Supplementary Methods

qRT-PCR validation

Total RNA from the normal squamous tissue of the 12 AA and 12 EA examined by ST 2.1 array, as well as additional normal squamous tissue from 9 AA and 9 EA were then used for real-time (qRT-PCR) validation of selected gene transcripts. Isolated RNA was converted to cDNA and used for qRT-PCR as previously described[1]. Primers for qRT-PCR reactions were designed using Primer-BLAST[2] (*GSTT2*: primers for cDNA Exons 1-2 5'-TGGGCCTAGAGCTGTTTCTT-3', 3'-CCAGGCTGTTGATCTGCAAG-5' and *GSTT1* primers for cDNA 5'- CTGGAGTTTGCTGACTCCCTC-3', 3'-GCTCGAAGGGAATGTCGTTCT-5') or previously published (*GAPDH* and/or *B-Actin*)[3]. Optimal annealing temperatures and reaction conditions were confirmed using the Cepheid SmartCycler (Cepheid, Sunnyvale, CA). Samples were run in duplicate using the ABI PRISM® 7900HT Sequence Detection System (Thermo Fisher, Waltham, MA) according to the manufacturer's instructions. Thermal cycling conditions consisted of 10min of initial denaturation at 95°C, and 40 cycles of 15sec of denaturation at 95°C, and 1min of annealing/extension at 60°C. One AA:NE sample failed to amplify for *GSTT2* PCR primers. Relative expression and fold-change of the genes of interest (*GSTT2* and *GSTT1*) were calculated by the ΔΔ-Ct method (subtracted the average of the Ct's for the reference gene *ACTB* from the Ct values of each target gene)[4].

GSTT2B deletion genotype detection

Using DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA), DNA was extracted from cell lines following the manufacturer's instructions. Normal squamous biopsy genomic DNA was extracted as described previously with modifications[5]. The *GSTT2B* genotype was determined using a three-primer set for PCR as previously described[6]. The PCR reaction contained a total volume of 50µl and consisted of 25µl of GoTaq Green Master Mix (Promega, Madison, WI), 10pmol of primers *GSTT2B*-6858 and *GSTT2B-2B*, 15pmol of primer *GSTT2B*-6857, and 25ng genomic DNA. The thermal cycling conditions for the reaction consisted of initial denaturation for 2min at 95°C, followed by 35 cycles of 30sec of denaturation at 95°C, 30sec of annealing at 60°C, and 45sec of extension at 72°C. PCR products were then separated on 2% agarose gels (Fisher) with ethidium bromide by electrophoresis in 1X TBE buffer with 100bp DNA ladder (Thermo Fisher) as a marker.

GSTT2/2B promoter genotype resolution and sequencing verification

The primers and PCR protocol from Marotta *et al* [7] was used to genotype the 17bp promoter duplication status. PCR products were loaded on 2% agarose gel (Fisher) and run overnight (>18hrs) at 20V to provide adequate resolution of resulting alleles (17 bp distinction between ~320bp products). Two examples of each promoter genotype were chosen for sequencing-based confirmation. PCR products were purified using

Qiaquick (Qiagen) columns as per manufacturer's instructions and submitted to the University of Michigan Sequencing Core for Sanger sequencing using the reverse primer. Resulting sequence reads were aligned using Geneious software (version 5.4.6; Biomatters Ltd, Auckland, New Zealand) and manually to match 17bp genotypes to qPCR results.

Cumene-hydroperoxide (cum-OOH) treatment of Het-1A and HeLa Cell line

Het-1A (an immortalized cell line from normal esophageal squamous mucosa) and HeLa (immortalized cell line from cervical cancer) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) (Cat. No. 11965092) with 10% fetal bovine serum and 1% antibiotic-antiseptic. Cells were plated in 35mm x 10mm polystyrene plates (# 430165) (Corning Incorporated, NY) with 4 coverslips (6mm) within the plate, at a starting number of 120,000 cells per plate. Plates and coverslips were prepared to ensure enough coverslips for duplicates for all experimental conditions. For knockdown of GSTT2, four siRNAs (siRNA-05 GCUAAGGAUGGUGAUUUC; siRNA-06 GCACCGUGGAUUUGGUCAA; siRNA-07 AGGCUAUGCUGCUUCGAAU, siRNA-08 GACACUGGCUGAUCUCAUG; (#LQ-011181-00-0005, Dharmacon, Lafayette, CO) directed at *GSTT2* were used at a concentration of 10nm. For controls, we used a non-target (NT) siRNA (10nm), as well as treated cells with only OPTI-MEM plus RNAimax (mock control) (Thermo Fisher). Cells were plated and transfected using Lipofectamine® RNAiMAX Transfection Reagent (as directed by RNAiMAX protocol, Cat. No. 13778030 (Thermo Fisher) on day 1, with the following groups: mock, NT, si05, si06, si07, and si08. On day 2, the media was changed and cells were transfected a second time. On day 3, the media and cell were incubated for 48hrs. After 48hrs, coverslips were transferred to randomly assigned wells in 24-well plates and cells were then treated with either 0μM or 100μM of cumene hydroperoxide (cum-OOH) (#247502; Sigma-Aldrich, St. Louis, MO) for 1hr. Experimental conditions were organized with each condition present on a plate such that there were duplicated 24 well plates representing the whole experiment. After treatment, the cells were fixed for 20min at -20 °C using 100% cold methanol, washed with PBS (3 x 5min) and allowed to stand in fresh PBS at 4°C before their use in the immunofluorescence protocol.

Immunofluorescence analysis of γ-H2AX

PBS was removed from the fixed samples and the protein-DNA was cross-linked with 500μL of 10% phosphate-buffered formalin for 20min at room temperature (RT). Cells were washed once with 500μL TBS followed by permeabilization with 100% cold methanol (-20°C) then incubated for 1hr in a blocking buffer containing 5% goat serum, 1% BSA, and 0.2% Triton x100 in TBS (500μl of blocking buffer per well). We diluted each primary antibody (γ-H2AX 1:1500 (#05-636, Millipore) GSTT2-1:500 (#514667, Santa Cruz) in TBS with 1% BSA such that 150μL was used for each cover glass and allowed to incubate on the cells overnight at 4 °C in a humidified chamber. The following day cells were washed 3 times with TBS-T for 5min. The cells were then incubated with secondary antibody for 1hr at room temperature (1:500

Alexa Fluor 488 goat anti-mouse IgG1, #A21121, Life Technologies; 1:200 Alexa Fluor 594 goat anti-mouse IgG2a, #A21135, Life Technologies), in 1% BSA, followed by three 5min washes with TBS-T. Coverslips were mounted onto slides using DAPI mounting solution (# P36935, Thermo Fisher) and stored at -20°C prior to microscopic and photographic imaging. For γ-H2AX response quantitation, each coverslip was conceptually cordoned into 3 quadrants, upper left, upper right and lower (left and right combined), with several high-resolution images (IX73:Inverted Microscope, Olympus, Life Science Solutions) taken to ensure that at least 50 DAPI stained nuclei were scored per quadrant. We chose to use the ratio of DNA repair active nuclei (>10 γ-H2AX positive nuclear foci) verse DAPI stained nuclei as our DNA damage metric. These per quadrant ratios were averaged and replicate plate averages were in-turn, averaged to provide per-experiment ratios for each experimental condition. The experiment was repeated twice, with the resulting daily averaged ratios considered as part of the 4-way ANOVA analysis model, along with terms for cum-OOH (0 vs 100uM), cell line (Het-1A vs HeLA) and knock down treatment (control, siRNA05, siRNA06). The ANOVA model used could be written as Yijkm = Aijk +Bm +Eijkm, where i is cell type, j is cum-OOH dose, k is treatment, and m is day, such that Aijk represents the interaction between cell line, cum-OOH and KD treatment measures, the Bm term added a component to account for experimental (daily) variation and Eijkm represents the random residual error across all components. Analyses were done via an Excel (Microsoft, Seattle) spreadsheet using the FDIST function. We generated at each between-group and difference of difference comparison *P* value, which we report without adjustment below. For simplicity, we chose to represent aspects of this comparison in three separate figures, such that **Figure 7(A and B)**, and **Supplementary Figure S7(I and J)** demonstrate the effect of Cum-OOH and GSTT2 siRNA knock down on squamous cell lines Het-1A and HeLa, respectively, while **Supplementary Figure S6(D)** compares the response of both cell lines to cum-OOH treatment, without the presence of GSTT2 specific siRNA. Between-experiment variation (term Bm above) proved to be a very minor determinant in between group comparison and showed minimal contributions to p-values. To determine this, we removed the Bm component from the model and compared the resulting two-group comparison *P*-values to the full model. A summary of ANOVA models and group comparison *P* values, with and without the between-experiment (Bm) component are given below.

Analysis summary of isRNA knockdown and cum-OOH treatment on γ-H2AX response in Het-1A and HeLa

	experiment day 1 average ratios						experiment day 2 average ratios						averages over the two days.					
Cu-OOH	**0uM**	**0uM**	**0uM**	**100uM**	**100uM**	**100uM**	**0uM**	**0uM**	**0uM**	**100uM**	**100uM**	**100uM**	**0uM**	**0uM**	**0uM**	**100uM**	**100uM**	**100uM**
condition	**control**	**siRNA05**	**siRNA06**	**control**	**siRNA05**	**siRNA06**	**control**	**siRNA05**	**siRNA06**	**control**	**siRNA05**	**siRNA06**	**control**	**siRNA05**	**siRNA06**	**control**	**siRNA05**	**siRNA06**
day	**1**	**1**	**1**	**1**	**1**	**1**	**2**	**2**	**2**	**2**	**2**	**2**	**average**	**average**	**average**	**average**	**average**	**average**
Het-1A	0.014	0.23	0.11	0.11	0.69	0.65	0.013	0.18	0.15	0.27	0.68	0.75	0.013	0.21	0.13	0.19	0.69	0.70
HeLa	0.011	0.19	0.22	0.17	0.66	0.55	0.040	0.30	0.11	0.18	0.44	0.46	0.026	0.24	0.17	0.18	0.55	0.51

4-Way ANOVA - full model	Full model is **Yijkm = Aijk +Bm +Eijkm**, where **i** is cell type (Het-1A vs HeLa), **j** is Cu-OOH dose (0μM vs 100μM), **k** is treatment (control, siRNA05, siRNA06), and **m** is day (1 vs 2).
3-Way ANOVA - no intra-day component	Model is **Yijkm = Aijk +Eijkm**, where **i** is cell type (Het-1A vs HeLa), **j** is Cu-OOH dose (0μM vs 100μM), **k** is treatment (control, siRNA05, siRNA06).

between group comparisons	**4-way ANOVA P values**	**3-way ANOVA P values**	**% change**	**Fold change**
si05 vs control, 0uM, Het-1A	0.026	0.020	19%	15.41
si06 vs control, 0uM, Het-1A	0.14	0.13	12%	9.84
si05 vs control, 0uM, HeLA	0.014	0.010	22%	9.50
si06 vs control, 0uM, HeLA	0.087	0.073	14%	6.51
si05 vs control, 100uM, Het-1A	3.93E-05	1.70E-05	49%	3.54
si06 vs control, 100uM, Het-1A	2.92E-05	1.23E-05	51%	3.63
si05 vs control, 100uM, HeLA	4.29E-04	2.30E-04	37%	3.09
si06 vs control, 100uM, HeLA	1.06E-03	6.13E-04	33%	2.86
si06 vs si05, 0uM, Het-1A	0.34	0.32	-7%	0.64
si06 vs si05, 0uM, HeLa	0.33	0.31	-8%	0.69
si06 vs si05, 100uM, Het-1A	0.83	0.82	2%	1.02
si06 vs si05, 100uM, HeLa	0.59	0.57	-4%	0.92
100 vs 0 uM, control, Het-1A	0.034	0.027	18%	14.58
100 vs 0 uM, control, HeLa	0.067	0.056	15%	6.94
100 vs 0 uM, si05, Het-1A	4.82E-05	2.13E-05	48%	3.35
100 vs 0 uM, si05, HeLa	1.80E-03	1.09E-03	31%	2.26
100 vs 0 uM, si06, Het-1A	9.96E-06	3.81E-06	57%	5.37
100 vs 0 uM, si06, HeLa	8.27E-04	4.70E-04	34%	3.05
Het-1A vs HeLa, 0uM control	0.87	0.87	-1%	0.52
Het-1A vs HeLa, 0uM siRNA05	0.62	0.61	-4%	0.84
Het-1A vs HeLa, 0uM siRNA06	0.65	0.63	-4%	0.79
Het-1A vs HeLa, 100uM control	0.83	0.82	2%	1.09
Het-1A vs HeLa, 100uM siRNA05	0.092	0.080	14%	1.25
Het-1A vs HeLa, 100uM siRNA06	0.024	0.019	20%	1.39

difference-of-differences	**4-way ANOVA P values**	**3-way ANOVA P values**
si05 vs control, 0uM, Het-1A vs HeLa	0.81	0.81
si06 vs control, 0uM, Het-1A vs HeLa	0.83	0.82
si05 vs control, 100uM, Het-1A vs Hel	0.27	0.25
si06 vs control, 100uM, Het-1A vs Hel	0.12	0.10
100 vs 0 uM, control, Het-1A vs HeLa	0.79	0.78
100 vs 0 uM, si05, Het-1A vs HeLa	0.12	0.11
100 vs 0 uM, si06, Het-1A vs HeLa	0.051	0.041
si05 vs control, 100 vs 0uM, Het-1A	0.016	0.012
si06 vs control, 100 vs 0uM, Het-1A	3.48E-03	2.24E-03
si05 vs control, 100 vs 0uM, HeLa	0.17	0.15
si06 vs control, 100 vs 0uM, HeLa	0.10	0.087

Western blotting following cum-OOH treatment

Cells were incubated and plated as described above (Treatments cum-OOH section). After 1 PBS wash, 40μl of Cell Signaling protein lysis buffer (1X) with proteinase inhibitor cocktail (PIC) (20μl per mL) was added followed by scraping the plates. Cell lysates were transferred to 1.5mL tubes, spun for 20min at 4°C and supernatant protein was quantified using protein assay. All Western lysates were prepared at 20μg total, with 20% 2-mercaptoethanol, and 4X sample buffer. Samples were resolved using SDS Novex gels (4-12% gradient) (Thermo Fisher) run for 2hrs at 125-126V. Nitrocellulose membranes were activated with 100% methanol for 1min and transfer was performed using Novex Transfer Buffer run overnight at 12V. After transfer, membranes were stained with Ponceau red for assessment of transfer efficiency. Blocking was performed by incubating the membranes in 5% milk/TBST or 5% BSA/TBST for 1hr at room temperature. Primary antibodies [GSTT2 (mouse monoclonal isotype IgG2a #514667; Santa Cruz, Dallas, TX); γ-H2AX (anti-phospho-histone H2AX (Ser139) isotype IgG1 #05-636; Millipore, Burlington, MA)] were diluted in

5% milk/TBST or 5% BSA/TBST and incubated over-night at 4°C. Membranes were then washed 3 times (5min per wash) with TBS-T. Secondary antibodies were diluted in 5% milk/TBST and added to the membranes for 1hr incubations at room temperature. The membranes were washed 3 times with TBST, 5min per wash, and then incubated with ECL for 5min before imaging using X-ray film.

Supplementary Methods References:

1. Ferrer-Torres D, Nancarrow DJ, Kuick R, et al. Genomic similarity between gastroesophageal junction and esophageal Barrett's adenocarcinomas. Oncotarget 2016;7:54867-54882.
2. Ye J, Coulouris G, Zaretskaya I, et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 2012;13:134.
3. Lin L, Bass AJ, Lockwood WW, et al. Activation of GATA binding protein 6 (GATA6) sustains oncogenic lineage-survival in esophageal adenocarcinoma. Proc Natl Acad Sci U S A 2012;109:4251-6.
4. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25:402-8.
5. Blin N, Stafford DW. General Method for Isolation of High Molecular-Weight DNA from Eukaryotes. Nucleic Acids Research 1976;3:2303-2308.
6. Zhao Y, Marotta M, Eichler EE, et al. Linkage disequilibrium between two high-frequency deletion polymorphisms: implications for association studies involving the glutathione-S transferase (GST) genes. PLoS Genet 2009;5:e1000472.
7. Marotta M, Piontkivska H, Tanaka H. Molecular trajectories leading to the alternative fates of duplicate genes. PLoS One 2012;7:e38958.

Supplementary Table S1. Summary of Sample Group Characteristics[1]

	N	Age >55 years	Gender Male	Smoked history[2]	GERD status[3]	BMI >25 overweight[4]	BE/EAC pathology
AA-NE	12	8 (67%)	12 (100%)	11 (92%)	2 (17%)	11 (92%)	0 (0%)
EA-NE	12	5 (42%)	12 (100%)	10 (83%)	4 (33%)	11 (92%)	0 (0%)
AA-NE:BE	8	7 (88%)	5 (63%)	7 (88%)	8 (100%)	7 (88%)	8 (100%)[5]
EA-NE:BE	8	7 (88%)	6 (75%)	6 (75%)	7 (88%)	6 (75%)	8 (100%)[6]

1 - clinical/demographic data was collected in conjunction with study enrollment consent

2 - individuals classified as positive recorded as current or former smoker

3 - individual classified as positive when reporting regurgitation or heart burn at least once per week, or with a medical diagnosis of GERD

4 - individuals reporting current weight as >25 classified as overweight

5 - one individual resected for EAC with prior BE history, one resected for HGD, one with LGD, five with non-dysplastic BE

6 - four individuals resected for EAC with prior BE history, four resected for EAC at initial presentation

Supplementary Table S2. Genes with FC > 3 and ANOVA *P* < 0.01 for AA-NE vs EA-NE comparison

			ANOVA *P*-value*				Fold change^			
Symbol	**Gene Name**	**Entrez Gene**	**AA-NE vs EA-NE**	**NE:BE vs EA-**	**NE:BE vs AA-NE**	**NE:BE vs EA-NE**	**AA-NE / EA-NE**	**AA-NE:BE / EA-NE:BE**	**AA-NE:BE / AA-NE**	**EA-NE:BE / EA-NE**
GSTT2	glutathione S-transferase theta 2	2953	0.00035	0.1372	0.0479	0.8557	5.15	2.17	0.39	0.92
GSTT2B	glutathione S-transferase theta 2B (gene/pseudogene)	653689	0.00157	0.0700	0.1038	0.5163	3.13	2.15	0.54	0.78
HLA-DPB1	major histocompatibility complex, class II, DP beta 1	3115	0.00724	0.9458	0.6697	0.0486	3.13	1.03	0.82	2.50
IGHD	immunoglobulin heavy constant delta	3495	0.00763	0.0307	0.0644	0.0039	5.07	0.21	0.29	7.24
IGHA1	immunoglobulin heavy constant	3493	0.00742	5.45E-06	0.0169	1.03E-06	3.77	0.05	0.27	21.53

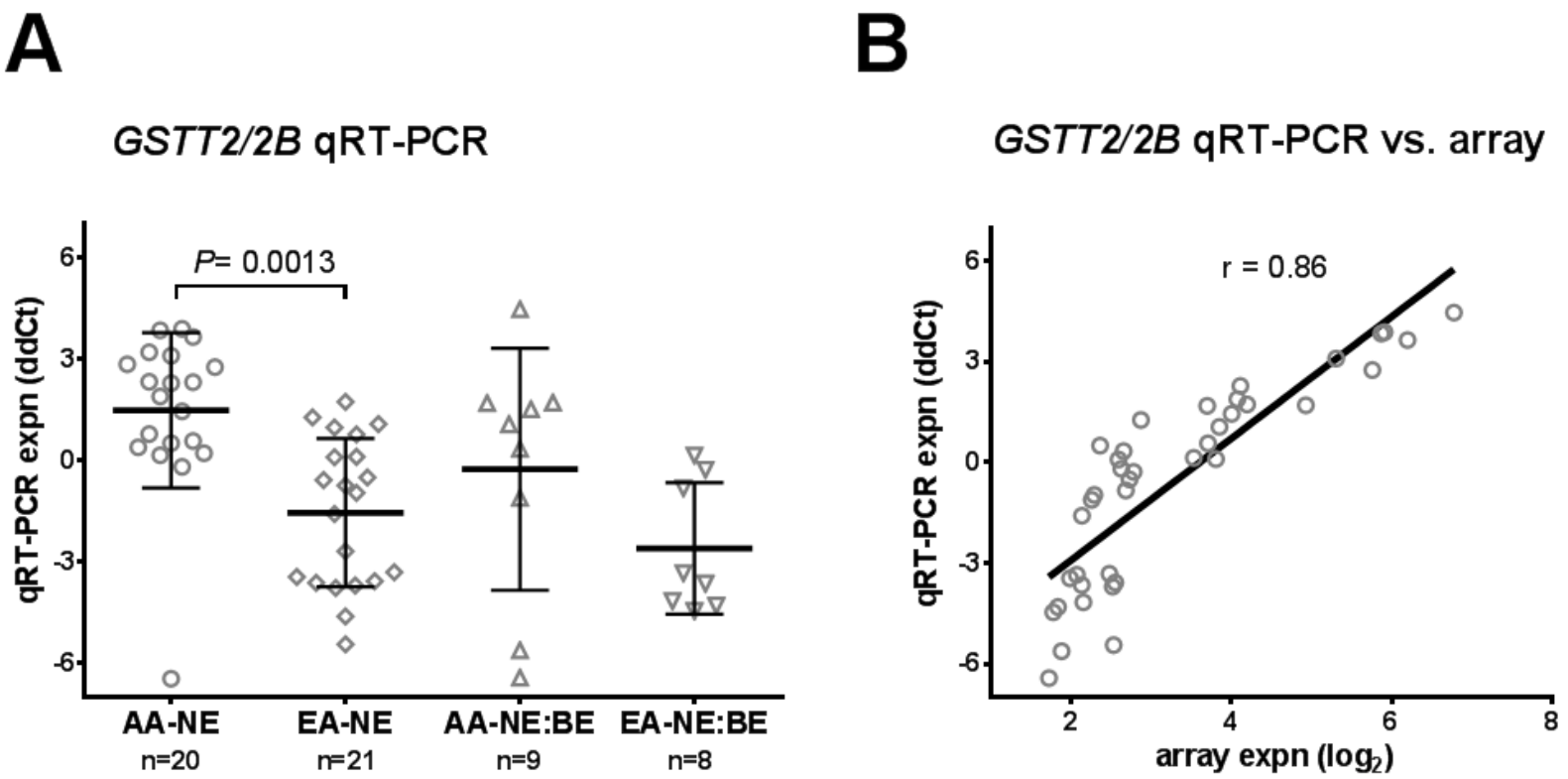
**P* value from one-way ANOVA comparison of AA-NE and EA-NE normalized log 2 Affy ST 2.1 expression data

^Mean group fold changes (FC) were calculated and presented as non-log ratios

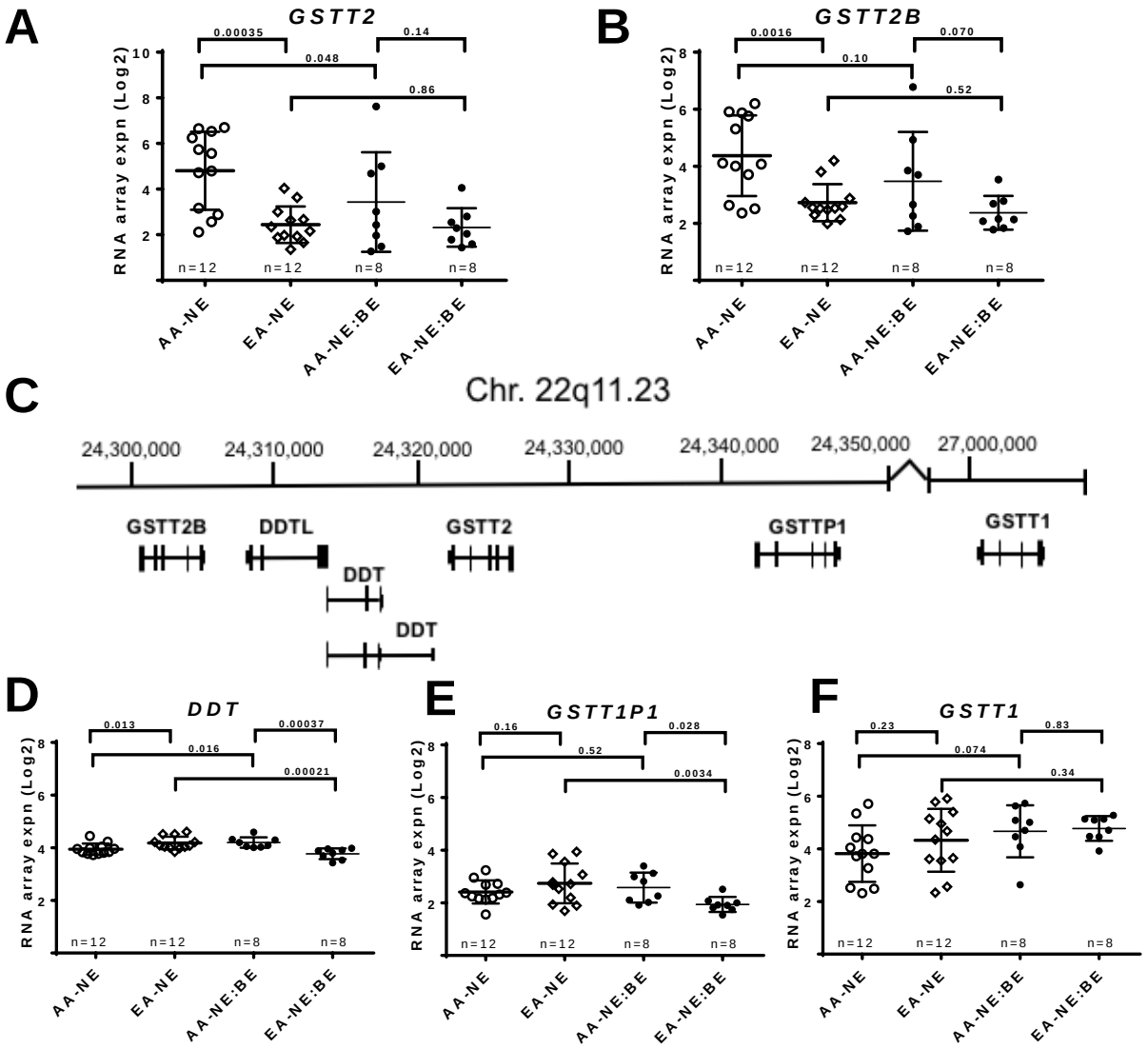
Supplementary Table S3. Promoter duplication allele and genotype frequency details across 1000G population groups

				Allele			Genotype			
Population description	**Pop. code**	**Super pop.**	**sample size* (subjects)**	**non-duplicated [=w]**	**duplicated [=D]**	**duplicated allele frequency**	**homozygous non-duplicates [=ww]**	**heterozygotes [=wD]**	**homozygous duplicates [=DD]**	**homozygous duplicate frequency**
African Carribbeans in Barbados	ACB	AFR	61	34	88	72.1%	15	4	42	68.9%
Americans of African Ancestry in SW USA	ASW	AFR	30	16	44	73.3%	7	2	21	70.0%
Esan in Nigera	ESN	AFR	63	25	101	80.2%	7	11	45	71.4%
Gambian in Western Divisons in The Gambia	GWD	AFR	76	53	99	65.1%	17	19	40	52.6%
Luhya in Webuye	LWK	AFR	48	14	82	85.4%	4	6	38	79.2%
Mende in Sierra Leone	MSL	AFR	60	51	69	57.5%	21	9	30	50.0%
Yoruba in Ibadan	YRI	AFR	57	39	75	65.8%	14	11	32	56.1%
Colombians from Medellin, Colombia	CLM	AMR	53	16	90	84.9%	4	8	41	77.4%
Mexican Ancestry from Los Angeles USA	MXL	AMR	24	0	48	100.0%	0	0	24	100.0%
Peruvians from Lima, Peru	PEL	AMR	53	4	102	96.2%	0	4	49	92.5%
Puerto Ricans from Puerto Rico	PUR	AMR	48	7	89	92.7%	1	5	42	87.5%
Chinese Dai in Xishuangbanna, China	CDX	EAS	68	8	128	94.1%	1	6	61	89.7%
Han Chinese in Beijing	CHB	EAS	58	5	111	95.7%	1	3	54	93.1%
Han Chinese South	CHS	EAS	44	2	86	97.7%	0	2	42	95.5%
Japanese in Tokyo	JPT	EAS	57	6	108	94.7%	2	2	53	93.0%
Kinh in Ho Chi Minh City, Vietnam	KHV	EAS	65	6	124	95.4%	2	2	61	93.8%
Utah residents with ancestry from Europe	CEU	EUR	58	3	113	97.4%	1	1	56	96.6%
Finnish in Finland	FIN	EUR	34	3	65	95.6%	0	3	31	91.2%
British from England and Scotland	GBR	EUR	47	3	91	96.8%	1	1	45	95.7%
Iberian populations in Spain	IBS	EUR	62	14	110	88.7%	3	8	51	82.3%
Toscani in Italia	TSI	EUR	67	16	118	88.1%	7	2	58	86.6%
Bengali from Bangladesh	BEB	SAS	60	0	120	100.0%	0	0	60	100.0%
Gujarati Indian from Houston, Texas	GIH	SAS	64	11	117	91.4%	5	1	58	90.6%
Indian Telugu from the UK	ITU	SAS	64	6	122	95.3%	2	2	60	93.8%
Punjabi from Lahore, Pakistan	PJL	SAS	63	2	124	98.4%	1	0	62	98.4%
Sri Lankan Tamil from the UK	STU	SAS	59	5	113	95.8%	0	5	54	91.5%

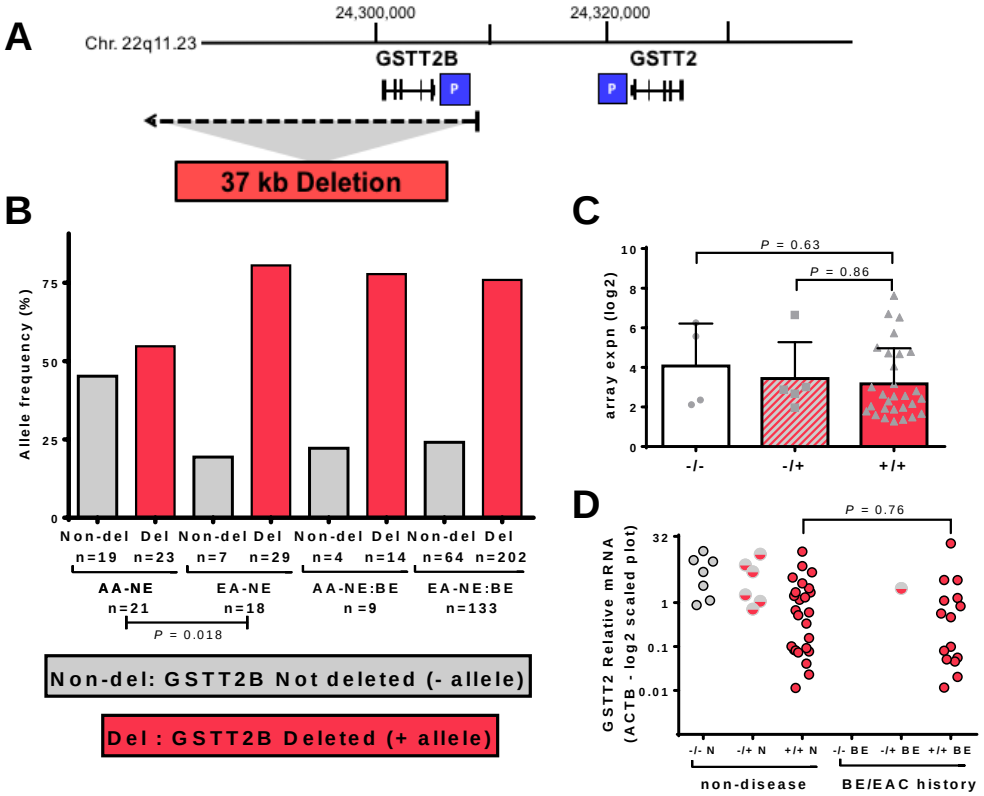
* not all samples yielded data for *GSTT2* promotor genotype status



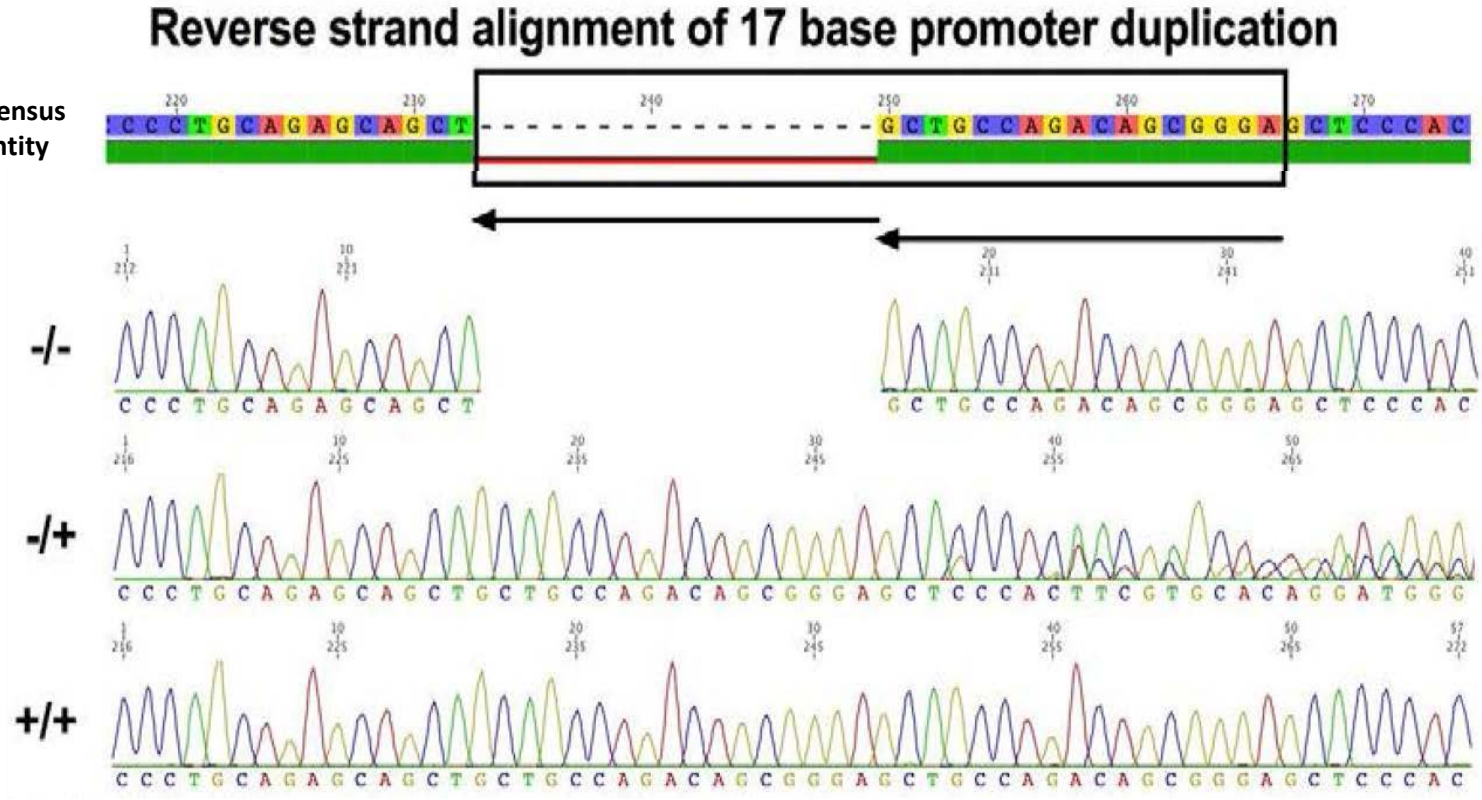
Supplementary Figure S1. PCR based validation of *GSTT2* expression and correlation to array data. (A) Validation of *GSTT2/2B* expression in the NE of AA and EA with and without Barrett's (BE) using qRT-PCR with a larger cohort confirms differential mRNA expression of *GSTT2/2B* in AA vs. EA (*P*=0.0013). One-way ANOVA was used with Tukey post-hoc adjustment for comparisons performed across the 4 sample groups. The significant *P* value for the AA-NE vs. EA-NE comparison is shown, while the other comparisons, AA-NE:BE vs. EA-NE:BE, AA-NE vs. AA-NE:BE and EA-NE vs. EA-NE:BE were not significant with *P* values of 0.21, 0.30 and 0.73 respectively. **(B)** Pearson-correlation analysis for *GSTT2/2B* comparing log2 normalized Human Gene 2.1 ST arrayed samples and relative qRT-PCR expression, with ddCt with *ACTB* as the reference gene (r=0.86, n=39).



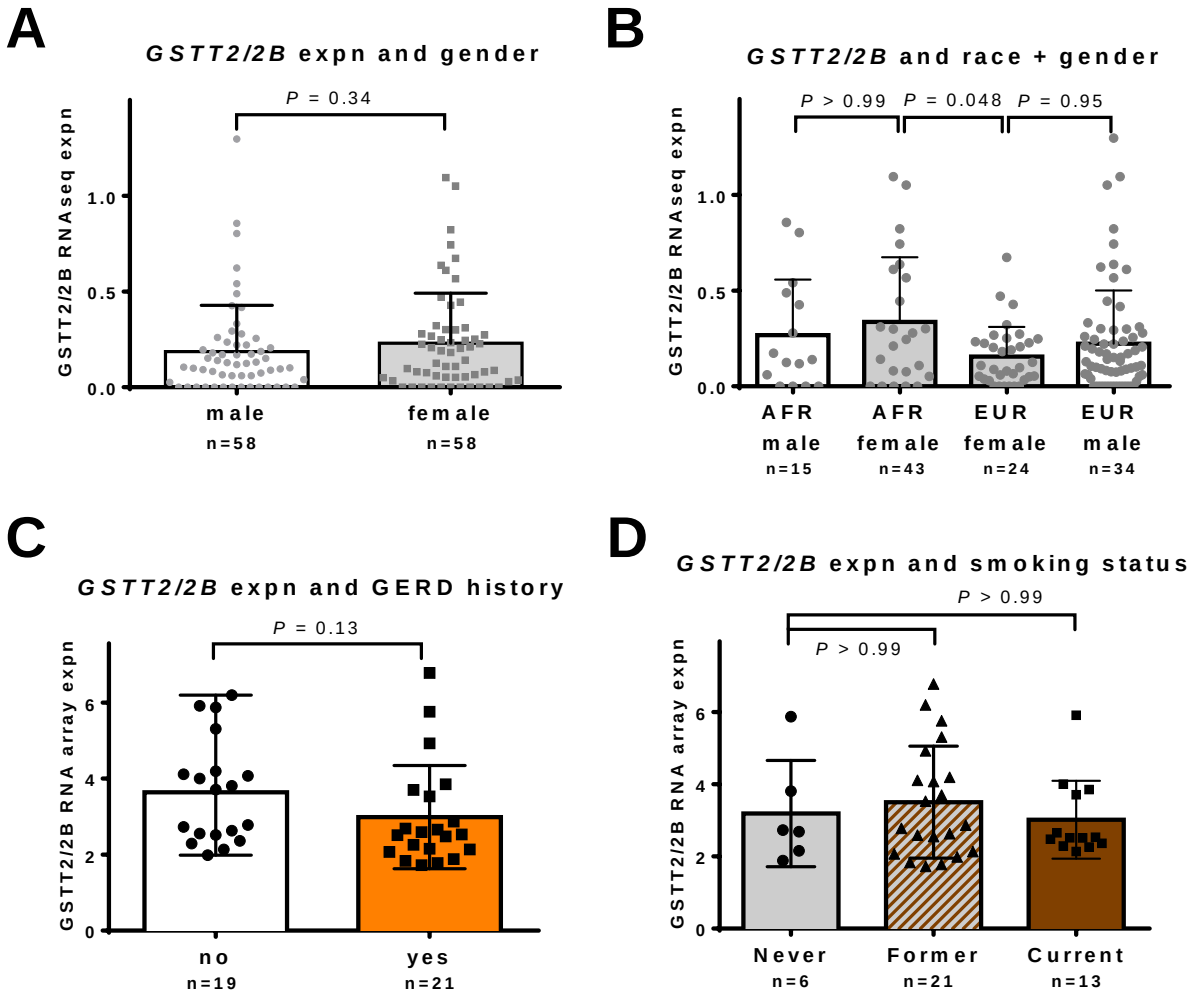
Supplementary Figure S2. Map of chromosome 22q11.1 region containing *GSTT2* and array analyses of neighboring genes. (A) *GSTT2* and **(B)** *GSTT2B* **(A)** array results of all four squamous (NE) tissue groups from AA and EA, showing differential expression between population control groups but not between disease groups. **(C)** Genomic map of the human *GSTT2B/GSTT2* region at chromosome 22q11.23 in hg19, including *GSTT* family members *GSTTP1* and *GSTT1*. **(B-D)** array analyses of the four squamous (NE) tissue groups from AA and EA indicate that neither *DDT* nor *GSTT* family members *GSTTP1* and *GSTT1* are not differentially expressed between control groups. One-way ANOVA with terms for the means of four groups to log-transformed gene expression data, as described in Materials and Methods.



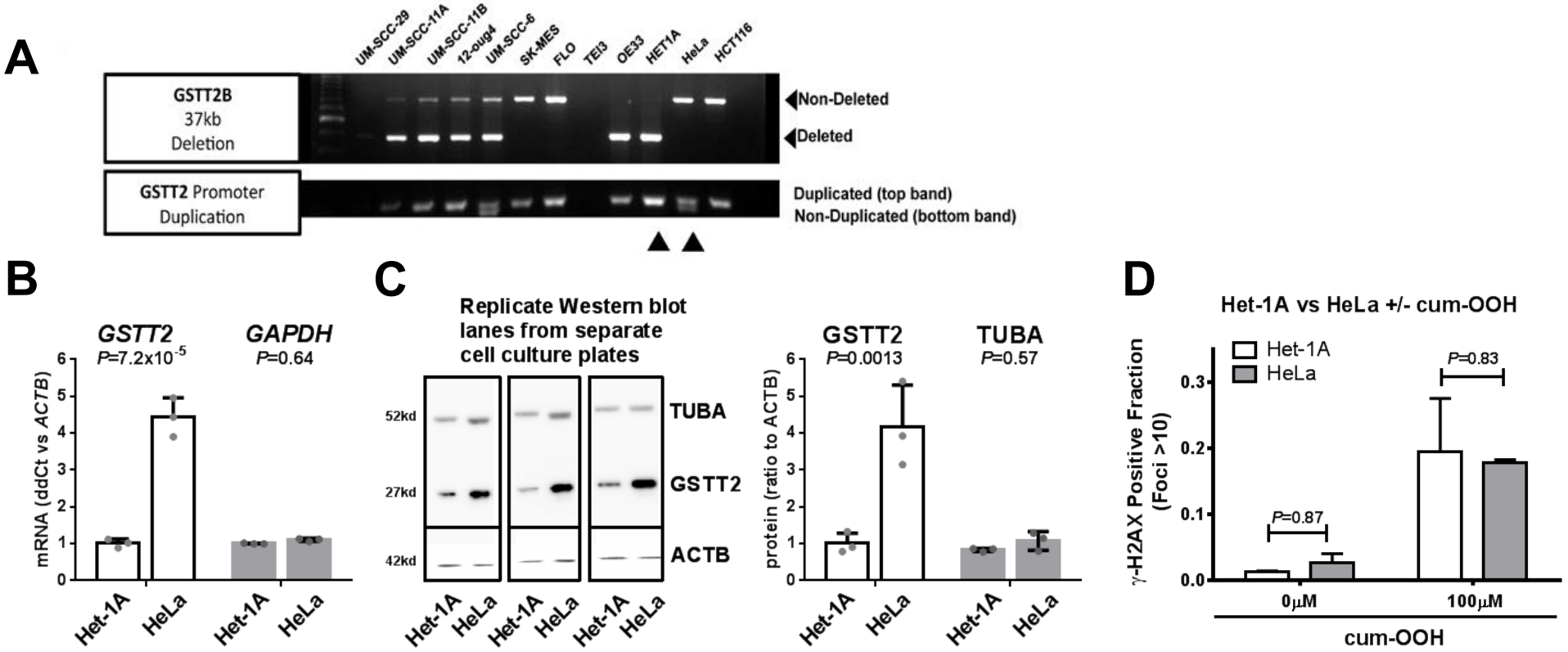
Supplementary Figure S3. Investigating *GSTT2B* deletion allele in esophageal squamous tissue samples from AA and EA, with and without a history of BE. (A) *GSTT2/2B* locus structure showing the 37kb deletion that removes the *GSTT2B* locus. "p" denotes gene promotor, highlighting the inverse orientation of the *GSTT2* genes. **(B)** Allele frequency of the 37kb *GSTT2B* deletion in NE of EA and AA with disease (BE) and without disease (NE). A trend towards an increased incidence of the deletion in EA was noted (two-sided Fisher Exact Test). Comparisons involving AA-NE:BE versus AA:NE or either EA group were not significant. The comparison between disease and non-disease EA groups was also not significant. **(C)** All arrayed NE samples combined show no evidence that *GSTT2B* deletion homozygosity (most common genotype; n=29) resulted in decreased expression relative to homozygous for the non-deleted genotype (n=4), or heterozygote (n=5) genotypes. Log2-adjusted *GSTT2* array expression results were not different when the 3 genotypes were compared with one-way ANOVA (*P*=0.65), nor pair-wise with Tukey post-hoc test for genotype combinations (0.24 0.55, 0.98 for '-/-' vs '-/+ & +/+', '-/-' vs '+/+' and '-/+' vs '-/- & +/+', respectively). Note that two samples failed to amplify for the *GSTT2B* allele test after three attempts. **(D)** The presence or absence of BE did not influence the effect of the *GSTT2B* deletion on *GSTT2/2B* mRNA expression (*P*=0.76 by one-way ANOVA), such that EA with either '+/+' or '-/+' genotypes had consistently lower expression relative to AA with the same genotype. ddCt qRT-PCR results are shown, where *ACTB* was used as the reference gene. Samples include 37 of the original 40 arrayed samples (1 mRNA and 2 DNAs failed to amplify, as noted) and an additional 17 samples, where an additional 2 samples failed to amplify for the *GSTT2* deletion genotype.



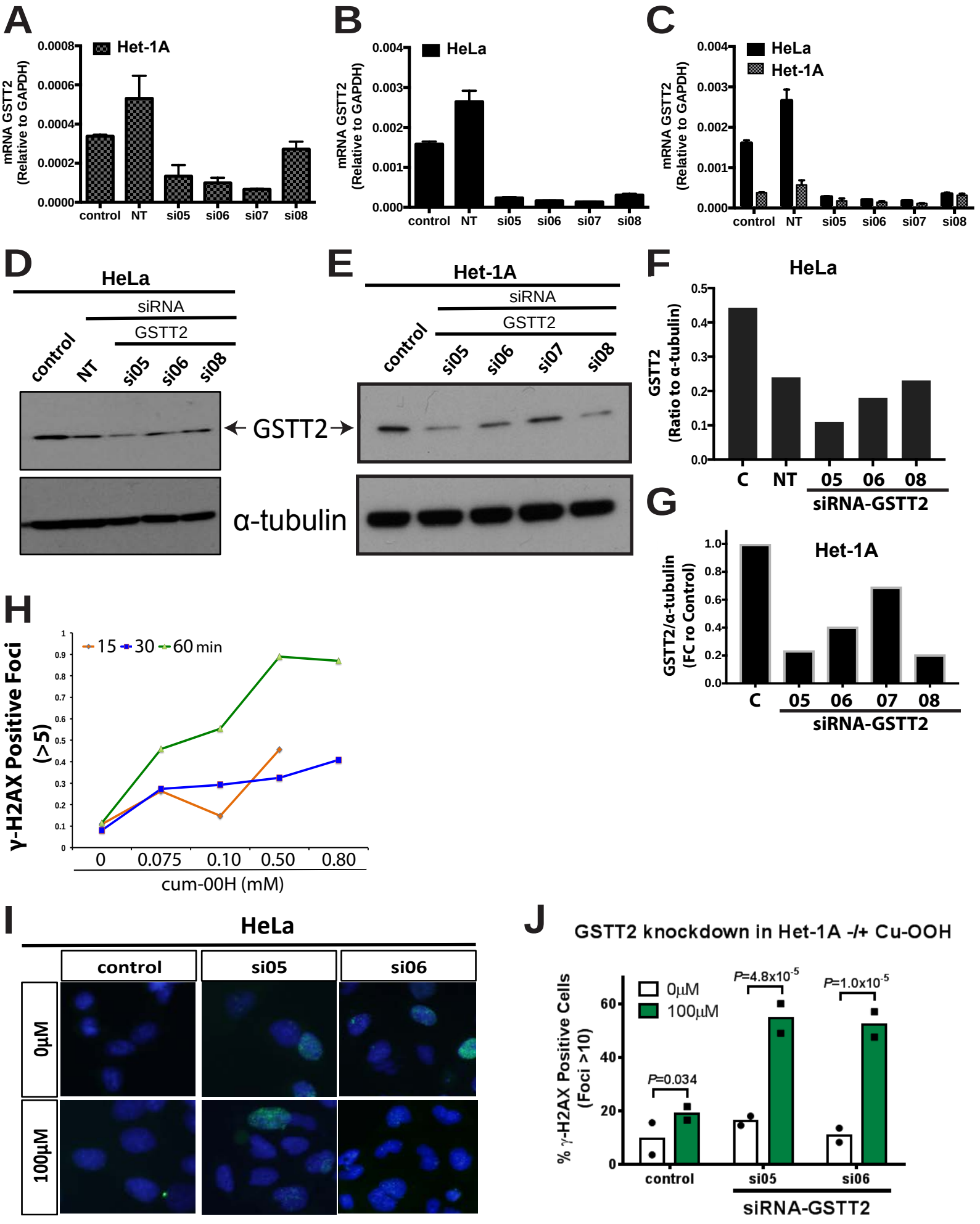
Supplementary Figure S4. Example 17bp *GSTT2* variant genotype sequences. Examples of the *GSTT2/2B* 17bp duplication genotypes using reverse strand sequences showing the alignment generated using Geneious and manual alignment of raw sequence traces.



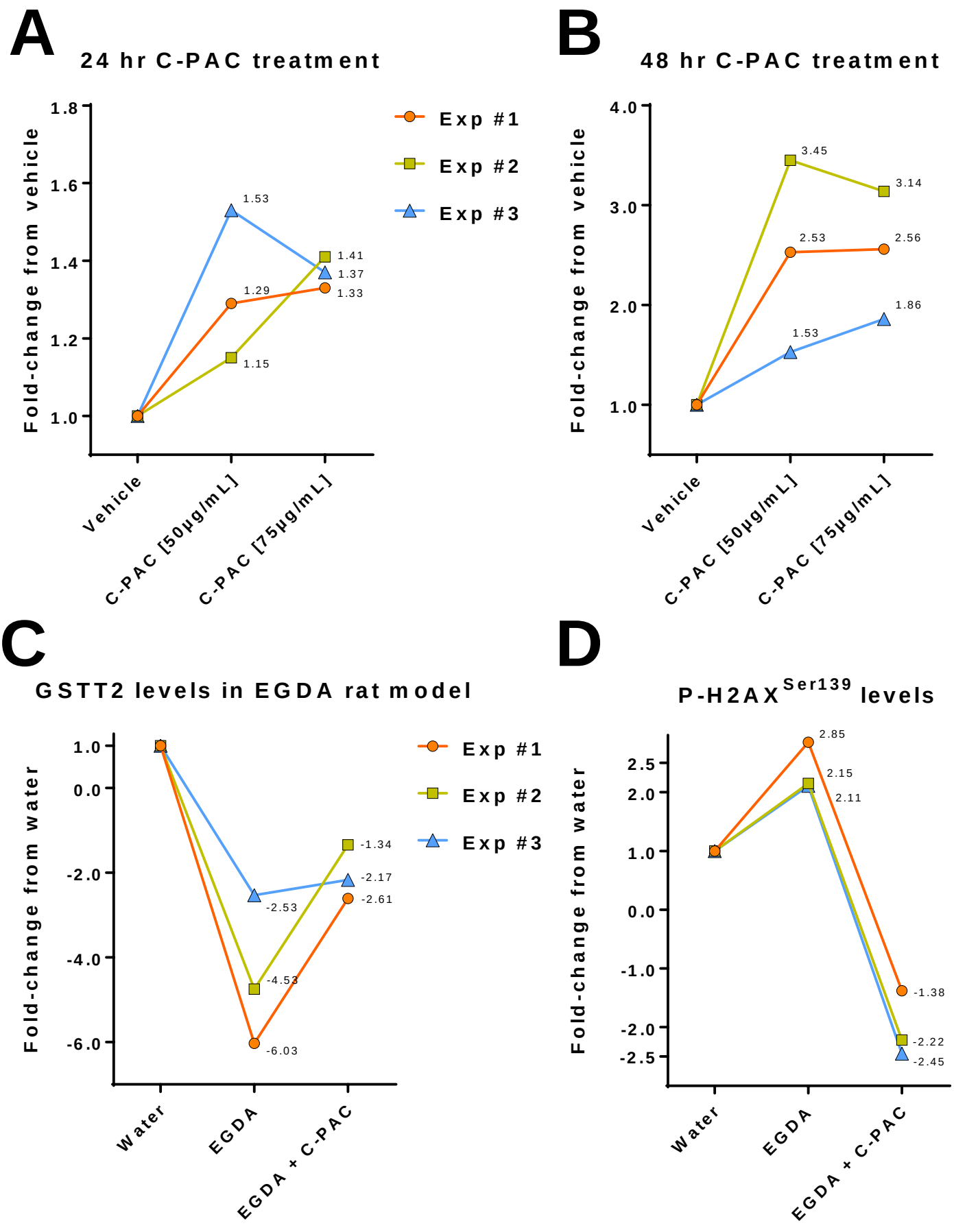
Supplementary Figure S5. Additional *GSTT2* mRNA analyses of potential confounding effects. Using 1000 Genomes population controls (n=116) we observe that differences in *GSTT2/2B* mRNA expression is not significantly different between **(A)** genders but were **(B)** race dependent, with a difference in expression levels seen between AFR and EUR females, but not males (*P*>0.99), nor between genders for either race as shown. RPKM values were obtained from previous 1000G/Hapmap publications, as described in the methods. Within the 40 arrayed NE samples differences in *GSTT2* expression were not explained in terms of either the presence of **(C)** GERD nor **(D)** smoking status where no association was seen with the overall ANOVA, nor pair-wise group tests, between *GSTT2* levels and cigarette usage. Log2-adjusted expression values for *GSTT2* were plotted in Prism, with one-way ANOVA was used in each panel. The Bonferroni post-hoc test was applied across the 4 and 3 pair-wise tests for race/gender and self-reported smoking status in panels B and D, respectively.



Supplementary Figure S6. Het-1A and HeLa susceptibility to DNA damage with and without Cu-OOH treatment. (**A**) *GSTT2*/2B 37kb deletion and promoter duplication genotyping in a cohort of 12 cell lines. Bottom triangles highlight the Het-1A and HeLa AA cell lines. (**B**) qRT-PCR and (**C**) western blots were performed on HeLa and Het-1A to measure endogenous GSTT2/2B levels, with mean and standard deviation of replicate plate expression comparisons, relative to ACTB mRNA or protein, with *GAPDH* and TUBA shown as similarly normalized housekeeping gene and protein respectively. Two-way ANOVA was used for both mRNA and protein comparisons between cell lines, with unadjusted two group comparson *P* values shown. **(D)** Quantification of γ-H2AX positive foci in Het-1A and HeLa cells, with and without cum-OOH treatment, with the average ratios (per plate matched γ-H2AX and DAPI cell counts) and standard deviations shown for two experiments, each consisting of duplicate measures. Nuclei with >10 positively-stained foci were considered as representing a cell positive for DNA damage. Presented measures are non-siRNA controls for the *GSTT2* knockdown studies presented in Sup Figure 7 to show the importance of GSTT2 in the cellular response to genotoxic stress. Therefore, *P* values comparing the level of DNA damage response between these cell lines, with and without cum-OOH, were derived from the 4-way ANOVA with terms of cum-OOH (0 vs 100uM), cell type (Het-1A vs HeLA), treatment (control, siRNA05, siRNA06) and experiment day, as detailed in Supplementary Methods. While both cell lines showed some DNA damage response following cum-OOH treatment, with *P*=0.034 and 15 fold for Het-1A and *P*=0.067 and 7 fold for HeLa, we saw no significant difference in the level of DNA damage response between these cell lines, either endogenous or cum-OOH induced, even though they have key differences in *GSTT2* genotypes. A two-tailed, paired t-test gave a similarly non-significant result (*P*=0.91) as the ANOVA model and fold changes were also nominal with 0.52 and 1.1 for the Het-1A/HeLa ratio of untreated and 100uM cum-OOH respectively.


A
Het-1A
mRNA GSTT2 (Relative to GAPDH)
0.0008
0.0006
0.0004
0.0002
0.0000
control
NT
si05
si06
si07
si08
B
HeLa
mRNA GSTT2 (Relative to GAPDH)
0.004
0.003
0.002
0.001
0.000
control
NT
si05
si06
si07
si08
C
HeLa
Het-1A
mRNA GSTT2 (Relative to GAPDH)
0.004
0.003
0.002
0.001
0.000
control
NT
si05
si06
si07
si08
D
HeLa
siRNA
GSTT2
control
NT
si05
si06
si08
GSTT2
α-tubulin
E
Het-1A
siRNA
GSTT2
control
si05
si06
si07
si08
F
HeLa
GSTT2 (Ratio to α-tubulin)
0.5
0.4
0.3
0.2
0.1
0.0
C
NT
05
06
08
siRNA-GSTT2
G
Het-1A
GSTT2/α-tubulin (FC ro Control)
1.0
0.8
0.6
0.4
0.2
0.0
C
05
06
07
08
siRNA-GSTT2
H
γ-H2AX Positive Foci (>5)
15
30
60 min
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.075
0.10
0.50
0.80
cum-00H (mM)
I
HeLa
control
si05
si06
0μM
100μM
J
GSTT2 knockdown in Het-1A -/+ Cu-OOH
% γ-H2AX Positive Cells (Foci >10)
0μM
100μM
P=0.034
P=4.8x10-5
P=1.0x10-5
60
40
20
0
control
si05
si06
siRNA-GSTT2

Supplementary Figure S7. Validation of *GSTT2* knockdown in Het-1A and HeLa cells. (**A-G**) Relative *GSTT2/2B* mRNA expression in Het-1A **(A, E, G)** and HeLa **(B, D, F)** cells after 48hr transfection with four siRNAs. Panel **C** shows the comparative knockdown between the two cell lines. (**H-J**) Quantification of positive foci for γ-H2AX in HeLa, with the averaged ratio for 2 experiments shown (2 replicates for each experiment, each with a matched γ-H2AX count). Nuclei with >10 positively-stained foci were considered as representing a cell positive for DNA damage. Values were normalized to the mean of the untreated control group, and two-way ANOVA, with Tukey post-hoc adjustment, was used to assess pair-wise comparisons between siRNA treatments and controls, and the cum-OOH treatment response. While data trend similar to Het-1A (Figure 4), *P* values for control and two siRNAs with and without cum-OOH were not significant (shown). Contrasts between 0μM cum-OOH treatments were also non-significant (P=0.32, 0.12 and 0.94 for comparisons of control to siRNA05, control to siRNA6 and siRNA05 to siRNA06 respectively) and only control to siRNA6 was significant (P=0.0092) at 100μM cum-OOH, but not control to siRNA05 or siRNA05 to siRNA06 (P=0.60 and 0.45, respectively).


A
24 hr C-PAC treatment
Fold-change from vehicle
1.8
1.6
1.4
1.2
1.0
Exp #1
Exp #2
Exp #3
1.53
1.41
1.37
1.29
1.33
1.15
Vehicle
C-PAC [50µg/mL]
C-PAC [75µg/mL]
B
48 hr C-PAC treatment
Fold-change from vehicle
4.0
3.0
2.0
1.0
3.45
3.14
2.53
2.56
1.86
1.53
Vehicle
C-PAC [50µg/mL]
C-PAC [75µg/mL]
C
GSTT2 levels in EGDA rat model
Fold-change from water
1.0
0.0
-2.0
-4.0
-6.0
Exp #1
Exp #2
Exp #3
-1.34
-2.17
-2.61
-2.53
-4.53
-6.03
Water
EGDA
EGDA + C-PAC
D
P-H2AX Ser139 levels
Fold-change from water
2.5
2.0
1.0
0.0
-1.0
-2.0
-2.5
2.85
2.15
2.11
-1.38
-2.22
-2.45
Water
EGDA
EGDA + C-PAC

Supplementary Figure S8. Westerns blot results of relative fold-change across triplicate experiments for C-PAC treatment of Het-1A and EGDA rats. Relative protein ratios (GSTT2 vs HSP60) following **(A)** 24 and **(B)** 48hr C-PAC treatment of Het-1A cells. Each ratio (GSTT2/HSP90) was normalized to within-experiment vehicle (set to 1), with relative ratio values shown for each experimental point. Western blots and statistics are presented in **Figure 4A-B**. **(C-D)** Western blot results showing C-PAC mitigation of GSTT2 loss coincides with protection against the DNA damaging effects of gastroduodenal reflux in an esophagogastroduodenal anastomosis (EGDA) rat model, with relative esophageal protein levels of **(C)** GSTT2 and **(D)** Phospho-H2AXSer139 protein levels (γ-H2AX) contrasted with and without C-PAC treatment. Ratios (HSP60 as reference) for GSTT2 and γ-H2AX were normalized to the assigned reference (water only) animal (ration set to 1), with relative ratio values shown for each experimental point. Western blots and statistics are presented in **Figure 4C-B**.

Supplemental Array Analyses

In order to investigate comparative expression profiles between our 4 arrayed NE sample groups we employed principal component analysis (PCA) to investigate variance distribution, as well as one-way ANOVA with unprotected Fisher's Least Significant Difference (LSD) (Fisher 1935; Hochberg and Tamhane 1987), as implemented in Prism 7 by Graphpad Software) for the 4 groups, combined with mean-group fold-change comparisons. Plots of the first 5 principal components (**Insert 1**) showed a clear separation of EA-NE:BE from the other 3 groups by PCA component 1, with poor separation between the other 3 groups (**Insert 1**). For the ANOVA analysis we only conducted the four key inter-group comparisons: between population control groups (EA-NE vs AA-NE), between disease groups (EA-NE:BE vs AA-NE:BE) and the normal-disease comparisons within each population (EA-NE:BE vs EA-NE and AA-NE:BE vs AA-NE). Using a threshold of $P<0.01$ we identified 375, 8,983, 10,304 and 769 differentially expressed genes for the EA-NE vs AA-NE, EA-NE:BE vs AA-NE:BE, EA-NE:BE vs EA-NE and AA-NE:BE vs AA-NE comparisons, respectively. We estimated false discovery rates of 0.61, 0.027, 0.023 and 0.31 for these comparisons, respectively, by analyzing 10,000 data sets in which the sample labels were randomly permuted, and averaged the number of qualifying probe-sets across the 10,000 data sets. These results suggest that the two comparisons involving the EA-NE:BE group resulted in significantly more genes that were expected by change.

A VENN diagram (Heberle et al. 2015) of differential genes obtained by between-group one-way ANOVA comparisons ($P<0.01$) shows that a majority of these genes (96.9%) involved EA-NE:BE comparisons to other groups (**Insert 2**). Thus both at the individual sample, and sample-groups levels, through supervised and unsupervised comparisons, the EA-NE:BE cohort had very distinct expression profiles.

While initially a surprise, there is a key difference between EA-NE:BE and AA-NE:BE groups that explains our results. All 8 EA-NE:BE individuals had progressed to cancer, as compared to 1 out of 8 AA-NE:BE individuals. Moreover, significant mRNA expression differences in esophageal squamous tissue from cancer patients has previously been noted, as compared to both BE patients who have not progressed, and individuals with no esophageal disease history (Brabender et al. 2005; Saadi et al. 2010). These manuscripts compared biopsied squamous material collected at least 4 cm from primary esophageal lesions or control subjects, and showed overlapping, but discreet, sets of expression changes in the squamous from subjects with cancer. We ranked the two sets of genes identified by these studies (Brabender et al. 2005; Saadi et al. 2010) according to how well they discriminated EA-NE:BE from our other sample groups (by fold-change and *P* values, as presented in **Insert Table 1**) as shown by heatmap in **Insert 3.** The Cox-2 gene (*PTGS2*) was over-expressed among the squamous samples from cancer patients, relative to non-EAC subjects, in all 3 studies (**Insert 3**, Brabender et al. 2005; Saadi et al. 2010). Cox-2 is a key enzyme in the production of prostaglandins, that are chiefly responsible for localized immune responses (reviewed by Ricciotti and Fitzgerald 2011). Increased Cox-2 levels have been a feature of EAC (reviewed in Liao et al. 2012), and other cancer studies (reviewed by Liu, Qu, and Yan 2015) and it is believed that a chronic reduction in PGE2 levels via Cox-2 modulation is the chief benefit of NSAID treatment linked to reduced cancer risk. Excluding *PTGS2*, we plotted averaged Z-scores for the five next best genes from each of the two previous studies against each other to highlight that these focused gene sets can also discriminate our squamous from cancer subset (EA-NE:BE) from other squamous groups. Six of the ten genes identified by Saadi and coworkers were strongly increased in the stroma of our cancer patient vs control groups (FC>2 with $P< 0.01$ in both EA-NE:BE vs AA-NE:BE and EA-NE:BE vs EA-NE comparisons; **Insert Table 1**). Similarly, 7/12 genes originally reported by Brabender et al., as significantly different between the squamous mucosa of cancer and non-cancer subjects, discriminated EA-NE:BE in our data ($P<0.01$), although only five of these showed the same fold-change direction in our cohort (**Insert Table 1**). As noted by Saadi and co-workers in their analysis of several older datasets, some of these differential NE-expressing genes (*PTGS2, SPARC, TSPAN1, TSPAN8 and MMP7* rather than *MMP1*) have shown increased expression in tissues spanning the EAC-related progression from normal squamous (from

individuals with no esophageal disease history) through BE to EAC tumor tissue (Botelho et al. 2010; Saadi et al. 2010), through many genes show differential expression between these tissues. Thus, in our data many genes distinguished the EA-NE:BE sample group from the other NE sample groups (**Insert 2**), including a subset of genes that overlap with previous studies (**Insert 3**) to highlight existence of an esophageal tissue field effect associated with disease progression. Several studies have observed this progression-related field effect with metabolic profiling (Yakoub et al. 2010; Reed et al. 2017). In addition, gene ontology analysis (DAVID; Huang et al. 2007) of genes over-expressed in the EA-NE:BE sample group (ANOVA $P<0.01$ and FC>2), as clustered by REVIGO (Supek et al. 2011), show over-representation of extracellular matrix reorganization, cell adhesion and angiogenesis (**Insert 4**), further suggesting the presence of a cancer field effect, or perhaps more accurately etiological field effect (Lochhead et al. 2015) where a combination of risk factor exposure and risk factor susceptibility resulted in chronic esophageal tissue damage culminating in the formation of EAC. We believe this diffuse expression signal is evidence for the chronic changes to the lower esophagus including the surrounding squamous tissue, as previously suggested (Saadi et al. 2010; Lochhead et al. 2015). It is for this reason that we weighted our initial energies towards the investigation of differences between our normal squamous sample groups, with only supporting evidence from our disease groups. By far the best of these candidates was *GSTT2*.

References:

Botelho, N. K., F. I. Schneiders, S. J. Lord, A. K. Freeman, S. Tyagi, D. J. Nancarrow, N. K. Hayward, D. C. Whiteman, and R. V. Lord. 2010. 'Gene expression alterations in formalin-fixed, paraffin-embedded Barrett esophagus and esophageal adenocarcinoma tissues', *Cancer Biol Ther*, 10: 172-9.

Brabender, J., P. Marjoram, R. V. Lord, R. Metzger, D. Salonga, D. Vallbohmer, H. Schafer, K. D. Danenberg, P. V. Danenberg, F. M. Selaru, S. E. Baldus, A. H. Holscher, S. J. Meltzer, and P. M. Schneider. 2005. 'The molecular signature of normal squamous esophageal epithelium identifies the presence of a field effect and can discriminate between patients with Barrett's esophagus and patients with Barrett's-associated adenocarcinoma', *Cancer Epidemiol Biomarkers Prev*, 14: 2113-7.

Fisher, R.A. 1935. *The Design of Experiments* (Oliver and Boyd: Edinburgh).

Heberle, H., G. V. Meirelles, F. R. da Silva, G. P. Telles, and R. Minghim. 2015. 'InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams', *BMC Bioinformatics*, 16: 169.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple comparison procedures* (John Wiley \& Sons, Inc.).

Huang, D. W., B. T. Sherman, Q. Tan, J. R. Collins, W. G. Alvord, J. Roayaei, R. Stephens, M. W. Baseler, H. C. Lane, and R. A. Lempicki. 2007. 'The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists', *Genome Biol*, 8: R183.

Liao, L. M., T. L. Vaughan, D. A. Corley, M. B. Cook, A. G. Casson, F. Kamangar, C. C. Abnet, H. A. Risch, C. Giffen, N. D. Freedman, W. H. Chow, S. Sadeghi, N. Pandeya, D. C. Whiteman, L. J. Murray, L. Bernstein, M. D. Gammon, and A. H. Wu. 2012. 'Nonsteroidal anti-inflammatory drug use reduces risk of adenocarcinomas of the esophagus and esophagogastric junction in a pooled analysis', *Gastroenterology*, 142: 442-52.

Liu, Bing, Liyan Qu, and Shigui Yan. 2015. 'Cyclooxygenase-2 promotes tumor growth and suppresses tumor immunity', *Cancer Cell International*, 15: 106.

Lochhead, P., A. T. Chan, R. Nishihara, C. S. Fuchs, A. H. Beck, E. Giovannucci, and S. Ogino. 2015. 'Etiologic field effect: reappraisal of the field effect concept in cancer predisposition and progression', *Mod Pathol*, 28: 14-29.

Reed, M. A., R. Singhal, C. Ludwig, J. B. Carrigan, D. G. Ward, P. Taniere, D. Alderson, and U. L. Gunther. 2017. 'Metabolomic Evidence for a Field Effect in Histologically Normal and Metaplastic Tissues in Patients with Esophageal Adenocarcinoma', *Neoplasia*, 19: 165-74.

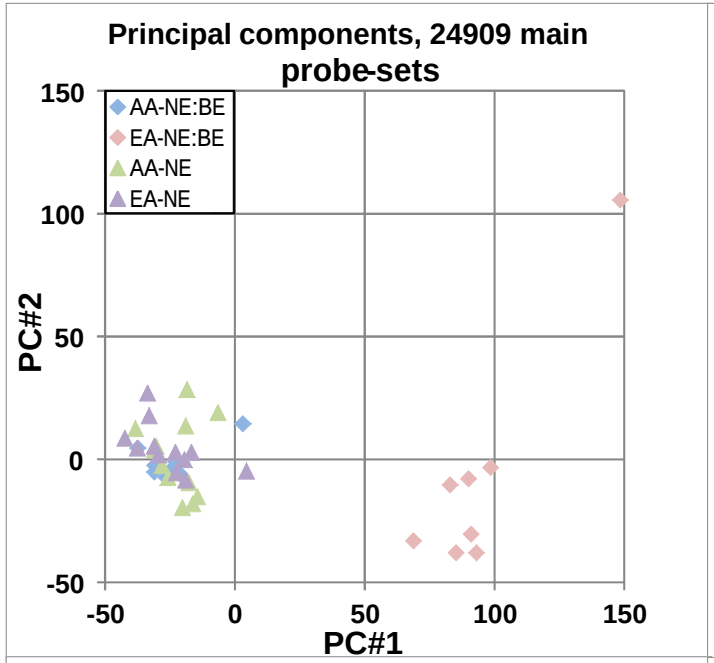
Ricciotti, E., and G. A. Fitzgerald. 2011. 'Prostaglandins and inflammation', *Arterioscler Thromb Vasc Biol*, 31: 986-1000.

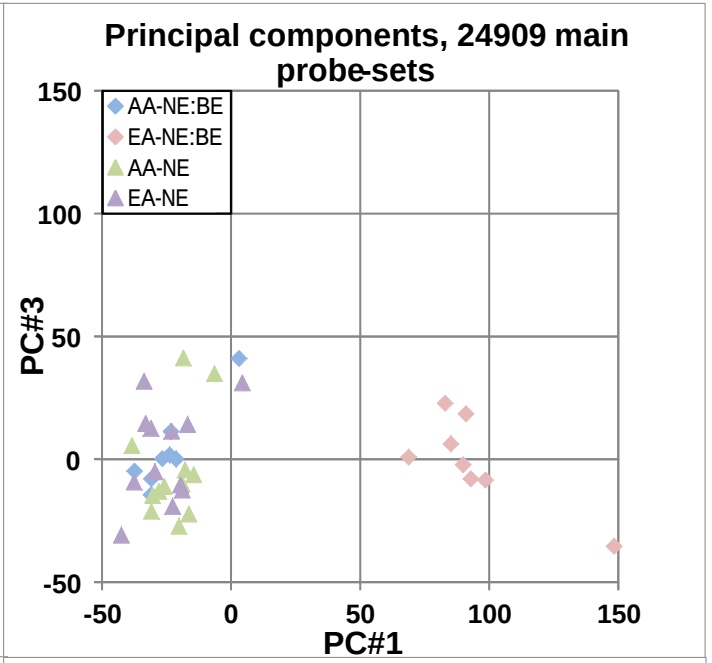
Saadi, A., N. B. Shannon, P. Lao-Sirieix, M. O'Donovan, E. Walker, N. J. Clemons, J. S. Hardwick, C. Zhang, M. Das, V. Save, M. Novelli, F. Balkwill, and R. C. Fitzgerald. 2010. 'Stromal genes

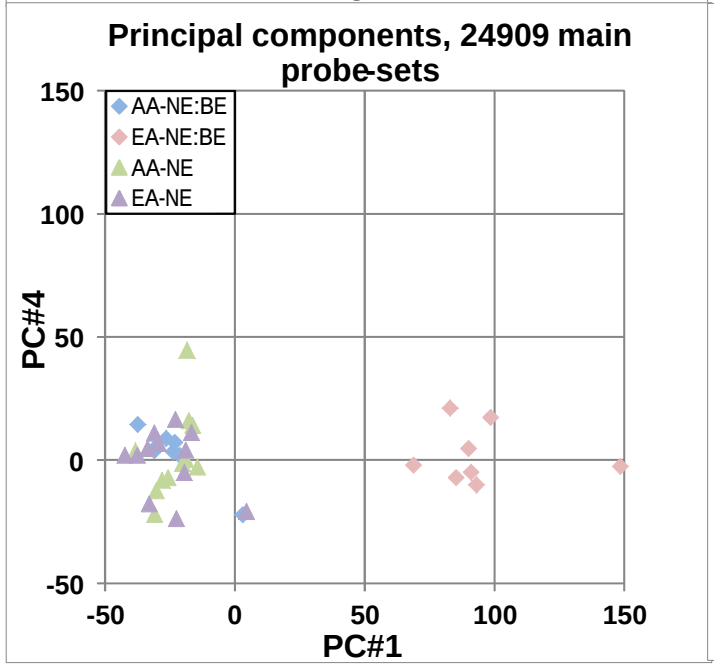
discriminate preinvasive from invasive disease, predict outcome, and highlight inflammatory pathways in digestive cancers', *Proc Natl Acad Sci U S A*, 107: 2177-82.

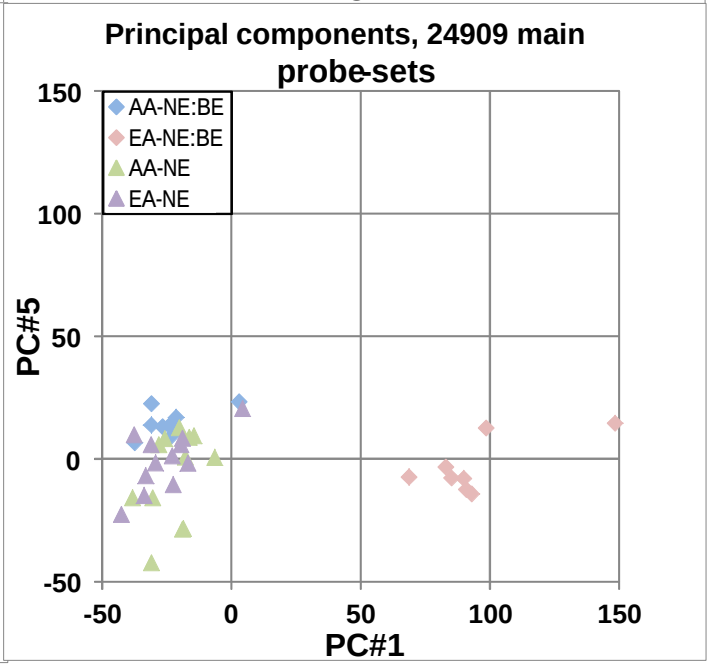
Supek, F., M. Bosnjak, N. Skunca, and T. Smuc. 2011. 'REVIGO summarizes and visualizes long lists of gene ontology terms', *PLoS One*, 6: e21800.

Yakoub, D., H. C. Keun, R. Goldin, and G. B. Hanna. 2010. 'Metabolic profiling detects field effects in nondysplastic tissue from esophageal cancer patients', *Cancer Res*, 70: 9129-36.

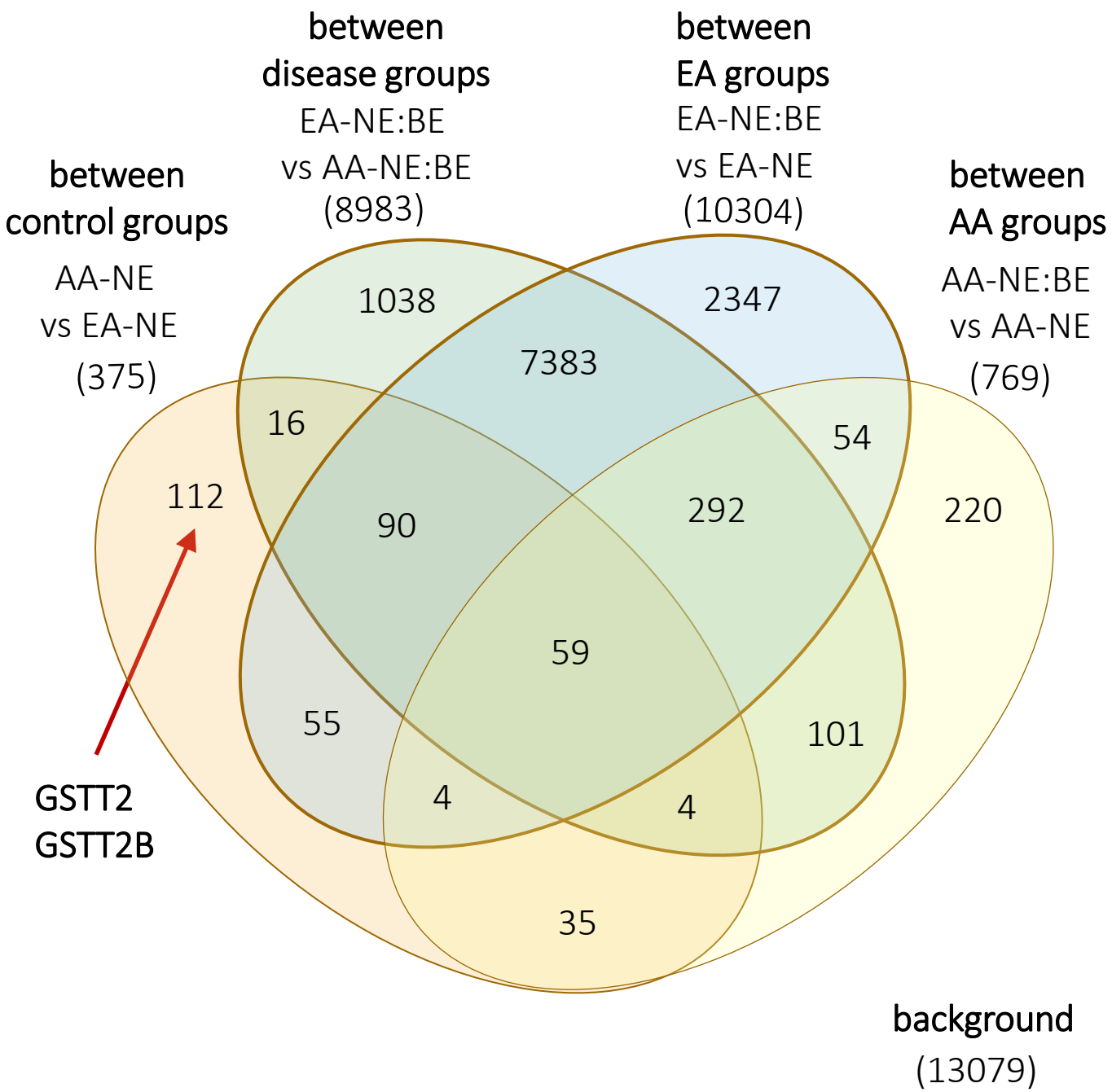




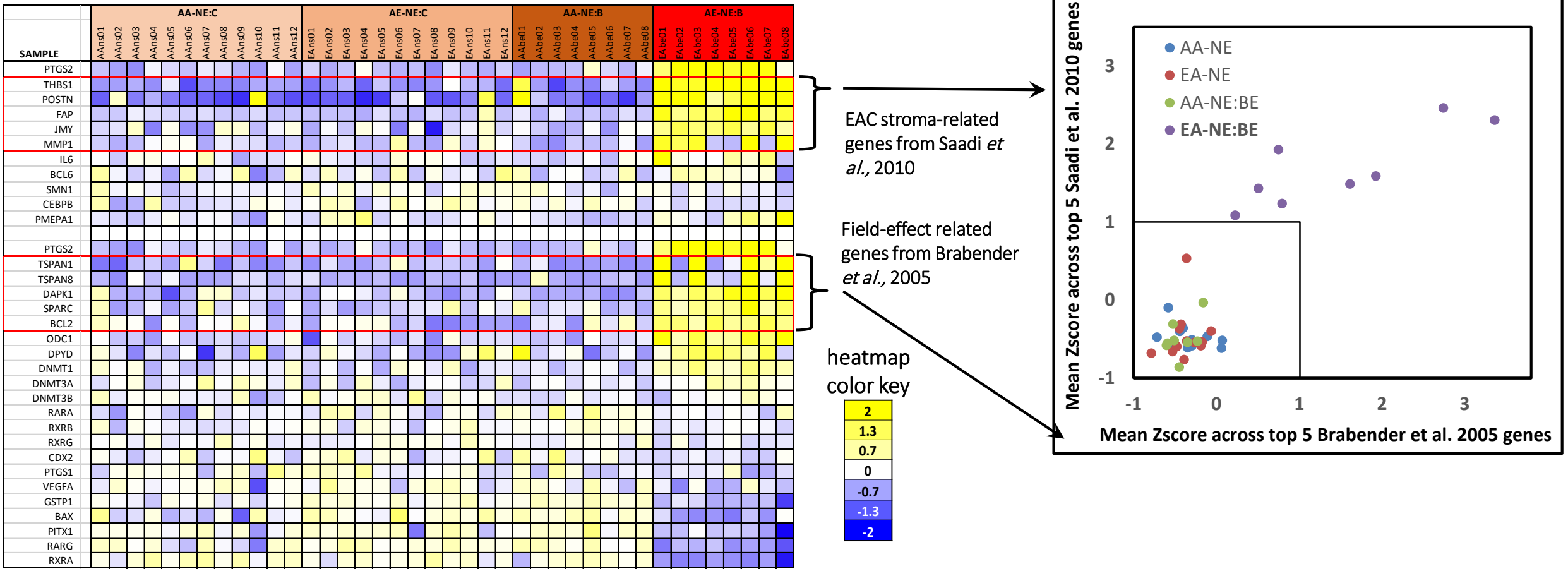




INSERT 1: Principal component analysis showed that EA-NE:BE individuals were distinct. PCA was conducted across all 24,909 probesets using mean-normalized expression. The top five unsupervised PCA components are shown where components 2, 3, 4 and 5 have been compared to the most variable component. In each case EA-NE:BE group form a distinct cluster compared to the other sample 3 groups, but chiefly through the PC#1 component.



INSERT 2: Paired between-group ANOVA shows that the EA-NE:BE is a distinct group. A breakdown of transcript numbers across the 4 between-group comparisons we conducted, with one-way ANOVA with Fisher's LDS (*P*>0.01) for; AA-NE vs EA-NE, EA-NE:BE vs AA-NE:BE, EA-NE:BE vs EA-NE, & AA-NE:BE vs AA-NE. Results show that 48.5% (24,909-13,079=11,830) of transcripts showed differences for one or more of the four comparisons, with 96.9% (11,463) involving the EA-NE:BE disease group. 64.4% (7,383) of these were common to population (EA-NE:BE vs EA-NE) and disease (EA-NE:BE vs AA-NE:BE) comparisons involving this group. Venn diagram was produced with InteractiVenn (Heberle et al. 2015).



INSERT 3: Analysis of previously identified genes differentially expressed in EAC squamous. In contrast to the 5 genes identified in AA-NE vs EA-NE (**Supplementary Table S2**), we show expression differences for genes previously reported as associated with an EAC-related tissue field-effect, are over-expressed in EA-NE:BE. Log2 expression values for genes reported by Brabender et al., 2005 and Saadi et al., 2010 were mean normalized to generate sample-group orientated (supervised) heat-maps using our Human Gene ST 2.1 expression data. Combining the most overexpressed of these genes from each study spatially delineated the 8 EA-NE:BE samples from the other 32 squamous samples. **Insert Table 1** shows the associated ANOVA *P* values and fold change ratios, relative to AE-NE:BE, as well as within-gene permutation-based *P* values.

Insert Table 1. Genes identified by Saadi et al., 2010 and Brabender et al., 2005 to be associated with an EAC progression field effect

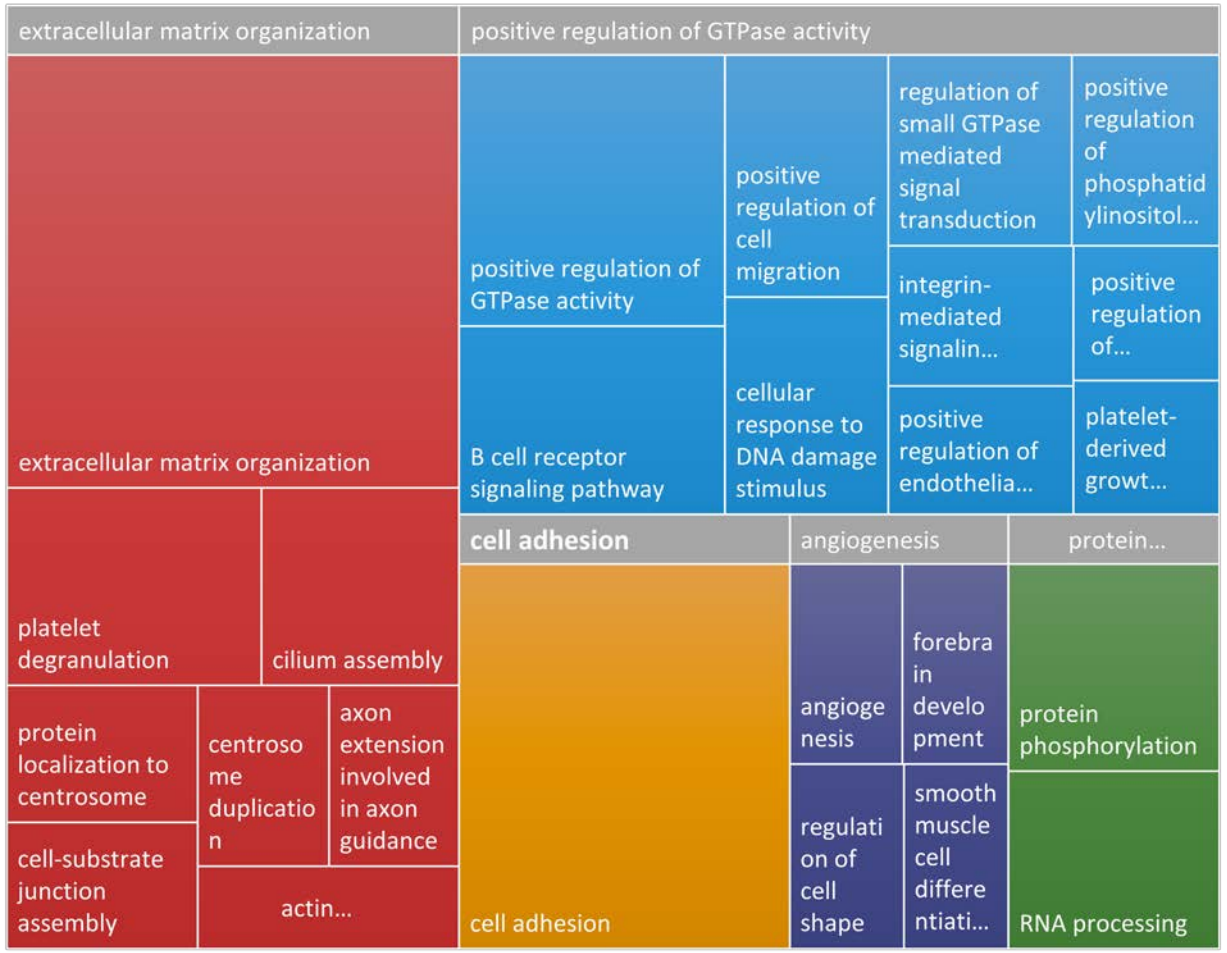
			ANOVA *P*-value*				Fold change^			
Symbol	Gene Name	Entrez Gene	AA-NE vs EA-NE	EA-NE:BE vs AA-NE:BE	AA-NE:BE vs AA-NE	EA-NE:BE vs EA-NE	AA-NE / EA-NE	EA-NE:BE / AA-NE:BE	AA-NE:BE / AA-NE	EA-NE:BE / EA-NE
Saadi et al 2010 - genes discriminating stroma of BE from stroma of EAC										
PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	5743	0.99112	2.93E-11	0.6219	6.99E-13	1.00	4.47	1.07	4.81
THBS1	thrombospondin 1	7057	0.87363	1.73E-13	0.4444	2.88E-15	0.97	11.52	1.16	13.03
POSTN	periostin, osteoblast specific factor	10631	0.85933	8.95E-05	0.4118	1.20E-06	1.07	8.71	1.45	13.58
FAP	fibroblast activation protein, alpha	2191	0.89871	3.38E-14	0.8909	2.41E-15	1.01	3.83	0.99	3.82
JMY	junction mediating and regulatory protein, p53 cofactor	133746	0.51382	1.19E-06	0.6321	1.56E-07	1.10	2.68	0.93	2.73
MMP1	matrix metallopeptidase 1 (interstitial collagenase)	4312	0.78121	1.25E-04	0.5744	4.12E-04	0.93	3.61	0.86	2.89
IL6	interleukin 6 (interferon, beta 2)	3569	0.79878	0.0024	0.4933	0.0036	1.03	1.63	0.91	1.53
BCL6	B-cell CLL/lymphoma 6	604	0.69507	0.2937	0.8822	0.5105	0.95	1.20	0.98	1.11
SMN1	survival of motor neuron 1, telomeric	6606	0.05156	0.5330	0.5619	0.0648	0.86	0.94	1.05	0.85
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta	1051	0.04566	0.0776	0.1691	0.0198	0.81	1.25	1.17	0.76
Brabender et al 2005 - genes discriminating squamous of non-disease subjects from squamous of patients with EAC										
PTGS2	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	5743	0.99112	2.93E-11	0.6219	6.99E-13	1.00	4.47	1.07	4.81
TSPAN1	tetraspanin 1	10103	0.76880	8.07E-05	0.8489	8.90E-05	0.89	7.89	0.92	6.50
TSPAN8	tetraspanin 8	7103	0.77094	2.06E-04	0.9217	3.89E-05	1.11	6.43	0.96	6.87
DAPK1	death-associated protein kinase 1	1612	0.93683	9.81E-10	0.2223	3.06E-09	1.01	5.17	0.80	4.17
SPARC	secreted protein, acidic, cysteine-rich (osteonectin)	6678	0.75082	3.06E-06	0.9854	2.38E-07	1.04	2.35	1.00	2.45
BCL2	B-cell CLL/lymphoma 2	596	0.03985	1.54E-06	0.4506	8.91E-09	1.30	2.40	0.90	2.81
ODC1	ornithine decarboxylase 1	4953	0.70043	4.00E-09	0.8569	2.93E-10	1.05	3.29	0.97	3.37
DPYD	dihydropyrimidine dehydrogenase	1806	0.54122	0.0114	0.5403	0.0051	0.89	1.90	1.15	1.93
DNMT1	DNA (cytosine-5-)-methyltransferase 1	1786	0.66680	1.16E-06	0.3490	1.17E-06	1.04	1.87	0.91	1.77
DNMT3A	DNA (cytosine-5-)-methyltransferase 3 alpha	1788	0.09592	0.0046	0.0528	0.0075	1.16	1.39	0.82	1.33
DNMT3B	DNA (cytosine-5-)-methyltransferase 3 beta	1789	0.15808	0.1548	0.3355	0.5048	0.88	1.17	0.91	0.93
RARA	retinoic acid receptor, alpha	5914	0.06717	0.0570	0.0021	0.6053	0.83	0.78	1.46	0.94
RXRB	retinoid X receptor, beta	6257	0.15916	0.0080	0.0117	0.0966	0.91	0.80	1.21	0.88
RXRG	retinoid X receptor, gamma	6258	0.54726	0.1008	0.9431	0.0188	0.97	0.89	1.00	0.85
CDX2	caudal type homeobox 2	1045	0.51735	0.0622	0.3408	0.0926	0.93	0.77	1.13	0.81
PTGS1	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	5742	0.84660	0.2682	0.4843	0.0416	0.98	0.86	0.92	0.77
VEGFA	vascular endothelial growth factor A	7422	0.50625	0.0433	0.5319	0.0297	0.92	0.73	1.09	0.74
GSTP1	glutathione S-transferase pi 1	2950	0.02690	2.09E-09	0.0018	1.08E-08	0.85	0.51	1.30	0.56
BAX	BCL2-associated X protein	581	0.00723	2.23E-05	0.0109	8.18E-06	0.71	0.49	1.43	0.50
PITX1	paired-like homeodomain 1	5307	0.74656	1.38E-04	0.2420	5.80E-04	0.95	0.45	1.22	0.53
RARG	retinoic acid receptor, gamma	5916	0.06468	9.99E-08	0.0672	2.55E-08	0.83	0.46	1.22	0.47
RXRA	retinoid X receptor, alpha	6256	0.35088	2.86E-09	0.4109	3.42E-10	1.09	0.43	0.92	0.43

**P*-value from ANOVA comparison of AA-NE and EA-NE normalized log2 Affy ST 2.1 expression data, with 4 group correction but not multiple testing correction

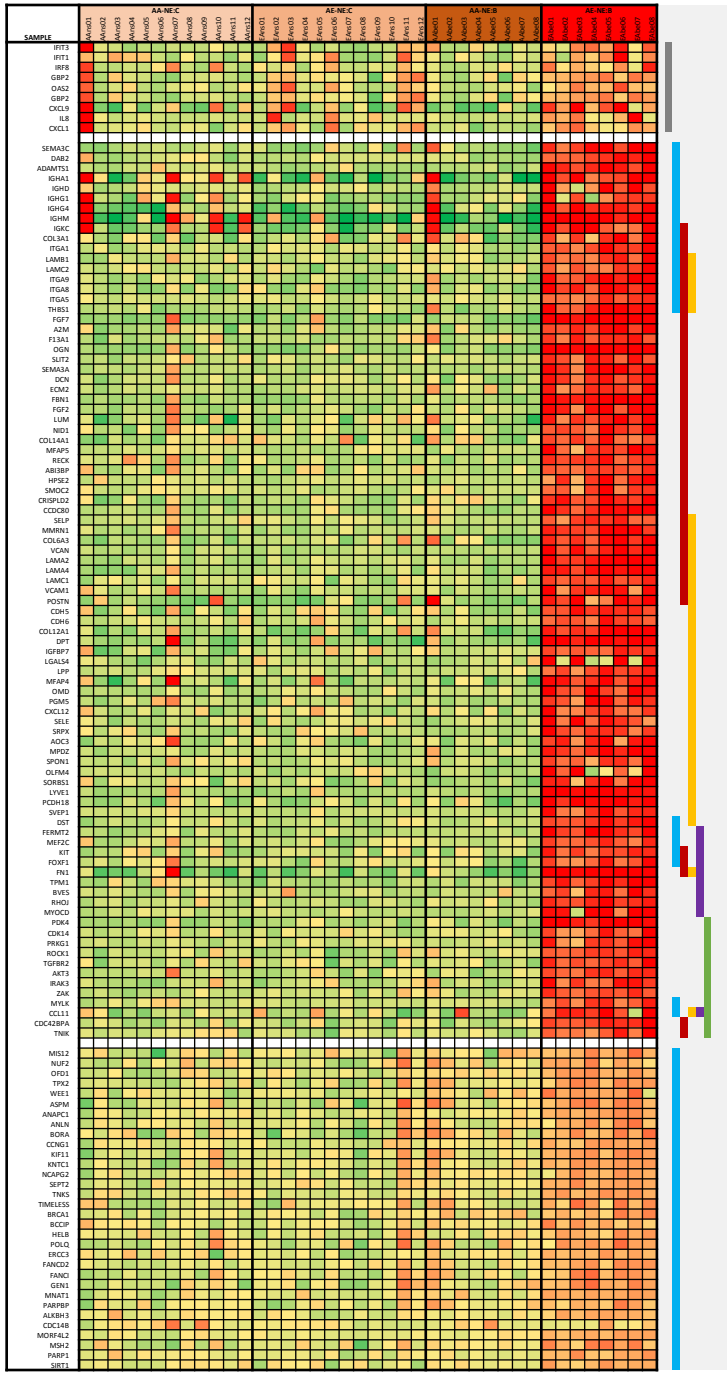
^Mean group fold changes (FC) were calculated and presented in non-log values

EA-NE:BE vs AA-NE:BE only

EA-NE:BE vs AA-NE:BE and EA-NE:BE vs EA-NE



EA-NE:BE vs EA-NE only



INSERT 4: Gene ontologies over-represented in the EA-NE:BE sample group. The most upregulated transcripts (ANOVA *P*<0.01 and FC >2-fold) common to both EA-NE:BE vs AA-NE:BE and EA-NE:BE vs EA-NE comparisons were interrogated via the DAVID web-application to identify the most over-represented GO ontologies (Benjamini adjusted *P*<0.05), visualized using REVIGO (Supek et al 2011). As expected, based on the high number of shared genes (Insert 2) the majority of ontologies were shared between the two comparisons, but with immune-related gene groups differential between the two disease populations, and proliferation/DNA damage ontologies more highly expressed in BE/EAC squamous compared to controls from the same race. We then used whole cohort mean normalized log2 expression to generate a heat-map of the differentially expressed from each of the significant gene ontologies.