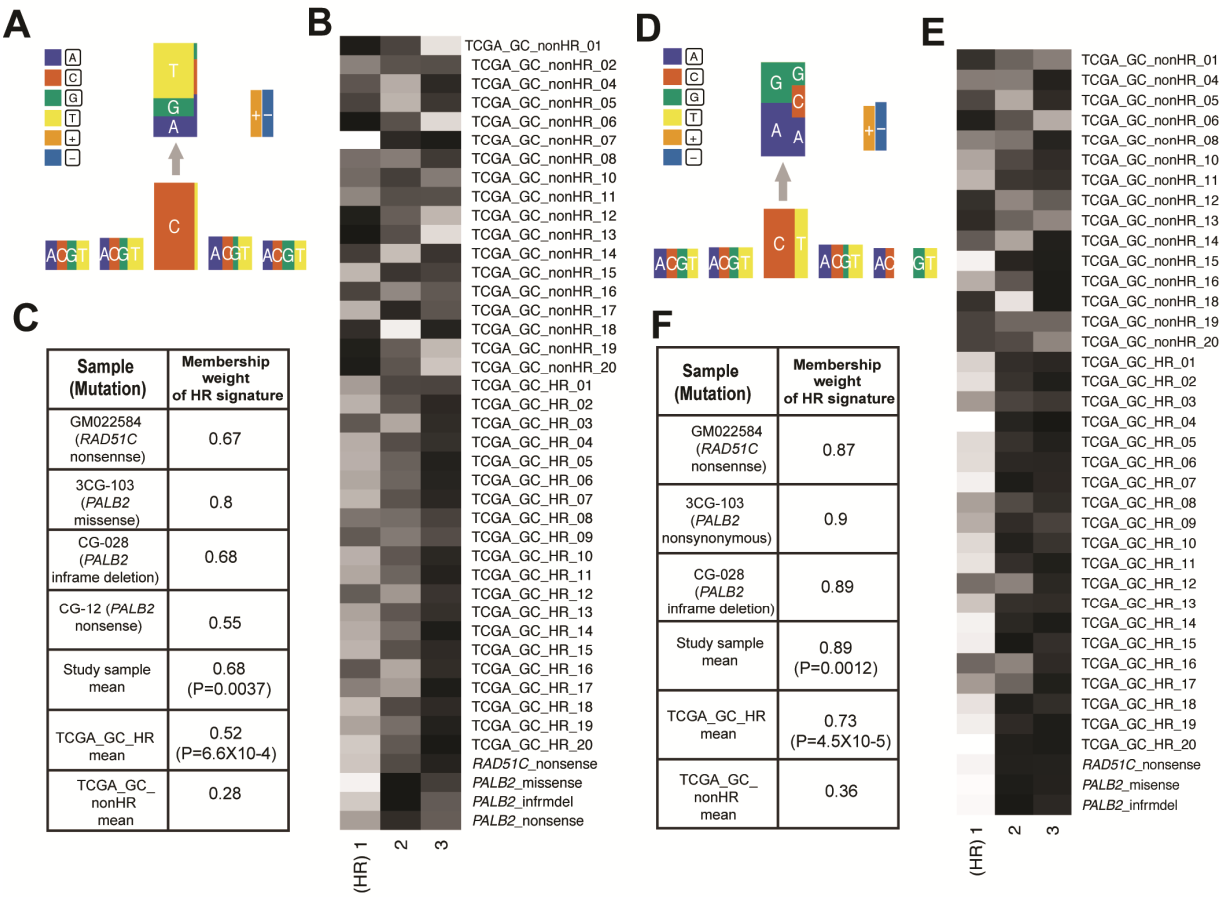
Supplementary Figure 2: Analysis of mutational signatures in tumor samples. We used whole exome sequence (WES) data from four *PALB2* and *RAD51C* mutation carriers (GM022584, 3CG-103, CG-028 and CG-12) and from 40 HR defective (TCGA_GC_HR, n=20) and HR proficient (TCGA_GC_nonHR, n=20) cases from the TCGA study. These analyses included all mutations (full analyses, A-C, left panel) and removal of C:G>T:A changes (restricted analyses, D-E, left panel) as our WES data was generated from archival tumors, which are known to accumulate artifactual C:G>T:A mutations. **A and D**. Logos of somatic HR signatures. The central base represents the frequency of the mutation, which is surrounded by the frequency of bases at positions -2 and -1 (left) and +1 and +2 (right). The top right bars indicate the frequency of such mutations in the + and – transcription strand polarities (see ref 14 for more details). **B and E**. Heatmaps of relative contribution or membership weights of each signature

within each sample. Dark shading indicates low contribution of the mutation signature and light shading represents high contribution of the mutation signature. Our four samples had highest membership weight to signature #1 (the HR signature) and clustered in the full (which included all mutations, panel B, right) and restricted (which excluded C:G>T:A changes, panel E, left) analyses with the TCGA HR-positive cases. The pattern involving signatures #2 (unknown cases but very similar to a previously reported signature by Shiraishi *et al* in gastric and colorectal tumors) and #3 (cytosine deamination) showed stronger membership weights with the non-HR samples. The *PALB2* nonsense mutation carrier and five TCGA_GC_nonHR samples were removed from the restricted analysis as they had few mutations after removal of C:G>T:A changes. **C and F.** Tables indicating membership weights for each sample. Table indicates the estimated fraction of mutations associated with the HR signature pattern. Study sample mean indicates mean membership weight of HR signature. P-value from Mann-Whitney two-sample U test compares membership weight of the Study sample mean or TCGA_GC_HR sample mean to TCGA_GC_nonHR sample mean (row 6 and 8 and row 7 and 8) respectively.

Supplementary Figure 1



Supplementary Figure 2



C

Sample (Mutation)	Membership weight of HR signature
GM022584 (*RAD51C* nonsennse)	0.67
3CG-103 (*PALB2* missense)	0.8
CG-028 (*PALB2* inframe deletion)	0.68
CG-12 (*PALB2* nonsense)	0.55
Study sample mean	0.68 (P=0.0037)
TCGA_GC_HR mean	0.52 (P=6.6X10-4)
TCGA_GC_ nonHR mean	0.28

F

Sample (Mutation)	Membership weight of HR signature
GM022584 (*RAD51C* nonsennse)	0.87
3CG-103 (*PALB2* nonsynonymous)	0.9
CG-028 (*PALB2* inframe deletion)	0.89
Study sample mean	0.89 (P=0.0012)
TCGA_GC_HR mean	0.73 (P=4.5X10-5)
TCGA_GC_ nonHR mean	0.36