MANUSCRIPT

AI-assisted Discovery of an Ethnicity-influenced Driver of Cell Transformation in Esophageal and Gastroesophageal Junction Adenocarcinomas

Authors: Pradipta Ghosh[1-4*†], Vinicius J. Campos[5*], Daniella T. Vo[6], Caitlin Guccione[7], Vanae Goheen-Holland[1, 3], Courtney Tindle[1, 3], Guilherme S. Mazzini[5, 8], Yudou He[1, 4], Ludmil Alexandrov[1, 4], Scott M. Lippman[2, 4], Richard R. Gurski[5,8,9], Soumita Das[3, 10], Rena Yadlapati[2*], Kit Curtius[2, 7*], and Debashis Sahoo[4, 6, 11*†]

Affiliations: [1]Department of Cellular and Molecular Medicine, University of California San Diego; [2] Department of Medicine, University of California San Diego; [3] HUMANOID Center of Research Excellence (CoRE), University of California San Diego; [4] Moore's Comprehensive Cancer Center, University of California San Diego; [5] Department of Gastrointestinal Surgery, Hospital de Clínicas de Porto Alegre, 2350 Ramiro Barcellos Street, Porto Alegre, RS, 90035-003; [6] Department of Pediatrics, University of California San Diego; [7] Division of Biomedical Informatics, University of California San Diego; [8] Postgraduate Program in Medicine; Surgical Sciences, Federal University of Rio Grande do Sul, 2400 Ramiro Barcellos Street, Porto Alegre, RS, 90035-003; [9] Medical School of Federal University of Rio Grande do Sul 2400 Ramiro Barcellos Street, Porto Alegre, RS, 90035-003; [10] Department of Pathology, University of California San Diego; [11] Department of Computer Science and Engineering, Jacob's School of Engineering, University of California San Diego.

[†]Equal contributions.

* CORRESPONDING AUTHOR CONTACT INFORMATION

Debashis Sahoo, Ph.D.; Associate Professor, Department of Pediatrics, University of California San Diego; 9500 Gilman Drive, MC 0703, Leichtag Building 132; La Jolla, CA 92093-0703. **Phone:** 858-246-1803: **Fax:** 858-246-0019: **Email:** dsahoo@ucsd.edu

Vinicius Campos, M.D.; Department of Gastrointestinal Surgery, Hospital de Clínicas de Porto Alegre, 2350 Ramiro Barcellos Street, Porto Alegre, RS, 90035-003. Email: viniciusjcampos@gmail.com

Rena Yadlapati, M.D.; Associate Professor, Division of Gastroenterology; Department of Medicine, University of California San Diego; 9500 Gilman Drive, MC 0703, La Jolla, CA 92093-0887. **Email:** ryadlapati@health.ucsd.edu

Kit Curtius, Ph.D.; Assistant Professor, Division of Biomedical Informatics, Department of Medicine, University of California San Diego; 9500 Gilman Drive, MC 0703, Biomedical Research Facility-2; La Jolla, CA 92093-0728. **Phone:** 619-777-3296: **Email:** kcurtius@health.ucsd.edu

Pradipta Ghosh, M.D.; Professor, Departments of Medicine and Cell and Molecular Medicine, University of California San Diego; 9500 Gilman Drive (MC 0651), George E. Palade Bldg, Rm 232; La Jolla, CA 92093. **Phone**: 858-822-7633: **Email:** prghosh@ucsd.edu

FOOTNOTES:

Abbreviations: BE, Barrett's metaplasia of the esophagus; EAC, Esophageal adenocarcinoma; GEJ, Gastroesophageal junction; Cau, Caucasian; AA, African American; SNP, single nucleotide polymorphism; *DARC*, Duffy antigen/ receptor for chemokines.

BRIEF SUMMARY

Background and context

Esophageal adenocarcinoma (EAC) is a highly lethal cancer among Caucasians, while African Americans are somewhat protected; what factors drive transformation with racial disparity remain unknown.

Findings

AI-enabled creation of the first computational map of neoplastic progression in the esophagus built and validated using transcriptomic datasets from diverse cohorts of human samples pinpointed CXCL8↔neutrophil tumor immune-microenvironment as a racially influenced driver of EACs and GEJ-ACs.

Limitations

Although cellular states during normal esophagus to EAC progression were identified, this study does not assess the mechanism(s) by which neutrophils accelerate tumor progression.

Impact

Computational tools pinpoint a racially influenced driver of cell transformation during BE→EAC progression; in doing so, it reveals new novel biology, informs disease modeling, therapeutic strategies, and biomarkers.

LAY SUMMARY:

By modeling diseases as networks, this work unravels a fundamental race-influenced immunologic driver of cell transformation in adenocarcinomas of the esophagus and the gastroesophageal junction.

KEY WORDS:

Barrett's metaplasia, network transcriptomics, Boolean implication relationships, SPT6, TP63, benign ethnic neutropenia (BEN), CXCL8/IL8, *Duffy* antigen (*DARC/ACKR1*).

SUMMARY (200-word limit)

Although Barrett's metaplasia of the esophagus (BE) is the only known precursor lesion to esophageal adenocarcinomas (EACs), drivers of the metaplasia→dysplasia→neoplasia cascade in the esophagus remains incompletely understood. Using an AI-guided network transcriptomics approach, in which EAC initiation and progression is modeled as networks to simplify complex multi-cellular processes, we first predict cellular continuum states and disease driving processes with an unprecedented degree of precision. Key AI-guided predictions are subsequently validated in a human organoid model and patient-derived biopsies of BE, a case-control study of genomics of BE progression, and in a cross-sectional study of 113 patients with BE and EACs. We find that all EACs must originate from BE, pinpoint a CXCL8/IL8↔neutrophil immune microenvironment as a driver of cellular transformation in both EACs and gastroesophageal junction-ACs, and that this driver is prominent in Caucasians (Cau), but notably absent in African Americans (AAs). Network-derived gene signatures, independent signatures of neutrophil processes, CXCL8/IL8, and an absolute neutrophil count (ANC) are associated with risk of progression. SNPs associated with ethnic changes in ANC modify that risk. Thus, findings define a racially influenced immunological basis for cell transformation and suggest that benign ethnic neutropenia in AAs may serve as a deterrent to BE→EAC progression.

INTRODUCTION

EACs are devastating cancers with high mortality (5% 5-y survival[1,2]). BE is the only known precursor lesion; it is not fully understood why only a small proportion of BE lesions progress to EACs[3]. Also unknown is why EACs display racial disparity[4-6]; African Americans (AAs) are ~4-5-fold less likely to get it than Caucasians (Cau). In fact, risk factors for EAC, e.g., long-segment BE and dysplastic-BE are also less frequent in AA than in Cau[7].

As to what propels the progression of BE to EACs, approaches to interrogate the genome (whole genome/exome sequencing, etc.) have revealed that BE displays early genomic instability, i.e., in copy number variation that is detectable later in EACs and have revealed patterns of progression (either 'born bad', gradual or catastrophic[3]) that may help to describe the evolution of EACs. Despite the insights gained, these genomic insights are yet to translate into prognostic biomarkers of the risk of BE→EAC progression, and answers to fundamental questions, i.e., what drives cellular transformation in BE lesions and if some of those drivers are racially influenced have remained elusive.

Modeling human diseases as networks simplify complex multi-cellular processes, helps identify patterns in noisy data that humans cannot find, and thereby improves precision in prediction. Although transcriptomic network studies in most cancers is a mainstream approach, the field of BE/EAC is notable for its absence. Here we present a network-based approach that uses artificial intelligence (AI) to identify continuum states (of tissues, cell types and processes, and signaling events and pathways) during the process of disease initiation and progression. Once built, the network model then guides subsequent validation (**Figure 1**). We demonstrate how this approach aid in the discovery of fundamental progressive timeseries events underlying complex human diseases and exploit such insights to deliver important insights. More specifically, we use this approach to not only identify key drivers of the metaplasia→dysplasia→neoplasia cascade in the esophagus, but also validate them in a recently described organoid model of BE[8], tissues and in a cross-sectional retrospective study on a cohort of patients with BE and EAC and analyze if/how race and/or ethnicity may impact these drivers (see study design in **Figure 1**).

RESULTS AND DISCUSSION

An AI-assisted study design

To identify cellular continuum states during pre-cancer to cancer progression, we chose a Boolean approach to building transcriptomic networks[9]. This approach has been used to create maps of evolving cellular states along any disease continuum and identify hitherto unforeseen cellular states in diverse tissues and contexts with high degrees of precision. For example, it helped pinpoint branchpoints in B/T cell differentiation[10,11], define progenitor cell hierarchy in blood[12-16], normal and neoplastic cell states in colorectal cancers[17,18], bladder cancers[19-22], and prostate cancers[23,24], and identify NK cell exhaustive states[25], universal cell proliferative[26] and macrophage[27] markers, and cell states in the mucosal barrier in IBD[28]. The superiority of the Boolean approach over other conventional approaches in accurately modeling gene expression changes and predicting clinical outcomes has been shown before[17,18,20,23] in head-to-head comparisons (see *Methods*). Because the Boolean approach relies on invariant relationships that are conserved despite heterogeneity in the samples used for the analysis, which often represent maximum possible diversity, i.e., the relationships can be thought of as general relationships among pairs of genes across all samples irrespective of their origin (normal or disease), laboratories/cohorts, different perturbations, and sometimes in multiple species including human, mouse and rat, and hence, considered conserved invariants. It is assumed that such 'invariants' are likely to be fundamentally important for any given process.

We used the Boolean approach to build maps of continuum states first during metaplastic progression in the normal esophagus (NE→BE), and subsequently during neoplastic transformation of the metaplastic epithelium (BE→EAC). Gene signatures were identified from each map, independently, using machine learning approaches. Signatures were validated in numerous independent cohorts (see **Figure 1**; top), prior to their use in validation studies; the latter included various experimental approaches on human tissues or tissue-derived organoids (**Figure 1**; middle). Map-derived gene signatures in conjunction with multiple independent gene signatures were used as precise and objective tools to navigate new biology, and to formulate and rigorously test new hypotheses, which led to a few notable findings (**Figure 1**; bottom).

A Boolean map of metaplastic progression in the esophagus

We used Boolean Network Explorer (*BoNE*)[29] to first create a model of progressive gene regulatory events that occur during metaplastic transition (**Figure 2A**). For model training and development, we used the largest (to our knowledge) well annotated transcriptomic dataset [n = 76: GSE100843[30]] derived from BE and proximal matched normal mucosa from squamous esophagus from 18 BE patients. Gene expression patterns were first simplified into 'clusters' within which each gene is equivalent to one another (**Figure 2B**). The clusters (nodes) were connected to one another (starting with the largest) based on the pattern of relationships between the clusters (edges), conforming to one of the six possible Boolean implication relationships (BIRs; **Figure 2C**). These efforts helped chart numerous Boolean paths (**Figure 2D**, *left*) within a network with directed edges (**Figure 2E**). Each

cluster was then evaluated for whether they belong to the healthy esophagus or diseased side (BE) depending on whether the average gene expression value of a cluster in heathy samples is up or down, respectively. Each path of connected gene clusters indicates a certain hierarchy in gene expression events, which translates to a progressive series of gene down/upregulation events, predicted to occur in sequence during the metaplastic process (**Figure 2D**, *right*).

We next introduced in *BoNE* machine learning that seeks to identify which of the gene clusters (nodes) connected by Boolean implication relationships (edges) are most optimal in distinguishing healthy from diseased samples. *BoNE* computes a score that naturally orders the samples; this score can be thought of as a continuum of states. Among all possible permutations and combinations, a set of two clusters emerged as most robust, which was further refined by an additional filtering step through a second 'training dataset' [GSE39491[31]; see **Supplemental Information 1**] which is comprised of BE and normal esophageal matched samples from 43 patients. Both training datasets were analyzed independently throughout the process. The resultant model of metaplastic transition pinpointed a time series of metaplasia (BE)-associated invariant events, in which downregulation of expression of 220 genes (*SPINK7*-cluster; **Figure 2F**) was invariably associated with a concomitant upregulation of 24 genes (*SLC44A4*-cluster, **Figure 2F**) in all samples in the training datasets. *SPINK7* (serine peptidase inhibitor, kazal type 7), is a key checkpoint in the esophageal keratinocyte stem cell, which regulates mucosal differentiation, barrier function, and inflammatory responses[32]. *SLC44A4* encodes a specific high-affinity regulated carrier-mediated uptake system for TPP in human colonocytes, involved in regulation of microbiota-generated thiamine[33]. The pattern of gene expression signature was sufficient to classify samples in seven independent validation cohorts and performed consistently well when doing so (ROC AUC 0.88-1.00; **Figure 2G**). For a complete list of genes in these clusters see **Supplemental Information 2**.

Reactome pathway analysis of the upregulated *SLC44A4*- and downregulated *SPINK7*-clusters along the path continuum revealed the most important biological processes that they control (**Figure 2H**). The downregulated pathways (**Figure 2H;** *bottom*) were cellular processes that are inherently associated with squamous epithelium, e.g., keratinization, cornified envelope formation, as expected. Other notable changes were TP53 expression and cell-cell adhesion proteins. These findings are consistent with emerging evidence from numerous independent studies which agree that aberrant TP53 IHC highly correlated with *TP53* mutation status (90.6% agreement) and was strongly associated with higher risk of neoplastic progression regardless of the presence/absence of dysplasia[34-36]. The findings are also in keeping with the reduction observed by IHC in cell adhesion proteins in BE lesions [E-cadherin, P-cadherin and the catenins which serve as adaptor proteins that enable the cadherins to achieve cell adhesion[37]]. The most notable cellular processes that were upregulated (**Figure 2H;** *top*) were related to oxygen delivery to the tissue, consistent with reports of Warburg and Crabtree effects in BE tissues[38].

BE map-derived signatures support an epigenetic cascade with loss of keratinocyte identity

We began by confirming that the network-derived signatures were recapitulated by our recently published[8] organoid model of BE, i.e., overlaps between upregulated DEGs and the *SLC44A4*-cluster and between downregulated DEGs and the *SPINK7*-cluster were significant (p-value 1.37 e-4 and 8.65 e-63, respectively; **Figure 2I**). This model is unique in that it does not represent established BE; instead, it recapitulates the process of metaplastic transformation from stratified squamous to intestinal epithelium. This model emerged serendipitously during studies interrogating fate determinants of human keratinocyte stem cells using an unbiased siRNA screen approach[39] (**Figure 3A**); loss of transcription elongation factor SPT6 emerged as a *bona fide* trigger for epithelial transcommitment from stratified squamous epithelium to a 'intestine-like' lineage (**Figure 3B**). Mechanistically, depletion of SPT6 (in conjunction with rescue studies) confirmed that such transdifferentiation was primarily due to stalled transcription and downregulated expression of TP63, the master regulator of keratinocyte fate and differentiation[40-43]. Computational approaches confirmed that this phenomenon of *transcommitment* faithfully recapitulates the metaplasia-specific signatures of BE[8], i.e., differentially expressed genes (DEGs) from nine independent BE datasets were captured in this model (**Figure 3C**). Transient acid exposure downregulated SPT6 and TP63, both transcript and protein[8]. The proposed working model was that acid exposure downregulates SPT6, which in turn inhibits TP63, which is sufficient for a cell fate switch from epidermis to a metaplastic columnar epithelium; the latter is indistinguishable from BE, both morphologically and by gene expression[8].

We hypothesized that SPT6 deletion may have altered the genome-wide chromatin accessibility to genes (beyond those genes that are targets of *TP63*) in the BE-map derived *SPINK7* and *SLC44A4* clusters. ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) studies on the SPT6-depleted BE organoid model confirmed that it is the DOWN-regulated genes in the *SPINK7*-cluster, which includes *TP63*, are significantly impacted when the histone chaperone SPT6 fails to interact with histone H3 and elongate Pol-II to regulate chromatin dynamics and mRNA biogenesis. Findings support our prior conclusions[39], in that the *SPT6→TP63* axis maintains keratinocyte identity (pathways enriched in the *SPINK7* cluster; **Figure 2H**), and that its loss permits transcommitment to a metaplastic intestine-like fate.

To analyze the translational relevance of these findings, we next asked if the *SPT6→TP63* axis is downregulated in the squamous esophageal lining in patients with BE. We prospectively enrolled patients with or without BE and collected biopsies from the distal esophagus, 2 cm above the GEJ or the BE segment. Immunohistochemistry studies on FFPE-embedded biopsies confirmed that, compared to non-BE subjects, both SPT6 and TP63 proteins were significantly suppressed in the esophageal squamous lining from patients who have progressed to BE (p values 0.8 e-9 and 0.9 e-7, respectively; **Figure 3D-E**). These findings support keratinocyte transdifferentiation as a trigger for BE, as proposed by multiple independent groups[44-47], but cannot rule out gastric cardia as alternative sites of initiation.

A Boolean map of EAC identifies an 'immune paradox' during cell transformation in EACs and GEJ-ACs

Next, we created a Boolean map of early cellular states that are encountered during BE→EAC transformation (**Figure 4A**). A sequence of gene expression changes emerged: a cluster of 471 genes (*LNX1*-cluster) were downregulated, with a concomitant and sequential upregulation of two clusters totaling another 61 genes (*IL10RA*- and *LILRB3*-clusters) invariably encountered in all samples (**Figure 4A**-*right*; **Figure 4B**). Machine learning approaches pinpointed the *LILRB3*- and *IL10RA*-clusters as sufficient to classify EACs from BE and do so reproducibly in 4 independent validation cohorts (**Figure 4C**). Because one of the strengths of the Boolean approach is its ability to extract using the logic of asymmetric invariant relations the fundamental timeseries of events in cross-sectional datasets[10], we saw that the UP-signatures are induced in a sequential manner (**Figure 4B**): While the *IL10RA*-cluster is up in non-dysplastic BE (**Figure 4D**-*left*), *LILRB3*-cluster is induced predominantly in EACs (**Figure 4D**-*middle*); the composite score of the combined signature shows progressive increase throughout BE→EAC transformation (**Figure 4D**-*right*). The genes in these clusters are listed in **Supplemental Information 3**

The degree of induction of the EAC-map derived signatures was indistinguishable in EACs and gastroesophageal junction adenocarcinomas (GEJ-ACs; GSE74553; **Figure 4E**-*left*), in that both tumor samples showed induction of the signature compared to normal esophagus (NE) and gastric (NG) samples. We validated this finding in another independent dataset (GSE96668; n = 60; **Figure 4E**-*right*), which was comprised of 11 normal GEJ tissues and 49 GEJ-ACs.

Reactome pathway analyses of these gene signatures revealed the set of cellular types and states that are progressively gained or lost (**Figure 4F-H**). Beyond the early Rac and Rho activation that were shown to be contributors during EAC initiation and progression[48] (**Figure 4G**; red font), the overwhelming and progressively increasing processes were that of innate reactive immune response and inflammatory cytokine signaling (**Figure 4G**), which was predominantly neutrophil centric and comprised of chemokines and receptors that specifically target the neutrophils (e.g., *CXCL8/IL8*, *CXCR1*, and *CXCL2*; see *LILRB3*-cluster in **Supplemental Information 3**). This is followed by an increase of an equally prominent alternative (immunotolerant/suppressive) immune response that is IL10, IL4/IL13-centric (**Figure 4H**). Such immunosuppressive states have indeed been reported in the setting of EAC by others[49], using a plethora of approaches, ranging from increased IL10 levels in the serum[50], to confirmed expression of IL10 protein/mRNA in the tissue[51,52]; all studies agree that higher immunosuppression carries worse prognosis. The upregulation of both reactive and tolerant immune responses, hence, paradoxical immune response, was associated with a concomitant loss of two known tumor suppressive processes in the esophagus: IL18 signaling[53-55], which plays a suppressive role uniquely in GI tract cancers[56], and the TP53 pathway (**Figure 4F**). It is noteworthy that loss of TP53 itself is known to orchestrate a distinct immune cell infiltration[57] and response.

These findings show that the dynamic immunologic landscapes predicted by the map conform to prior knowledge. They also increase confidence in the map's ability to formulate new hypotheses and make predictions, which we set out to do next.

A map of cellular continuum states confirms that all EACs must evolve through BE

We used the concept of Boolean invariant logic to create a model that captures both the early (metaplastic) and late (transformation) steps of cellular state change. First, we found that BE signatures (NE>>BE map; **Figure 2D**) are also induced in the EAC samples across diverse cohorts (**Figure 5A**). In fact, we found that the Boolean Implication *SPINK7* high => *SLC44A7* low, which defines metaplastic transition in the normal epithelium (**Figure 2E**), is an invariant relationship in the most diverse global human GSE119087 (n = 25955) dataset (**Figure 5B**), suggesting that this pattern may be fundamentally important. Using *SLC44A4* as a seed gene in a dataset that was comprised of normal esophagus, BE and EAC samples, we found that it shares an invariant relationship with one of the genes in the *LILRB3* cluster, *CXCL8*. *CXCL8* high => *SLC44A4* high in an invariant Boolean implication relationship in normal esophagus (NE), BE and EAC samples where each sample type is mostly confined to one quadrant (NE, bottom-left; BE, top-left; and EAC, top-right). This model suggests that if EACs must originate from the esophagus, they must do so via the metaplastic BE intermediate; only 3 genes could nearly accurately classify the samples (**Figure 5C**) and show progressive expression changes along the continuum (**Figure 5D**). These findings were validated in a second cohort pooled from multiple independent datasets (**Figure S1A**). By contrast, esophageal squamous cell carcinomas (ESCCs; > 400 samples pooled), did not conform to the Boolean logic-based model of the BE→EC continuum, and as expected, the model confirms that ESCCs do not transition through metaplastic BE states (**Figure S1B**). Datasets that include NE, gastric cardia, BE and EACs/GEJ-ACs could not be pooled (due to incompatible array platforms); this hindered our ability to objectively test if BE→EACs originate in NE or from the gastric cardia; nor could we test if GEJ-ACs originate in BE.

We analyzed the overlap between our network-derived signatures for BE (**Figure 2E-F; Supplemental Information 2**) or EAC (**Figure 4A; Supplemental Information 3**) and the 274 genes that were reported as aberrantly methylated in 6 independent EAC and/or BE methylation studies (i.e., an integrated methylome of BE; summarized in Krause et. al[58]). These genes can separate NE from EAC/BE, but not EAC from BE, indicating that the same clusters of methylation profiles exist in BE and EAC[58]. Significant overlaps were seen between the 274 genes and the BE-associated *SLC44A4*-cluster (p value 1.59e-10), but not the EAC-associated *LILRB3*- and *IL10RA*-clusters (**Figure 5E**), suggesting that the methylome in EAC is imprinted early during evolution through BE.

Two 'waves' of an IL8↔neutrophil-centric inflammation is encountered during cell transformation

We noted that among the genes that were upregulated during cell transformation (i.e., the BE→EAC map; **Figure 4A**) the *LILRB3*-cluster was enriched for one major cytokine/chemokine pathway and its target immune cell, neutrophils. It contained the *CXCL8* (henceforth, IL8) gene and the receptor for IL8, *CXCR1* (see complete gene list in **Supplemental Information 3**); IL8 was sufficient to classify BE and EAC (**Figure 5C-D**). Discovered as the first chemokine activator of neutrophils, IL8 displays a distinct target specificity for the neutrophil, with only

weak effects on other blood cells[59-61]. The *LILRB3*-cluster also contained *CXCL2*; the CXCL2-CXCR2 axis helps in the recruitment of tumor-associated neutrophils[62].

We asked, how this IL8↔neutrophil centric inflammation varies during the metaplasia-dysplasia-neoplasia cascade by comparing pairwise each sequential step, i.e., normal esophagus (NE) vs non-dysplastic BE (NDBE), NDBE vs dysplastic BE (BE-D), BE-D vs EAC. The induction of BE and EAC map-derived signatures, as well as IL8 and its two signaling receptors (CXCR1/2) (**Figure 5F**) and numerous pathologic neutrophil processes (**Figure 5G; Figure S2**) was significant in EACs (**Figure 5F-G**, *row iv*) and in GEJ-ACs (**Figure 5F-G**, *row vi*). The patterns of induction of all the gene signatures (see **Supplemental Information 4** for gene lists) were virtually indistinguishable between EACs and GEJ-ACs (**Figure 5F-G**, *row v* and **Figure S2C-F-I-L**). Upregulation was observed in two phases: Early during metaplastic transformation from NE to NDBE (*row i*) and later during transformation from BE-D to EACs (*row iii*), but not during BE to BE-D (*row ii*) (see also **Figure S2** for violin plots). These results show that the normal→metaplasia→dysplasia→neoplasia cascade is associated with a staircase waveform of IL8 and neutrophil processes. Because the IL8↔neutrophil axis is known to alter tumor immune landscape[63] and increased genomic instability[64], we conclude that a biphasic IL8↔neutrophil interplay is a specific association, and presumably, a potential driver of cell transformation along the BE↔EAC continuum.

Gene signatures reveal pro-tumor neutrophil infiltrates, benefit from immune checkpoint therapies

We next asked if the increased neutrophil processes were reflective of N1 or N2 profile of neutrophils, which represent immunostimulating (anti-tumor) and immunosuppressive (pro-tumor) subsets, respectively, in cancer[65]. To this end, we analyzed a tumor-associated neutrophil (TAN[66]) signature that measures the protumor N2 TANs. Both EACs (**Figure 5G**-*last column*, *row iv*) and GEJ-ACs (**Figure 5G**-*last column*, *row vi*) were associated with an induction of TAN signature. Similarly, upregulation in TAN signature was noted in pairwise comparisons of NE vs NDBE (**Figure 5G**-*last column*, *row i*) and NDBE vs DBE (**Figure 5G**-*last column*, *row ii*). ROC AUC and p values are displayed for each pairwise comparison in **Figure 5G**.

The induction of pro-tumorigenic TAN signatures was associated also with the pan-cancer marker of adaptive immune resistance, an 18-gene "tumor inflammation signature" (TIS[67,68]) and its 5-gene subset (that remained significantly higher in responders after correction for multiple testing[67]) (**Figure 5F***; last two columns*). Higher TIS was demonstrated to retrospectively predict benefit of anti-PD-1 therapy in various cancers[67,68]. These findings are consistent with reports of higher TIS score in EACs[69], regardless of treatment status and that Pembrolizumab was found to be effective in esophageal and GEJ cancers (KEYNOTE-590 trial[70,71] and reviewed in[72]). EAC signatures, neutrophil processes and TIS signatures positively and strongly correlated across all EAC and GEJ-AC datasets analyzed (*r* ranging from 0.8-0.99 or TIS vs EAC signatures; **Figure S3**). These findings suggest that pro-tumorigenic neutrophils may drive adaptive immune resistance.

Currently available animal models of BE→EAC transformation rarely recapitulate human disease

Animal models of diseases have both merits and limitations[73,74]. Because 'mice are not men'[75,76], especially when it comes to their innate immune system[76,77], and EACs and GEJ-ACs are associated with a prominent immune signature, we asked how well currently available EAC models recapitulate the human disease. To model how EAC-associated risk factors, i.e., obesity/BMI and IL8-induction enhance cell transformation, mice challenged with high-fat diet (HFD[78-80]) or overexpressing IL8[80] have been developed. Neither model induce our network-derived BE/EAC signatures (**Figure 5H**, *row i-iii*; **Figure S4A-C**); nor did they display induction of the neutrophil signatures we observed in human tissues (compare human-**Figure 5G** with murine- **Figure 5I**, *row i-iii*). The signatures were, however, induced in a transgenic interleukin1-β (IL1β)-overexpression model (GSE24931; **Figure 5H-I**; *rows iv-vi;* **Figure S4D-F**); in this model, the human IL1β cDNA was inserted downstream of an Epstein-Barr virus (ED-L2) promoter that targets the oral cavity, esophagus, and squamous forestomach[81]. These mice develop chronic inflammation that progresses to BE and EAC; progression was accelerated by exposure to bile acids. Findings show that a combination of inflammation and bile acids, the latter are components of gastroduodenal reflux that has been linked to BE→EAC progression[81,82]. Most importantly, the bile acid-accelerated model (GSE24931) recapitulated the neutrophil processes that were encountered in most human datasets of EACs and GEJ-ACs (**Figure 5G**).

Absolute neutrophil count (ANC) and neutrophil signatures prognosticate outcome in BE/EAC

We next asked how neutrophil counts and signatures impact the risk of transformation in BE and EAC outcome. First, we retrospectively analyzed data of patients with BE [NDBE, n = 72, DBE, n = 11] diagnosed between 2013 and 2017 and patients with EACs [n=30] diagnosed between 2005 and 2017, with a complete blood count within 6 months from diagnostic endoscopy (see *Methods*; **Figure 6A**). Patients were followed up for outcome until December 2021. As published before by numerous groups[83-85], the neutrophil:lymphocyte ratio (NLR) progressively increased during NDBE→DBE→EAC progression (**Figure 6B**). However, the increase in NLR in DBE and the extent of increase in NLR in EAC were largely driven by progressive increases in the absolute neutrophil counts (ANC; **Figure 6C**) and not due to reduced absolute lymphocyte counts (ALC; **Figure 6D**). Platelet count (**Figure 6E**) and total leukocyte count (**Figure 6F**) were significantly increased in patients diagnosed with EACs; the latter is largely driven by neutrophilia and despite significant lymphopenia. ANC remained the most significant variable that tracked the risk of NDBE→DBE→EACs progression in both univariate (**Figure 6G**) and multivariate (**Figure 6H**) analyses.

The prognostic role of the EAC signature (**Figure S5A**), the neutrophil degranulation (signature derived from the EAC cluster; **Figure S5B**), the neutrophil chemoattractant, CXCL8/IL8 (**Figure S5C**), and neutrophil abundance, estimated in tumor tissues by transcripts of the marker, CD16[86] (*FCGR3A/B*, Fc Gamma Receptor IIIa; **Figure S5D**) were also confirmed in the EAC datasets, curated from the TCGA. While the EAC and neutrophil signatures retained their prognostic impact also in ESCCs, CXCL8 did not (**Figure S5A-D**; right column). None of these signatures prognosticated outcome in gastric adenocarcinomas (GC; **Figure S5A-D**;

right column). A combined CD16 and CXCL8/IL8 composite score was prognostic for EACs (**Figure 6I**), but not ESCC or GCs (**Figure 6J-K**). These results show that high ANC, high CXCL8/IL8 and high intra-tumoral neutrophil-activation signatures may all contribute as disease drivers. They also suggest that IL8-driven trafficking of neutrophils into the tissue may be a unique driver in EAC, but not in ESCC or GCs.

Caucasians, but not African Americans mount IL8- and neutrophil-centric inflammation

The discovery of the only known race-influenced mechanism that protects cells against DNA damage, i.e., high glutathione S-transferase theta 2 (GSTT2) in AAs compared to Cau[87] (**Figure S7A**), was made using a unique dataset containing histologically normal esophageal squamous lining-derived transcriptomics from Cau and AA subjects with normal (N) or BE diagnosis (BE) (**Figure 7A**; see **Supplemental Information 1**; GSE77563[87]). We leveraged this dataset to ask how our findings differ along the race/gender divide. We made 3 observations: **a)** the BE (**Figure 7B**-*left*) or EAC (**Figure 7B**-*right*) signatures were not different in the squamous lining of the esophagus as baseline (compare AA-N vs Cau-N); **b)** a diagnosis of BE in Cau, but not AA was associated with an induction of the EAC signatures in the histologically normal proximal squamous lining (**Figure 7B**-*right;* compare AA-BE vs Cau-BE); **c)** When these signatures, and all other signatures of tumor microenvironment and neutrophil processes were analyzed systematically, we found that the changes in gene signatures was seen in both genders (**Figure 7C-D**, compare bottom 3 rows; **Figure S6**); however, Caucasian men [Cau-BE (M)] accounted for the most significant changes in the signatures across the board. Thus, the nature of the progressive inflammation and cell states identified by our network approach is induced preferentially in Cau subjects with BE, in whom the risk of BE/EAC is ~4-5-fold higher than AAs. Finally, we found that the TIS[67,68] is induced in Cau subjects with BE, but not AA (**Figure 7C**), suggesting that these inhibitors may stall cell transformation in Cau subjects.

It is noteworthy that the protective high levels of expression of GSTT2 in AAs over Cau[87] (see legend, **Figure S7A**) is observed both in healthy controls as well as in subjects with BE/EAC (**Figure S7B**). GSTT2 expression inversely correlated with the EAC signature (**Figure S7C**) and its subset of neutrophil degranulation signature (**Figure S7D**). These findings indicate that the neutrophil/IL8 centric inflammation that is mounted exclusively during BE→EAC progression in the Cau subjects (**Figure 7B**-*right*) occurs in the vulnerable setting of low GSTT2 in the same subjects. Because GSTT2 protects against DNA damage, findings suggest that racial differences in GSTT2 may protect AAs from or predispose Cau to more DNA damage in the setting of the IL8-neutrophil centric inflammation.

SNPs that increase or decrease ANC are oppositely enriched during BE→EAC progression

We next asked if race and/or ethnicity may intersect directly with ANC and the risk of BE→EAC progression. Populations of African descent are known to have low ANCs compared to others[88] while those of Hispanic/Latino (HL) descent have high ANCs compared to non-Hispanic white (NHW[89]). Single nucleotide polymorphisms

(SNPs) that either increase (such as in the US Hispanic/Latino population[90]) or decrease (such as in the US AA population[91,92]) ANCs have been identified. In the case of AAs, a study[92] that pooled >6,000 AA samples from 3 studies [the Jackson Heart Study (JHS), the Atherosclerotic Risk in Communities (ARIC) Study, and the Health, Aging and Body Composition (Health ABC) Study], the homozygous SNP (rs2814778, which disrupts a binding site for the GATA1 erythroid transcription factor, resulting in a ACKR1-null phenotype (**Figure 8A**) was found to be predictive of low ANC in AAs above beyond the previously described admixture association. This study established a *bona fide* phenotype (low ANC) for this genetic variant and demonstrated that the causal variant must be at least 91% different in frequency between AA vs Cau. In the absence of a longitudinal study that monitors BE progression in AA vs Cau subjects, we leveraged a case-control study[93] that included 80 patients with BE (40 who progressed to EACs and 40 who did not, i.e., non-progressors); this cohort was selected from a larger case-cohort study within the Seattle Barrett's Esophagus Program (SBEP) at the Fred Hutchinson Cancer Research Center[94]. Germline data from this cohort was analyzed for the occurrence of 10 SNPs that increase or decrease the ANC, as determined in various studies in the US (see complete list in **Supplemental Information 6**). All 3 genes (*DARC/ACKR1*, *ABCC1* and *HMMR*), impacted by 4 of the 6 tested risk alleles, including *rs2814778* [the allele maximal risk, across studies, and the one that confers the risk of benign ethnic neutropenia (BEN[95])] were significantly enriched among non-progressors, whereas 3 of the 4 protective alleles were significantly enriched among the progressors (**Figure 8B**). Among other SNPs that are associated with drug induced neutropenia in the non-US population, only one was significant (i.e., *CYP39A1*; **Supplemental Information 6**). The opposing patterns of enrichment and de-enrichment of neutropenia-protective and risk alleles, respectively, among BE→EAC progressors was significant (Welch's t-test p = 0.03978; **Figure 8B**). As expected, somatic mutations in genes within the EAC clusters (**Figure S7A**) or on genes associated with neutrophil function/number (**Figure S7B**) were more frequent among progressors compared to non-progressors, and the frequency increased further in tumors with high mutation burden.

Ethnic neutropenia may reduce EAC risk

We hypothesized that if the IL8↔neutrophil axis drives cell transformation and hence, the risk of EACs, then ethnic prevalence of polymorphisms that impact either IL8 signaling and/or neutrophil abundance might influence such risk. We found that ethnicities with the lowest incidence of BE/EAC (i.e., AAs), also have BEN[95], the most common form of neutropenia worldwide. In BEN, a homozygous SNP (*rs2814778*) impacts the functions of *DARC (only in RBCs),* which encodes is a 7-TM receptor[96] that selectively scavenges "inflammatory" chemokines e.g., IL8 and *CCL5*, both of which enhance neutrophil recruitment[97-99] (**Figure 8A**). The AA *Duffy* polymorphism has explained many mysterious racial disparities in modern medicine, e.g., high susceptibility to HIV[91], but protection against *P. vivax* malaria[92], high sickle cell-related complications, aging phenotypes[100], and acute lung injury[101] among AAs. Its impact on neutrophil count is generalizable also to Yemenite Jews[102] and Hispanics/Latinos[90]. We looked for the '*Duffy* negativity phenomenon' which was first described in malaria[103]; this phenomenon refers to the geographic distribution of the *Duffy* negative genotype, $Fy^{a-/b-}$, predominantly in sub-Saharan Africa (≥95% Duffy negativity frequency; CI 75%-95%; **Figure 8C**) which coincides with the

phenotype of near complete protection from *P. vivax* in the same regions (**Figure 8D**-*bottom panel*). It is noteworthy that *Duffy* negativity does not offer protection from *P. falciparum* (**Figure 8D**-*top panel*). A strikingly similar contrasting pattern was seen when we compared the global age-adjusted incidence rates of ESCCs and EACs (**Figure 8E**); the African and Saudi Arabian regions the *Duffy* negative genotype, $Fy^{a-/b}$ (**Figure 8C**) showed low incidence of EACs, but moderate to high incidence of ESCCs. These findings suggest that BEN could offer selective protection from EACs (just as it does for *P. vivax*) in individuals of African descent. Findings also suggest that BEN, which is widely prevalent in other ethnicities (e.g., Africans[104-106], AAs, Arabs[107], Yemenite and black Ethiopian Jews[105], and to a lesser extent also in Latinos[90]) but < 1% in the non-Hispanic white population in the US[89] is a possible risk modifier (protective) for BE→EAC progression.

CONCLUSIONS

The major discovery we report here, using an off-the-beaten path network transcriptomic approach, are insights into the cellular continuum states during the metaplasia→dysplasia→neoplasia cascade in adenocarcinomas of the esophagus and gastroesophageal junction. Findings enable us to draw four major conclusions, some with immediate and impactful translational relevance.

All EACs must originate from BE: First, our findings support a long-suspected tenet, originally proposed in 2000[108],, i.e., that all EACs arise in BE. That tenet was substantiated *via* recent multiscale computational modeling studies[109], through single-cell genomics and lineage tracking studies[110], and now through a Boolean logic-based model (current work). These three approaches independently verify that BE is the invariant precursor to EACs. The same Boolean logic-based model confirmed that, unlike EACs, ESCCs do not transition through metaplastic BE states (as expected). TP63 and SPT6 were suppressed in the squamous lining of the esophagus proximal to the BE segment (in **Figure 3D-E**), and BE/EAC signatures and neutrophil inflammatory milieu was similarly observed in the normal squamous lining of the esophagus proximal to BE/EAC lesions (in **Figure 7A-B**) suggests that the histologically so-called 'normal' esophageal lining is abnormal by all molecular (protein and gene expression) metrics among patients with BE/EAC. This evidence lends support to a transcommitted esophageal keratinocyte being a cell of origin of BE and EAC/GEJ-ACs, as has been suggested by several independent groups[44-47].

Neutrophilic assault assists cell transformation: Second, we pinpointed a 2-step waveform of an IL8-neutrophil centric tumor immune microenvironment that is increased first during metaplastic transition and more prominently later during neoplastic transformation (see **Figure 8F** legend). This biphasic-wave pattern may explain why long-standing BE carries low risk of EAC[111], presumably because a 2nd wave/'hit' of immune storm is required for cell transformation. Induction of IL8 and its receptors, CXCR1/2[49,112-116], increased presence of neutrophils in BE tissues[117] and circulation[83,84] have been reported and increased NETosis within EACs in the setting of peripheral leukocytosis has been shown to carry worse prognosis[118]. In showing that the IL8-neutrophil centric tumor microenvironment is prominently induced in Cau, but not in AAs, which mirrors the ~4-5 fold higher risk of EACs in Cau compared to AAs[119], our findings suggest that this inflammatory microenvironment is likely to be a driver event. Besides the tumor environment, we showed through multivariate analyses that a high ANC in circulation is an independent determinant of NDBE→DBE→EAC progression. We provide epidemiologic evidence and propose that ethnic neutropenia is a possible risk modifier (protective) for BE→EAC progression in AAs, Arabs, and other ethnicities. The fact that peripheral neutrophilia and intratumoral signatures of neutrophil processes are aligned, i.e., both carry poor prognosis in EACs is not unusual because prior studies have confirmed that both peripheral neutrophilia and the infiltration of TANs are associated with worse outcome in diverse cancers[120]. Although not investigated here, the insights into the IL8↔neutrophil centric immune microenvironment aid in formulating testable hypotheses on disease pathogenesis and prioritize a class of neutrophil-targeted therapeutics. For example, TAN infiltration may accelerate adenocarcinomas by impacting a slew of pro-cancer processes (reviewed in[121-123]), and hence, anti-neutrophil drugs could prove beneficial as both

single agent as well as adjuvant therapy. Because neutrophil infiltration in tumor is associated with key features of resistance to immune checkpoint inhibitors[124], and high neutrophil counts with high tumor mutation burden in diverse cancers (including esophageal) is known to reduce the efficacy of the checkpoint inhibitors[125], our results predict that neutrophil-targeted therapeutics will synergize with checkpoint inhibitors.

Ethnic disparity in EACs/GEJ-ACs* may stem from *DARC polymorphism: Third, our studies shed valuable insights into how *DARC* polymorphism *rs2814778* may shape the risk of EACs. For example, neutrophil processes are high in both EACs and ESCCs, and yet, the existence of a *Duffy* negativity phenomenon in EACs (that the Fy^{a-b-} genotype is associated with low incidence of EACs, but not ESCCs) suggests that the *Duffy* polymorphism impacts EACs through mechanisms other than being the most important genetic determinant of BEN. *DARC* polymorphism *rs2814778* is known to impact serum levels of MPO[126,127] and IL8[128] and their co-existence with IL8 polymorphisms is known to impact the prevalence of periodontitis in certain ethnicities[129]. Because EACs, but not ESCCs, significantly induce *CXCL8*, it is possible that the infiltration of TANs in EACs is gradient driven, and that gradient is maintained by the RBC-localized scavenger of *CXCL8*, *DARC* (serving as a cytokine 'sink'[130-132]). We cannot rule out that *DARC*-mediated chemokine sequestration may modulate tumor microenvironment, angiogenesis and metastasis, which has been described in other cancers (reviewed in[91]). As for the observed gender disparity (higher risk of EACs/GEJ-ACs in men), a more reactive neutrophil state[133] or prevalence of autoimmune or idiopathic neutropenia[89,134] may offer some explanation, but ethnic neutropenia due to *DARC* polymorphism *rs2814778* is just as important of a determinant of low ANC in females as in men[135].

EACs and GEJ-ACs behave similarly: Fourth, by showing that EACs and GEJ-ACs share similar gene signatures and tumor microenvironments, our study objectively establishes the degree of similarity between the two tumors, both with rapidly rising incidence, at a fundamental molecular level. This finding is in keeping with the fact that, much like EACs, GEJACs are also associated with short and long segments of BE suggesting that they arise from underlying metaplastic epithelium[136]. It is also consistent with the fact that the entity of familial BE has been seen in ~7.3% of patients presenting with BE, EACs, or GEJACs[137]. It is possible that much like EACs, GEJ-ACs evolve through the intestinal metaplastic continuum and that neutrophil-targeted or other immune checkpoint therapies that emerge in EACs are expected to have crossover benefits in GEJ-ACs.

STUDY LIMITATIONS

The biggest limitation of this study and our approach is the availability of data (or lack thereof). For example, due to lack of datasets with normal esophagus, gastric cardia, BE and EAC/GEJ-ACs all in one cohort, the logical model of EAC and GEJ-AC initiation we cannot rule out that cardia→BE→EAC is another plausible path. ANC, serum, and tumor IL8 expression, TAN and tumor neutrophil signatures were not evaluated in the same cohort; well-curated prospective studies will be necessary to connect these dots – independently and as composite measures -- with outcome. Such studies will help understand the role of IL8-gradient (serum to tumor),

neutrophilia and TAN infiltration. Although this study proposes BEN as a modifier of EAC risk, the impact of BEN on BE→EAC progression requires a study that is controlled for all variables (including race, smoking, BMI, etc) and genotyped (for *DARC* polymorphism rs2814778) and phenotyped (monitored for ANC and BE/EAC progression). Because smokers in all ethnic groups had a higher neutrophil count than non-smokers[89] (greatest effect on increments of ANC among Cau), it is possible that neutrophil processes and neutrophilia intersects with the established smoking-related risk of EACs and GEJ-ACs[138]; this was not investigated here. This study also did not address the mechanism(s) by which peripheral neutrophils/TANs may accelerate adenocarcinomas in the esophagus and the GE-junction and how *DARC* polymorphism *rs2814778* may offer protection. Tumor-neutrophil co-culture studies incorporating IL8 gradients and titrated neutrophil counts may yield valuable insights into how neutrophil infiltration impact tumor mutation burden.

METHODS (1000 words)

Detailed methods for computational modeling, AI-guided prediction and validation and description of validation models are presented in Supplementary Online Materials and mentioned in brief here.

Computational Approach

An overview of the key approaches is shown in **Figure 1**. Modeling continuum states within the metaplasia→dysplasia→neoplasia cascade was performed using Boolean Network Explorer (*BoNE)*[29]. We created an asymmetric gene expression network, first for metaplastic progression from normal esophagus to BE, and separately, for the dysplastic→neoplastic cascade during BE to EAC progression, using a computational method based on Boolean logic [9,10,139]. To build the BE/EAC network, we analyzed two publicly available colon-derived transcriptomic datasets [GSE100843[30] and GSE39491[31] for BE and E-MTAB-4045 for EACs (see **Supplemental Information 1**). These two datasets ('test cohorts') were independently analyzed and the resultant signatures were kept separate from each other at all times. A Boolean Network Explorer (*BoNE*; see Supplementary Methods) computational tool was introduced, which uses asymmetric properties of Boolean implication relationships (BIRs as in MIDReG algorithm[9]) to model natural progressive time-series changes in major cellular compartments that initiate, propagate and perpetuate cellular state change and are likely to be important for BE/EAC progression. *BoNE* provides an integrated platform for the construction, visualization and querying of a network of progressive changes much like a disease map (in this case, BE and EAC-maps) in three steps: First, the expression levels of all genes in these datasets were converted to binary values (high or low) using the *StepMiner* algorithm[140]. Second, gene expression relationships between pairs of genes were classified into one-of-six possible BIRs and expressed as Boolean implication statements; two symmetric Boolean implications "equivalent" and "opposite" are discovered when two diagonally opposite sparse quadrants are identified and four asymmetric relationships, each corresponding to one sparse quadrant. While conventional symmetric analysis of transcriptomic datasets can recognize the latter 2 relationships, such approach ignores the former. BooleanNet statistics is used to assess the significance of the Boolean implication relationships[9]. Prior work[28] has revealed how the Boolean approach offers a distinct advantage from currently used conventional computational methods that rely exclusively on symmetric linear relationships from gene expression data, e.g., Differential, correlation-network, coexpression-network, mutual information-network, and the Bayesian approach. The other advantage of using BIRs is that they are robust to the noise of sample heterogeneity (i.e., healthy, diseased, genotypic, phenotypic, ethnic, interventions, disease severity) and every sample follows the same mathematical equation, and hence is likely to be reproducible in independent validation datasets. The heterogeneity of samples in each of the datasets used in this study is highlighted in **Supplemental Information 1**. Third, genes with similar expression architectures, determined by sharing at least half of the equivalences among gene pairs, were grouped into clusters and organized into a network by determining the overwhelming Boolean relationships observed between any two clusters [9,10]. In the resultant Boolean implication network, clusters of genes are the nodes, and the BIR between the clusters are the directed edges; *BoNE* enables their

discovery in an unsupervised way while remaining agnostic to the sample type. All gene expression datasets were visualized using Hierarchical Exploration of Gene Expression Microarrays Online (HEGEMON) framework[141].

Human subjects:

For assessing the impact of absolute neutrophil count (ANC) on BE/EAC diagnosis and outcome of EACs, we retrospectively analyzed patients with biopsies reporting BE between 2013 and 2017 and with a complete blood count within 6 months from the endoscopy, as well as patients with esophageal adenocarcinoma (EAC) at a tertiary care center in Brazil (Hospital de Clínicas dePorto Alegre). Cases (n = 113) were classified as non-dysplastic BE (NDBE, n = 72), dysplastic BE (DBE, n = 11) and EAC (n = 30)[83]. Data collected from endoscopic histopathological reports were histologic diagnosis and adherence to Seattle Protocol[142,143]. The study was approved by the Brazilian National Committee on Research Ethics (CONEP), registered by number CAAE-81068617.2.0000.5327.

For collecting esophageal biopsies for IHC we enrolled patients undergoing endoscopies as a part of their routine care and follow-up at UC San Diego's Center for Esophageal Diseases. Patients were recruited and consented using a study proposal (Project ID#200047) approved by the Human Research Protection Program (HRPP) Institutional Review Board of the University of California, San Diego (Project ID#200047).

Human keratinocyte-derived BE organoids were created[39] and validated[144] previously using an approved IRB (Project ID # 190105) that covers human subject research at the UC San Diego HUMANOID Center of Research Excellence (CoRE). For all the deidentified human subjects, information including age, gender, and previous history of the disease, was collected from the chart following the rules of HIPAA. The study design and the use of human study participants was conducted in accordance with the criteria set by the Declaration of Helsinki.

ACKNOWLEDGEMENTS

This work was supported by NIH AI141630 and CA100768 (to P.G), GM138385 (to D.S), T32 GM8806 (to D.V) and AI155696, UG3TR003355 and UG3TR002968 (to D.S, P.G and S.D), P30 CA023100 (KC and CG), and K23 DK125266 (RY). P.G, D.S, R.Y and S.D were also supported by funds from the Torey Pines. KC and CG were also supported by a UCSD Academic Senate Grant (RG103468). We thank Donald Pizzo (UCSD Pathology Histologic Biomarkers Core), Madeline Gerytak and Vanessa Castillo (UCSD) for technical and logistical support.

AUTHOR CONTRIBUTIONS TO MANUSCRIPT

DS and PG conceptualized the project; DS and DV, under the supervision of PG and DS carried out computational analyses; DS contributed all software used in this work; RY provided access to patients and secured all tissue samples used in this study; VGH carried out the immunohistochemical studies under the supervision of PG; CT and SD were responsible for the BE organoid model studies; CG, PH and KC carried out the mutation and SNP analyses in BE datasets, accessed through the pre-cancer genome atlas (PCGA) pipeline co-created by LA and SML. DS and PG prepared figures for data visualization and wrote the original draft of the manuscript. All authors provided input and edited and revised the manuscript. All co-authors approved the final version of the manuscript. PG coordinated and supervised all parts of the project and administered the project.

DATA AVAILABILITY

All data is available in the main text or the supplementary materials. Publicly available datasets used are enlisted in **Supplemental Information 1**. The software codes are publicly available at the following links: https://github.com/sahoo00/BoNE and https://github.com/sahoo00/Hegemon.

STATEMENT ON CONFLICT OF INTERESTS

RY is a consultant for Medtronic (Institutional), Ironwood Pharmaceuticals (Institutional) and Phathom Pharmaceuticals, and has research support from Ironwood Pharmaceuticals. She is also on the advisory Board with Stock Options at RJS Mediagnostix. These entities had no influence on study design or conclusions.

SUPPLEMENTAL INFORMATION (Excel datasets)

Supplemental Information 1: Excel sheet with list of all transcriptomic datasets used in this work.

Supplemental Information 2: Excel sheet with list of BE-map derived genes and reactome pathways.

Supplemental Information 3: Excel sheet with list of EAC-map derived genes and reactome pathways.

Supplemental Information 4: Excel sheet with list of all neutrophil process-related signatures.

Supplemental Information 5: Excel sheet with hematologic parameters assessed in a cross-sectional retrospective study on 113 BE and EAC patients.

Supplemental Information 6: Excel sheet with SNP frequency among progressors and non-progressors in a case-control genomics study [the Seattle Barrett's Esophagus Program (SBEP)].

REFERENCES

1. Rustgi, A.K. & El-Serag, H.B. Esophageal carcinoma. *N Engl J Med* **371**, 2499-2509 (2014).

2. Xie, S.H. & Lagergren, J. Risk factors for oesophageal cancer. *Best Pract Res Clin Gastroenterol* **36-37**, 3-8 (2018).

3. Killcoyne, S. & Fitzgerald, R.C. Evolution and progression of Barrett's oesophagus to oesophageal cancer. *Nat Rev Cancer* (2021).

4. Islami, F., DeSantis, C.E. & Jemal, A. Incidence Trends of Esophageal and Gastric Cancer Subtypes by Race, Ethnicity, and Age in the United States, 1997-2014. *Clin Gastroenterol Hepatol* **17**, 429-439 (2019).

5. Xie, S.H., Rabbani, S., Petrick, J.L., Cook, M.B. & Lagergren, J. Racial and Ethnic Disparities in the Incidence of Esophageal Cancer in the United States, 1992-2013. *Am J Epidemiol* **186**, 1341-1351 (2017).

6. Cook, M.B., Chow, W.H. & Devesa, S.S. Oesophageal cancer incidence in the United States by race, sex, and histologic type, 1977-2005. *Br J Cancer* **101**, 855-859 (2009).

7. Khoury, J.E.*, et al.* African Americans with Barrett's esophagus are less likely to have dysplasia at biopsy. *Dig Dis Sci* **57**, 419-423 (2012).

8. Vo, D.T.F., MacKenzie R.; Tindle, Courtney; Anandachar, Mahitha Shree; Das, Soumita; Sahoo, Debashis; Ghosh, Pradipta. SPT6 loss permits the transdifferentiation of keratinocytes into an intestinal fate that resembles Barrett's metaplasia. *iScience* **24**(2021).

9. Sahoo, D., Dill, D.L., Gentles, A.J., Tibshirani, R. & Plevritis, S.K. Boolean implication networks derived from large scale, whole genome microarray datasets. *Genome Biol* **9**, R157 (2008).

10. Sahoo, D.*, et al.* MiDReG: a method of mining developmentally regulated genes using Boolean implications. *Proc Natl Acad Sci U S A* **107**, 5732-5737 (2010).

11. Inlay, M.A.*, et al.* Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes Dev* **23**, 2376-2381 (2009).

12. Chen, J.Y.*, et al.* Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. *Nature* **530**, 223-227 (2016).

13. Chan, C.K.*, et al.* Identification and specification of the mouse skeletal stem cell. *Cell* **160**, 285-298 (2015).

14. Dimov, I.K.*, et al.* Discriminating cellular heterogeneity using microwell-based RNA cytometry. *Nat Commun* **5**, 3451 (2014).

15. Chan, C.K.*, et al.* Clonal precursor of bone, cartilage, and hematopoietic niche stromal cells. *Proc Natl Acad Sci U S A* **110**, 12643-12648 (2013).

16. Seita, J.*, et al.* Gene Expression Commons: an open platform for absolute gene expression profiling. *PLoS One* **7**, e40321 (2012).

17. Dalerba, P.*, et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* **29**, 1120-1127 (2011).

18. Dalerba, P.*, et al.* CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. *N Engl J Med* **374**, 211-222 (2016).

19. Cheah, M.T., *et al.* CD14-expressing cancer cells establish the inflammatory and proliferative tumor microenvironment in bladder cancer. *Proc Natl Acad Sci U S A* **112**, 4725-4730 (2015).

20. Volkmer, J.P., *et al.* Three differentiation states risk-stratify bladder cancer into distinct subtypes. *Proc Natl Acad Sci U S A* **109**, 2078-2083 (2012).

21. Shin, K., *et al.* Hedgehog signaling restrains bladder cancer progression by eliciting stromal production of urothelial differentiation factors. *Cancer Cell* **26**, 521-533 (2014).

22. Sin, M.L.Y., *et al.* Deep Sequencing of Urinary RNAs for Bladder Cancer Molecular Diagnostics. *Clin Cancer Res* **23**, 3700-3710 (2017).

23. Sahoo, D., *et al.* Boolean analysis identifies CD38 as a biomarker of aggressive localized prostate cancer. *Oncotarget* **9**, 6550-6561 (2018).

24. Bhamre, S., Sahoo, D., Tibshirani, R., Dill, D.L. & Brooks, J.D. Temporal changes in gene expression induced by sulforaphane in human prostate cancer cells. *Prostate* **69**, 181-190 (2009).

25. Sahoo, D., *et al.* AI-guided discovery of the invariant host response to viral pandemics. *EBioMedicine* **68**, 103390 (2021).

26. Dabydeen, S.A., Desai, A. & Sahoo, D. Unbiased Boolean analysis of public gene expression data for cell cycle gene identification. *Mol Biol Cell* **30**, 1770-1779 (2019).

27. Dang, D., *et al.* Computational Approach to Identifying Universal Macrophage Biomarkers. *Frontiers in Physiology* **11**(2020).

28. Sahoo, D., *et al.* Artificial intelligence guided discovery of a barrier-protective therapy in inflammatory bowel disease. *Nat Commun* **12**, 4246 (2021).

29. Sahoo, D., *et al.* Artificial intelligence guided discovery of a barrier protective therapy in inflammatory bowel disease. *Nat Commun* (2021).

30. Cummings, L.C., *et al.* A nonrandomized trial of vitamin D supplementation for Barrett's esophagus. *PLoS One* **12**, e0184928 (2017).

31. Hyland, P.L., *et al.* Global changes in gene expression of Barrett's esophagus compared to normal squamous esophagus and gastric cardia tissues. *PLoS One* **9**, e93219 (2014).

32. Azouz, N.P., *et al.* The antiprotease SPINK7 serves as an inhibitory checkpoint for esophageal epithelial inflammatory responses. *Sci Transl Med* **10**(2018).

33. Nabokina, S.M., Ramos, M.B., Valle, J.E. & Said, H.M. Regulation of basal promoter activity of the human thiamine pyrophosphate transporter SLC44A4 in human intestinal epithelial cells. *Am J Physiol Cell Physiol* **308**, C750-757 (2015).

34. Redston, M., *et al.* Abnormal TP53 Predicts Risk of Progression in Patients With Barrett's Esophagus Regardless of a Diagnosis of Dysplasia. *Gastroenterology* (2021).

35. Kastelein, F., *et al.* Aberrant p53 protein expression is associated with an increased risk of neoplastic progression in patients with Barrett's oesophagus. *Gut* **62**, 1676-1683 (2013).

36. Snyder, P., *et al.* Aberrant p53 Immunostaining in Barrett's Esophagus Predicts Neoplastic Progression: Systematic Review and Meta-Analyses. *Dig Dis Sci* **64**, 1089-1097 (2019).

37. Bailey, T.*, et al.* Altered cadherin and catenin complexes in the Barrett's esophagus-dysplasia-adenocarcinoma sequence: correlation with disease progression and dedifferentiation. *Am J Pathol* **152**, 135-144 (1998).

38. Suchorolski, M.T., Paulson, T.G., Sanchez, C.A., Hockenbery, D. & Reid, B.J. Warburg and Crabtree effects in premalignant Barrett's esophagus cell lines with active mitochondria. *PLoS One* **8**, e56884 (2013).

39. Li, J.*, et al.* SPT6 promotes epidermal differentiation and blockade of an intestinal-like phenotype through control of transcriptional elongation. *Nat Commun* **12**, 784 (2021).

40. Truong, A.B., Kretz, M., Ridky, T.W., Kimmel, R. & Khavari, P.A. p63 regulates proliferation and differentiation of developmentally mature keratinocytes. *Genes Dev* **20**, 3185-3197 (2006).

41. Yang, A.*, et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714-718 (1999).

42. Senoo, M., Pinto, F., Crum, C.P. & McKeon, F. p63 Is essential for the proliferative potential of stem cells in stratified epithelia. *Cell* **129**, 523-536 (2007).

43. Crum, C.P. & McKeon, F.D. p63 in epithelial survival, germ cell surveillance, and neoplasia. *Annu Rev Pathol* **5**, 349-371 (2010).

44. Xian, W.*, et al.* The Cellular Origin of Barrett's Esophagus and Its Stem Cells. *Adv Exp Med Biol* **1123**, 55-69 (2019).

45. Minacapelli, C.D.*, et al.* Barrett's metaplasia develops from cellular reprograming of esophageal squamous epithelium due to gastroesophageal reflux. *Am J Physiol Gastrointest Liver Physiol* **312**, G615-G622 (2017).

46. Wang, D.H. & Souza, R.F. Transcommitment: Paving the Way to Barrett's Metaplasia. *Adv Exp Med Biol* **908**, 183-212 (2016).

47. Xian, W., Ho, K.Y., Crum, C.P. & McKeon, F. Cellular origin of Barrett's esophagus: controversy and therapeutic implications. *Gastroenterology* **142**, 1424-1430 (2012).

48. Dulak, A.M.*, et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet* **45**, 478-486 (2013).

49. Lagisetty, K.H.*, et al.* Immune determinants of Barrett's progression to esophageal adenocarcinoma. *JCI Insight* **6**(2021).

50. Karstens, K.F.*, et al.* Systemic interleukin 10 levels indicate advanced stages while interleukin 17A levels correlate with reduced survival in esophageal adenocarcinomas. *PLoS One* **15**, e0231833 (2020).

51. Karstens, K.F.*, et al.* Anti-inflammatory microenvironment of esophageal adenocarcinomas negatively impacts survival. *Cancer Immunol Immunother* **69**, 1043-1056 (2020).

52. Qian, Q., Wu, C., Chen, J. & Wang, W. Relationship between IL-10 and PD-L1 in esophageal carcinoma tissues and IL-10 down-regulates PD-L1 expression via Met signaling pathway. *Journal of Gastrointestinal Oncology* **11**, 337-355 (2020).

53. Xu, X.*, et al.* Downregulation of HuR Inhibits the Progression of Esophageal Cancer through Interleukin-18. *Cancer Res Treat* **50**, 71-87 (2018).

54. Zhu, J., *et al.* Association of the interleukin-18 receptor 1 and interleukin-18 receptor accessory protein polymorphisms with the risk of esophageal cancer. *Biomed Rep* **4**, 227-235 (2016).

55. Feng, X., *et al.* Interleukin-18 Is a Prognostic Marker and Plays a Tumor Suppressive Role in Colon Cancer. *Dis Markers* **2020**, 6439614 (2020).

56. Fabbi, M., Carbotti, G. & Ferrini, S. Context-dependent role of IL-18 in cancer biology and counter-regulation by IL-18BP. *J Leukoc Biol* **97**, 665-675 (2015).

57. Zhang, H., *et al.* The TP53-Related Signature Predicts Immune Cell Infiltration, Therapeutic Response, and Prognosis in Patients With Esophageal Carcinoma. *Front Genet* **12**, 607238 (2021).

58. Krause, L., *et al.* Identification of the CIMP-like subtype and aberrant methylation of members of the chromosomal segregation and spindle assembly pathways in esophageal adenocarcinoma. *Carcinogenesis* **37**, 356-365 (2016).

59. Bickel, M. The role of interleukin-8 in inflammation and mechanisms of regulation. *J Periodontol* **64**, 456-460 (1993).

60. Zeilhofer, H.U. & Schorr, W. Role of interleukin-8 in neutrophil signaling. *Curr Opin Hematol* **7**, 178-182 (2000).

61. Bernhard, S., *et al.* Interleukin 8 Elicits Rapid Physiological Changes in Neutrophils That Are Altered by Inflammatory Conditions. *J Innate Immun* **13**, 225-241 (2021).

62. Jablonska, J., Wu, C.F., Andzinski, L., Leschner, S. & Weiss, S. CXCR2-mediated tumor-associated neutrophil recruitment is regulated by IFN-beta. *Int J Cancer* **134**, 1346-1358 (2014).

63. Teijeira, A., *et al.* IL8, Neutrophils, and NETs in a Collusion against Cancer Immunity and Immunotherapy. *Clin Cancer Res* **27**, 2383-2393 (2021).

64. Haqqani, A.S., Sandhu, J.K. & Birnboim, H.C. Expression of interleukin-8 promotes neutrophil infiltration and genetic instability in mutatect tumors. *Neoplasia* **2**, 561-568 (2000).

65. Wang, X., Qiu, L., Li, Z., Wang, X.Y. & Yi, H. Understanding the Multifaceted Role of Neutrophils in Cancer and Autoimmune Diseases. *Front Immunol* **9**, 2456 (2018).

66. Shaul, M.E., *et al.* Tumor-associated neutrophils display a distinct N1 profile following TGFbeta modulation: A transcriptomics analysis of pro- vs. antitumor TANs. *Oncoimmunology* **5**, e1232221 (2016).

67. Damotte, D., *et al.* The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. *J Transl Med* **17**, 357 (2019).

68. Ayers, M., *et al.* IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* **127**, 2930-2940 (2017).

69. Wagener-Ryczek, S., *et al.* Immune profile and immunosurveillance in treatment-naive and neoadjuvantly treated esophageal adenocarcinoma. *Cancer Immunol Immunother* **69**, 523-533 (2020).

70. Kato, K., *et al.* KEYNOTE-590: Phase III study of first-line chemotherapy with or without pembrolizumab for advanced esophageal cancer. *Future Oncol* **15**, 1057-1066 (2019).

71. Sun, J.M., *et al.* Pembrolizumab plus chemotherapy versus chemotherapy alone for first-line treatment of advanced oesophageal cancer (KEYNOTE-590): a randomised, placebo-controlled, phase 3 study. *Lancet* **398**, 759-771 (2021).

72. Shah, M.A.*, et al.* KEYNOTE-975 study design: a Phase III study of definitive chemoradiotherapy plus pembrolizumab in patients with esophageal carcinoma. *Future Oncol* **17**, 1143-1153 (2021).

73. Of men, not mice. *Nat Med* **19**, 379 (2013).

74. Hellman, S. & Hellman, D.S. Of mice but not men. Problems of the randomized clinical trial. *N Engl J Med* **324**, 1585-1589 (1991).

75. Warren, H.S.*, et al.* Mice are not men. *Proc Natl Acad Sci U S A* **112**, E345 (2015).

76. Mestas, J. & Hughes, C.C. Of mice and not men: differences between mouse and human immunology. *J Immunol* **172**, 2731-2738 (2004).

77. Seok, J.*, et al.* Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A* **110**, 3507-3512 (2013).

78. Kunze, B.*, et al.* Notch signaling drives development of Barrett's metaplasia from Dclk1-positive epithelial tuft cells in the murine gastric mucosa. *Sci Rep* **11**, 4509 (2021).

79. Kunze, B.*, et al.* Notch Signaling Mediates Differentiation in Barrett's Esophagus and Promotes Progression to Adenocarcinoma. *Gastroenterology* **159**, 575-590 (2020).

80. Munch, N.S.*, et al.* High-Fat Diet Accelerates Carcinogenesis in a Mouse Model of Barrett's Esophagus via Interleukin 8 and Alterations to the Gut Microbiome. *Gastroenterology* **157**, 492-506 e492 (2019).

81. Quante, M.*, et al.* Bile acid and inflammation activate gastric cardia stem cells in a mouse model of Barrett-like metaplasia. *Cancer Cell* **21**, 36-51 (2012).

82. Matsuzaki, J.*, et al.* Bile acids increase levels of microRNAs 221 and 222, leading to degradation of CDX2 during esophageal carcinogenesis. *Gastroenterology* **145**, 1300-1311 (2013).

83. Campos, V.J., Mazzini, G.S., Juchem, J.F. & Gurski, R.R. Neutrophil-Lymphocyte Ratio as a Marker of Progression from Non-Dysplastic Barrett's Esophagus to Esophageal Adenocarcinoma: a Cross-Sectional Retrospective Study. *J Gastrointest Surg* **24**, 8-18 (2020).

84. Peleg, N.*, et al.* Neutrophil to lymphocyte ratio and risk of neoplastic progression in patients with Barrett's esophagus. *Endoscopy* **53**, 774-781 (2021).

85. Zhang, X.*, et al.* Prognostic role of neutrophil-lymphocyte ratio in esophageal cancer: A systematic review and meta-analysis. *Medicine (Baltimore)* **97**, e13585 (2018).

86. Morris, M.C., Nadeem Khan, M. & Pichichero, M.E. A PCR-based method for quantifying neutrophils in human nasal secretions. *J Immunol Methods* **447**, 65-70 (2017).

87. Ferrer-Torres, D.*, et al.* Constitutively Higher Level of GSTT2 in Esophageal Tissues From African Americans Protects Cells Against DNA Damage. *Gastroenterology* **156**, 1404-1415 (2019).

88. Thobakgale, C.F. & Ndung'u, T. Neutrophil counts in persons of African origin. *Curr Opin Hematol* **21**, 50-57 (2014).

89. Hsieh, M.M., Everhart, J.E., Byrd-Holt, D.D., Tisdale, J.F. & Rodgers, G.P. Prevalence of neutropenia in the U.S. population: age, sex, smoking status, and ethnic differences. *Ann Intern Med* **146**, 486-492 (2007).

90. Jain, D.*, et al.* Genome-wide association of white blood cell counts in Hispanic/Latino Americans: the Hispanic Community Health Study/Study of Latinos. *Hum Mol Genet* **26**, 1193-1204 (2017).

91. Rappoport, N., Simon, A.J., Amariglio, N. & Rechavi, G. The Duffy antigen receptor for chemokines, ACKR1,- 'Jeanne DARC' of benign neutropenia. *Br J Haematol* **184**, 497-507 (2019).

92. Reich, D.*, et al.* Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet* **5**, e1000360 (2009).

93. Thomas, P.*, et al. Research Square* (2021).

94. Li, X.*, et al.* Temporal and spatial evolution of somatic chromosomal alterations: a case-cohort study of Barrett's esophagus. *Cancer Prev Res (Phila)* **7**, 114-127 (2014).

95. Charles, B.A.*, et al.* Analyses of genome wide association data, cytokines, and gene expression in African-Americans with benign ethnic neutropenia. *PLoS One* **13**, e0194400 (2018).

96. Nibbs, R.J. & Graham, G.J. Immune regulation by atypical chemokine receptors. *Nat Rev Immunol* **13**, 815-829 (2013).

97. Pruenster, M. & Rot, A. Throwing light on DARC. *Biochem Soc Trans* **34**, 1005-1008 (2006).

98. Bonecchi, R.*, et al.* Non-signaling chemokine receptors: mechanism of action and role in vivo. *J Neuroimmunol* **198**, 14-19 (2008).

99. Lee, J.S.*, et al.* Duffy antigen facilitates movement of chemokine across the endothelium in vitro and promotes neutrophil transmigration in vitro and in vivo. *J Immunol* **170**, 5244-5251 (2003).

100. Reiner, A.P.*, et al.* Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. *Am J Hum Genet* **76**, 463-477 (2005).

101. Kangelaris, K.N.*, et al.* The association between a Darc gene polymorphism and clinical outcomes in African American patients with acute lung injury. *Chest* **141**, 1160-1169 (2012).

102. Rappoport, N.*, et al.* Correlation between 'ACKR1/DARC null' polymorphism and benign neutropenia in Yemenite Jews. *Br J Haematol* **170**, 892-895 (2015).

103. Guerra, C.A., Snow, R.W. & Hay, S.I. Mapping the global extent of malaria in 2005. *Trends Parasitol* **22**, 353-358 (2006).

104. Rippey, J. Leucopenia in West Indians and Africans. *The Lancet* **290**, 44 (1967).

105. Weingarten, M.A., Pottick-Schwartz, E. & Brauner, A. The epidemiology of benign leukopenia in Yemenite Jews. *Israel journal of medical sciences* **29**, 297-299 (1993).

106. Shoenfeld, Y., Alkan, M.L., Asaly, A., Carmeli, Y. & Katz, M. Benign familial leukopenia and neutropenia in different ethnic groups. *European journal of haematology* **41**, 273-277 (1988).

107. Denic, S.*, et al.* Prevalence, phenotype and inheritance of benign neutropenia in Arabs. *BMC Hematology* **9**, 1-8 (2009).

108. Sabel, M.S., Pastore, K., Toon, H. & Smith, J.L. Adenocarcinoma of the esophagus with and without Barrett mucosa. *Arch Surg* **135**, 831-835; discussion 836 (2000).

109. Curtius, K., Rubenstein, J.H., Chak, A. & Inadomi, J.M. Computational modelling suggests that Barrett's oesophagus may be the precursor of all oesophageal adenocarcinomas. *Gut* **70**, 1435-1440 (2021).

110. Nowicki-Osuch, K.*, et al.* Molecular phenotyping reveals the identity of Barrett's esophagus and its malignant transition. *Science* **373**, 760-767 (2021).

111. Peters, Y.*, et al.* Incidence of Progression of Persistent Nondysplastic Barrett's Esophagus to Malignancy. *Clin Gastroenterol Hepatol* **17**, 869-877 e865 (2019).

112. Ogura, M.*, et al.* Clinical significance of CXCL-8/CXCR-2 network in esophageal squamous cell carcinoma. *Surgery* **154**, 512-520 (2013).

113. Krzystek-Korpacka, M.*, et al.* Elevation of circulating interleukin-8 is related to lymph node and distant metastases in esophageal squamous cell carcinomas--implication for clinical evaluation of cancer patient. *Cytokine* **41**, 232-239 (2008).

114. Nguyen, G.H.*, et al.* Inflammatory and microRNA gene expression as prognostic classifier of Barrett's-associated esophageal adenocarcinoma. *Clin Cancer Res* **16**, 5824-5834 (2010).

115. Alvarez, H.*, et al.* Widespread hypomethylation occurs early and synergizes with gene amplification during esophageal carcinogenesis. *PLoS Genet* **7**, e1001356 (2011).

116. Lukaszewicz-Zajac, M., Paczek, S. & Mroczko, B. The significance of chemokine CXCL-8 in esophageal carcinoma. *Arch Med Sci* **16**, 475-480 (2020).

117. Goldblum, J.R. Barrett's esophagus and Barrett's-related dysplasia. *Mod Pathol* **16**, 316-324 (2003).

118. Zhang, H.*, et al.* Preoperative leukocytosis is associated with increased tumor-infiltrating neutrophil extracellular traps and worse outcomes in esophageal cancer. *Ann Transl Med* **8**, 441 (2020).

119. Thrift, A.P. & El-Serag, H.B. Sex and Racial Disparity in Incidence of Esophageal Adenocarcinoma: Observations and Explanations. *Clin Gastroenterol Hepatol* **14**, 330-332 (2016).

120. Shaul, M.E. & Fridlender, Z.G. Tumour-associated neutrophils in patients with cancer. *Nat Rev Clin Oncol* **16**, 601-620 (2019).

121. Jaillon, S.*, et al.* Neutrophil diversity and plasticity in tumour progression and therapy. *Nat Rev Cancer* **20**, 485-503 (2020).

122. Hedrick, C.C. & Malanchi, I. Neutrophils in cancer: heterogeneous and multifaceted. *Nat Rev Immunol* (2021).

123. Masucci, M.T., Minopoli, M. & Carriero, M.V. Tumor Associated Neutrophils. Their Role in Tumorigenesis, Metastasis, Prognosis and Therapy. *Front Oncol* **9**, 1146 (2019).

124. Faget, J., Peters, S., Quantin, X., Meylan, E. & Bonnefoy, N. Neutrophils in the era of immune checkpoint blockade. *J Immunother Cancer* **9**(2021).

125. Valero, C.*, et al.* Pretreatment neutrophil-to-lymphocyte ratio and mutational burden as biomarkers of tumor response to immune checkpoint inhibitors. *Nat Commun* **12**, 729 (2021).

126. Phuah, C.L.*, et al.* Genetic variants influencing elevated myeloperoxidase levels increase risk of stroke. *Brain* **140**, 2663-2672 (2017).

127. Reiner, A.P.*, et al.* Genome-wide and gene-centric analyses of circulating myeloperoxidase levels in the charge and care consortia. *Hum Mol Genet* **22**, 3381-3393 (2013).

128. Moreno Velasquez, I.*, et al.* Duffy antigen receptor genetic variant and the association with Interleukin 8 levels. *Cytokine* **72**, 178-184 (2015).

129. Sippert, E.A., de Oliveira e Silva, C., Visentainer, J.E. & Sell, A.M. Association of duffy blood group gene polymorphisms with IL8 gene in chronic periodontitis. *PLoS One* **8**, e83286 (2013).

130. Fukuma, N.*, et al.* A role of the Duffy antigen for the maintenance of plasma chemokine concentrations. *Biochem Biophys Res Commun* **303**, 137-139 (2003).

131. Lee, J.S.*, et al.* The Duffy antigen modifies systemic and local tissue chemokine responses following lipopolysaccharide stimulation. *J Immunol* **177**, 8086-8094 (2006).

132. Mayr, F.B.*, et al.* Duffy antigen modifies the chemokine response in human endotoxemia. *Crit Care Med* **36**, 159-165 (2008).

133. Gupta, S.*, et al.* Sex differences in neutrophil biology modulate response to type I interferons and immunometabolism. *Proc Natl Acad Sci U S A* **117**, 16481-16491 (2020).

134. Newburger, P.E. & Dale, D.C. Evaluation and management of patients with isolated neutropenia. *Semin Hematol* **50**, 198-206 (2013).

135. Grann, V.R.*, et al.* Duffy (Fy), DARC, and neutropenia among women from the United States, Europe and the Caribbean. *Br J Haematol* **143**, 288-293 (2008).

136. Cameron, A.J., Lomboy, C.T., Pera, M. & Carpenter, H.A. Adenocarcinoma of the esophagogastric junction and Barrett's esophagus. *Gastroenterology* **109**, 1541-1546 (1995).

137. Chak, A.*, et al.* Familiality in Barrett's esophagus, adenocarcinoma of the esophagus, and adenocarcinoma of the gastroesophageal junction. *Cancer Epidemiol Biomarkers Prev* **15**, 1668-1673 (2006).

138. Cook, M.B.*, et al.* Cigarette smoking and adenocarcinomas of the esophagus and esophagogastric junction: a pooled analysis from the international BEACON consortium. *J Natl Cancer Inst* **102**, 1344-1353 (2010).

139. Sahoo, D. The power of boolean implication networks. *Front Physiol* **3**, 276 (2012).

140. Sahoo, D., Dill, D.L., Tibshirani, R. & Plevritis, S.K. Extracting binary signals from microarray time-course data. *Nucleic Acids Res* **35**, 3705-3712 (2007).

141. Sahoo, D. Artificial Intelligence guided Discovery of a Barrier-Protective Therapy in Inflammatory Bowel Disease. *GitHub/sahoo00/Hegemon* (2021).

142. Reid, B.J., Blount, P.L., Feng, Z. & Levine, D.S. Optimizing endoscopic biopsy detection of early cancers in Barrett's high-grade dysplasia. *Am J Gastroenterol* **95**, 3089-3096 (2000).

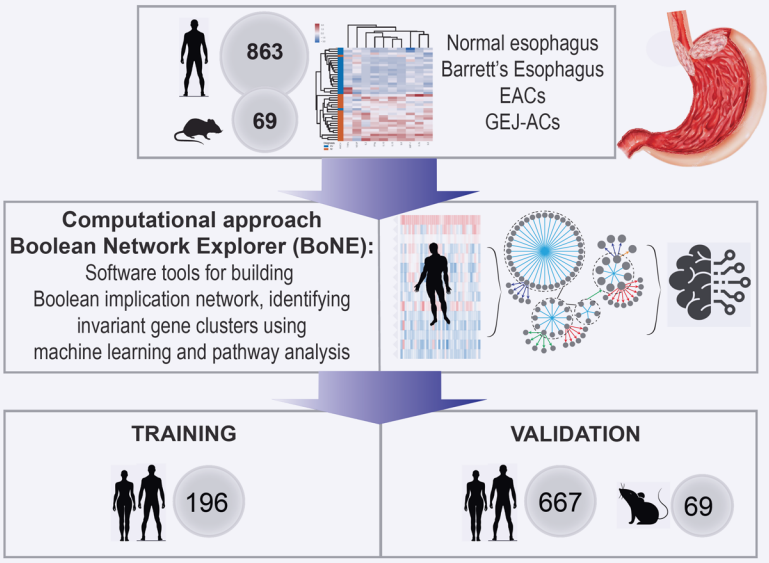
143. Levine, D.S.*, et al.* An endoscopic biopsy protocol can differentiate high-grade dysplasia from early adenocarcinoma in Barrett's esophagus. *Gastroenterology* **105**, 40-50 (1993).

144. Vo, D.T.*, et al.* SPT6 loss permits the transdifferentiation of keratinocytes into an intestinal fate that resembles Barrett's metaplasia. *iScience* **24**, 103121 (2021).

FIGURES AND LEGENDS

Step 1- A NETWORK TRANSCRIPTOMIC APPROACH TO IDENTIFY THE CELL STATES AND PROCESSES DURING THE INITIATION AND PROGRESSION OF ESOPHAGEAL ADENO-CARCINOMAS

Artificial Intelligence- guided Discovery of Patterns of Progressive Gene Expression Changes during Barret's metaplasia of the Esophagus (BE) and during the progression of BE to Esophageal Adenocarcinomas (EACs).



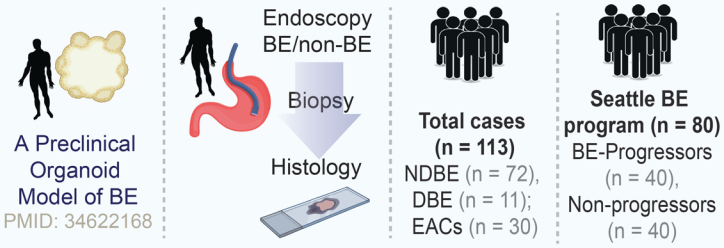
Step 2- VALIDATION OF FINDINGS

(i) RNA Seq and ATAC seq studies on a human pre-clinical model of BE.

(ii) IHC on esophageal biopsies.

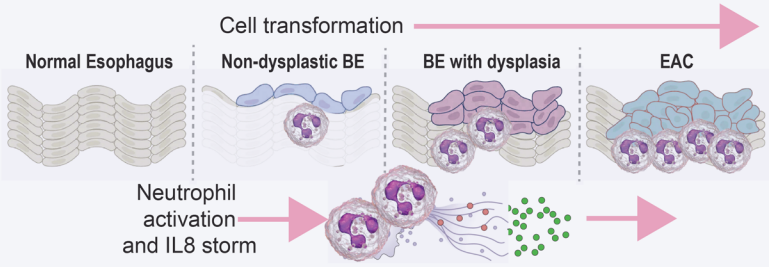
(iii) Cross-sectional retrospective study (n = 113).

(iv) Case-control genomics study (n = 40).



Step 3- KEY FINDINGS-I: PRECISE IDENTIFICATION OF THE NATURE OF THE TUMOR MICRO-ENVIRONMENT in E/GEJ-ACs

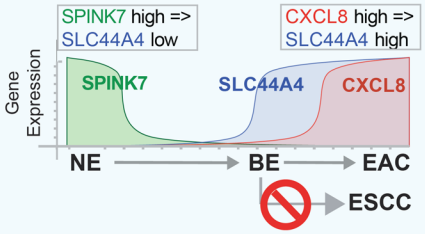
Both EACs and GEJ-ACs share a CXCL8/IL8 and neutrophil-centric immune infiltration. Mouse models of EAC rarely replicate this.



Step 4- KEY FINDINGS-II: INSIGHTS INTO THE CELL OF ORIGIN OF EACs

Boolean logical model of the metaplasia (NE>BE)⟶ dysplasia cascade (BE>EAC) revealed that all EACs (but not ESCCs) must arise from the metaplastic BE epithelium.

A Boolean logical model of the metaplasia ⟶ dysplasia ⟶ neoplasia cascade in the esophagus



Step 5- KEY FINDINGS-III: INSIGHTS INTO THE BASIS FOR RACIAL/ETHNIC DISPARITY

Neutropenia-associated SNPs are enriched in BE-non-progressors; neutrophilia-associated SNPs are enriched in BE-progressors.
Benign ethnic neutropenia (BEN) may protect AAs (compared to Cau).

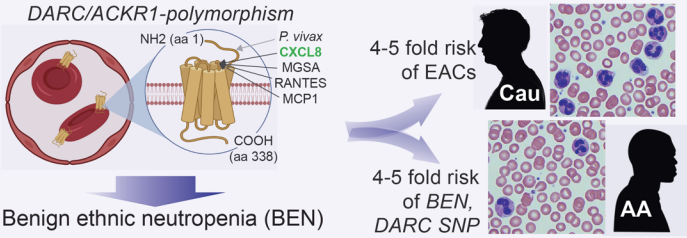


Figure 1. Study Design.

Step 1: A database containing 932 gene–expression data from both human (n = 863) and mouse (n = 69) samples was mined to build a validated Boolean implication network-based computational model of disease continuum states during the metaplastic→dysplastic→neoplastic cascade in the squamous epithelial lining of the esophagus. Paths, clusters, and list of genes in the network-based model are first prioritized by machine learning approaches and used subsequently to discover cell types and cellular states that fuel the cascade.

Step 2: Schematic displays the four different models and approaches were used to validate the network-derived findings.

Step 3-5: The network-based model reveals three key findings.

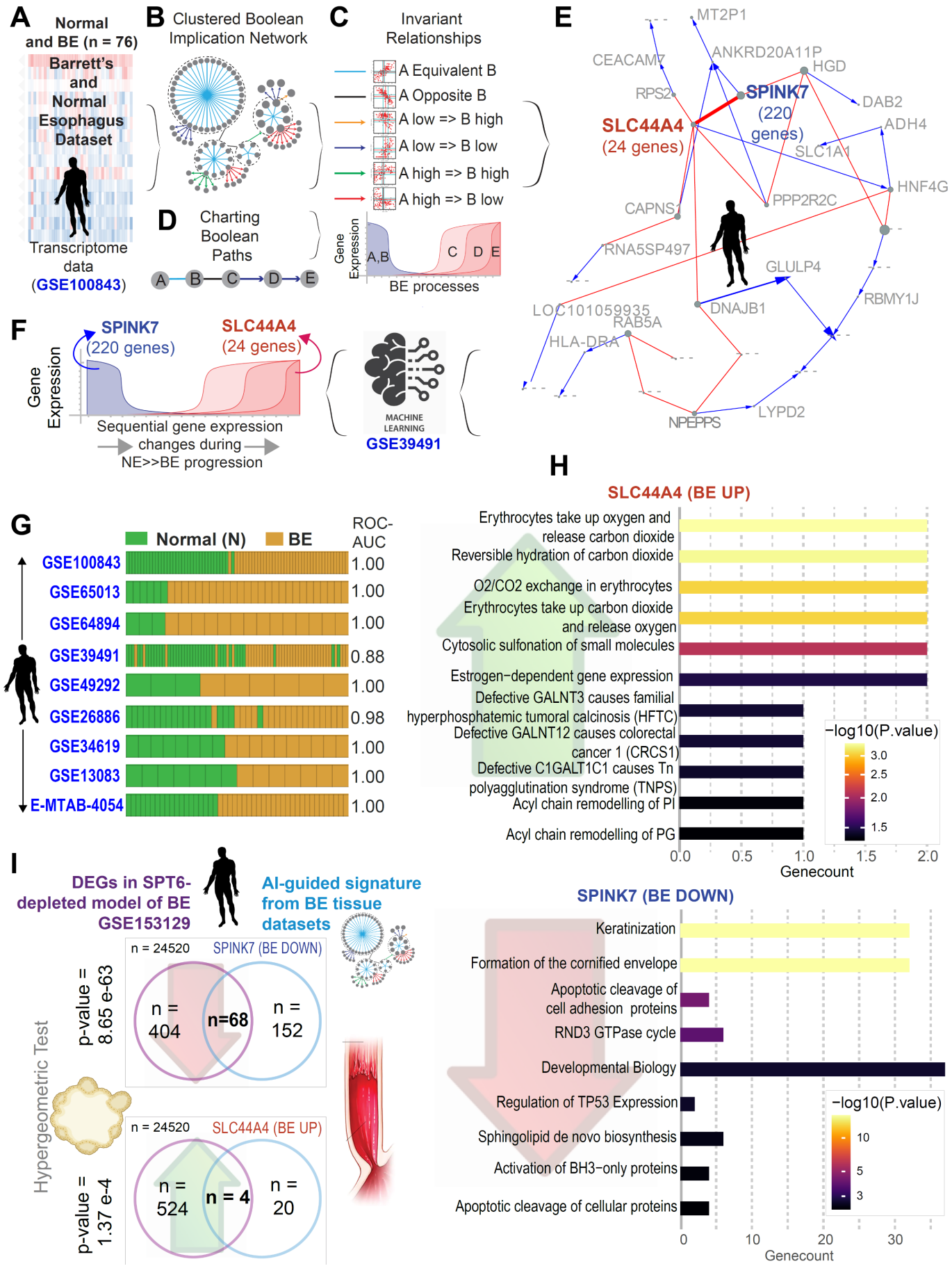


Figure 2. Generation and validation of Boolean network map of Barrett's metaplasia of the esophagus.

A-D. Schematics outline the workflow and training datasets used to create a Boolean map of changing expression of gene clusters during normal esophagus (NE) to BE transition using *BoNE*[29]. We applied *BoNE* to analyze dataset GSE100843 (A). The Boolean network (B) contains the six possible Boolean relationships (C) in the form of a directed graph with genes having similar expression profiles organized into clusters, and relationships between gene clusters represented as color-coded edges (B). Boolean cluster relationships are used to chart disease paths (D-left); and individual gene expression changes along a boolean path, illustrating gene expression dynamics within the normal to BE to EAC continuum (D-right).

E. Graph showing invariant patterns (down or up-regulation) of gene expression changes during NE>BE progression. Two clusters of genes (*SPINK7* and *SLC44A4*) identified by machine learning are indicated in bold.

F. Gene clusters in E were filtered through a second dataset [GSE39491] to refine the gene signature (left). The resultant signature involves progressive downregulation of *SPINK7*-cluster with a concomitant upregulation of *SLC44A4*-cluster *(right)*.

G. Bar plots showing sample classification accuracy across diverse BE validation datasets, with corresponding ROC-AUC values.

H. Reactome pathway analysis of *SPINK7*- and *SLC44A4*-clusters were performed to identify the signaling pathways and cellular processes that are enriched during metaplastic progression in esophagus.

I. Gene signatures UP- or DOWN-regulated in BE-map (which is built using tissue transcriptomics) are compared against differentially expressed genes (DEGs) in our recently published SPT6-depleted organoid model of BE[8]. Hypergeometric statistical analyses show significant overlaps in both UP and DOWN-regulated genes.

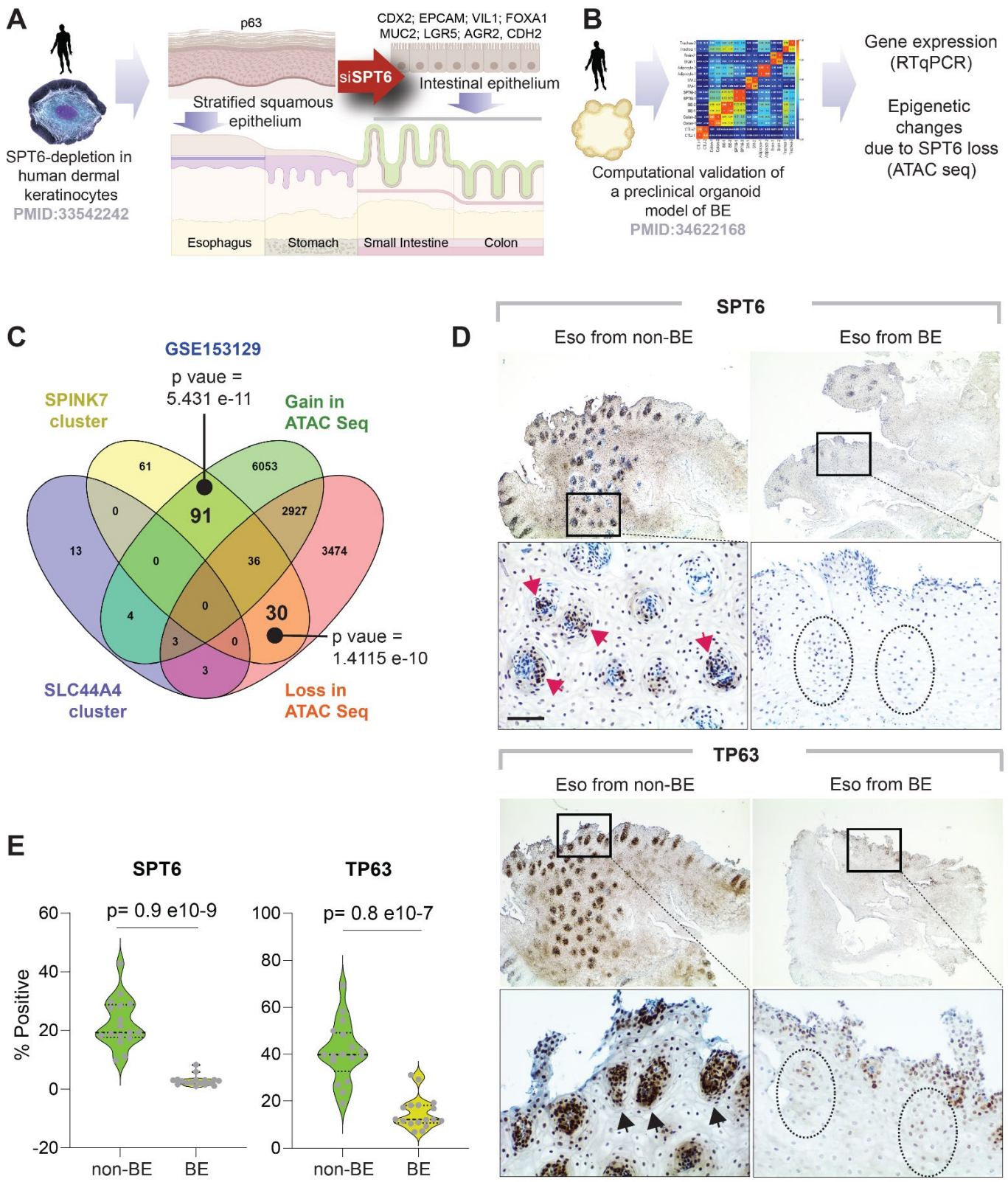


Figure 3. Validation of the Boolean map of Barrett's esophagus using a human organoid model of squamous metaplasia.

formed stratified squamous epithelium, siRNA mediated transient depletion of SPT6 in keratinocytes (SPT6i) grew as 'intestine-like' monolayers. RNA seq studies on those monolayers confirmed that this model recapitulates metaplastic gene signatures in BE, and not normal gut differentiation that is observed in the healthy gut lining.

C. Gene clusters from the BE-map were analyzed for overlap with those affected in the organoid BE model by loss or gain of access due to chromatin remodeling upon SPT6 loss, as identified by ATAC Seq. Only significant *p values*, as determined using hypergeometric analyses, are displayed.

D. Esophageal biopsies obtained 2 cm above the GEJ from Caucasian males who had BE (Eso from BE) or had not (Eso from non-BE) progressed to BE were analyzed for SPT6 and TP63 expression by IHC. Red and black arrowheads point to crypts staining positive. Interrupted circles highlight crypts with little/no expression. Fields representative from n = 3 subjects are shown; boxed regions above are magnified below; Scale bar = 100 μm.

E. Violin plots display the % cells positive for staining in regions of interest (ROIs, marked with arrows on left and interrupted ovals on the right), as determined by the ImageJ plug-in, IHC profiler.

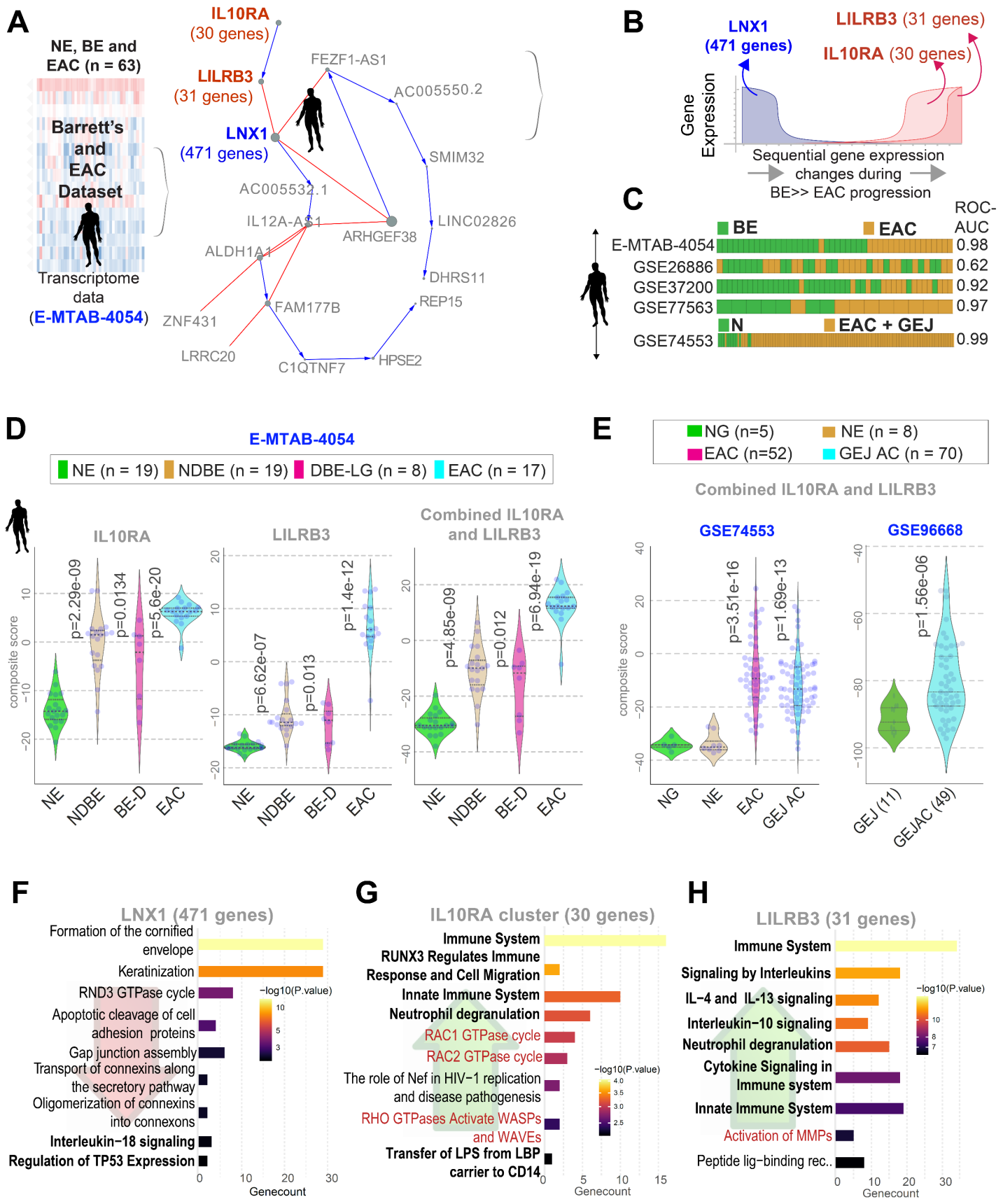


Figure 4. Generation and validation of Boolean network map of BE to EAC progression.

B. Graph showing invariant patterns (down or up-regulation) of gene expression changes during BE>EAC transformation. Three clusters of genes identified by machine learning are indicated.

C. Bar plots showing sample classification accuracy across diverse validation datasets, with corresponding ROC-AUC values.

D. Violin plots showing the composite scores of upregulated gene clusters in normal esophagus (NE), non-dysplastic BE (NDBE), dysplastic BE (DBE), and EACs. *P* values indicate comparison of each sample type against the normal esophagus (NE), as determined by Welch's t-test.

E. Violin plots showing the composite scores of upregulated gene clusters in normal esophagus (NE), normal gastric (NG), normal GEJ (GEJ) and GEJ-ACs. *P* values indicate comparison of each sample type against the normal esophagus (NE; left, GSE74553) or GEJ (right, GSE96668), as determined by Welch's t-test.

F-H. Pathway analyses (reactome.org) of gene clusters reveal cellular processes that are UP or DOWN during cell transformation in 3 clusters derived from the EAC map in B. Red fonts indicate likely epithelial processes.

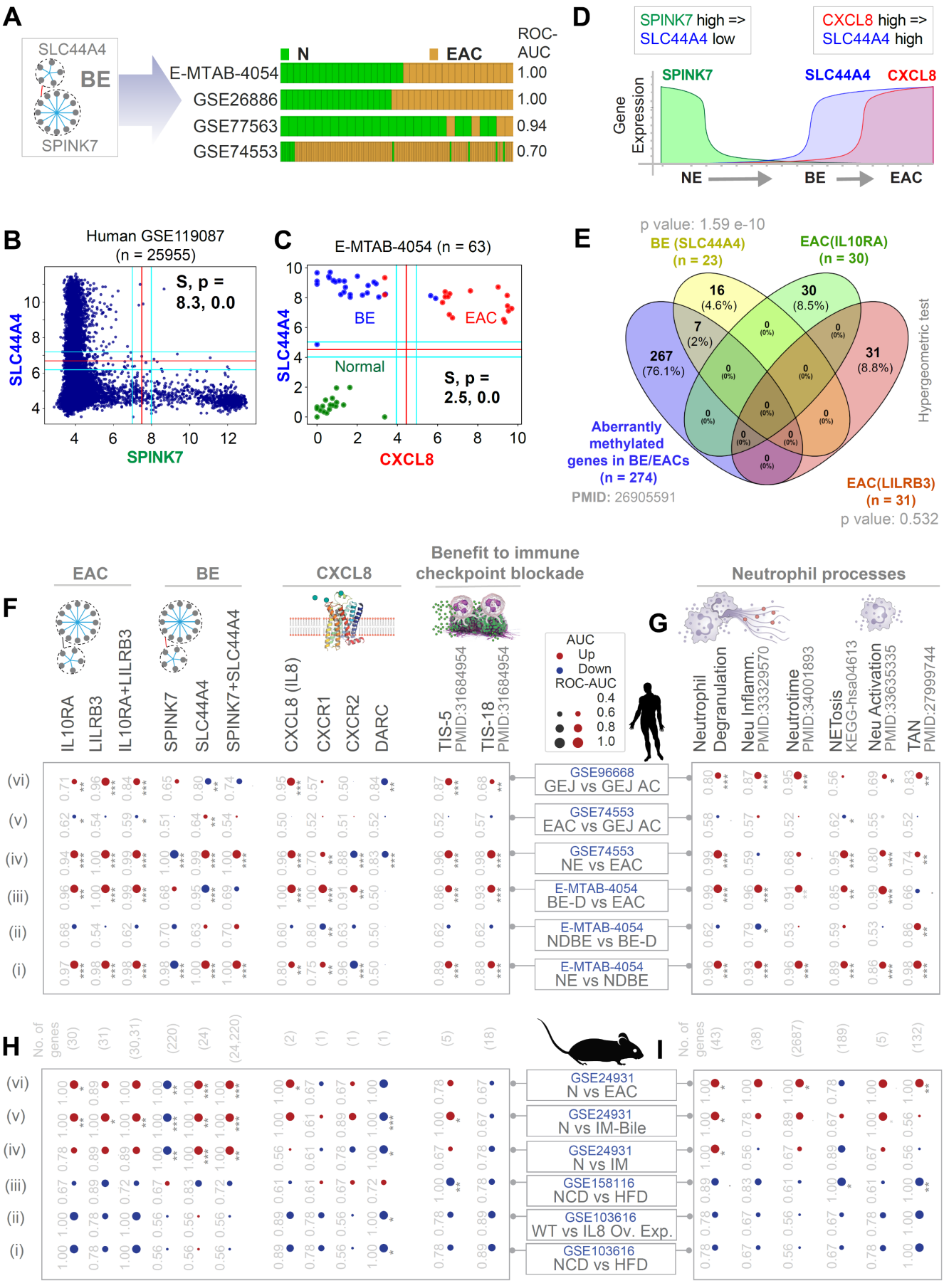


Figure 5. A Boolean logical model of cellular states during NE→BE→EAC progression.

A. Bar plots showing that BE signatures (NE>>BE map; **Figure 2F-G**) can distinguish normal esophagus (NE) and EAC samples, indicating that these signatures are also induced in EAC samples. Corresponding ROC AUC values are displayed.

B. A scatterplot for *SPINK7* and *SLC44A4* expression in global human GSE119087 (n = 25955) dataset. Boolean Implication SPINK7 high => SLC44A7 low (S=8.3, p=0.0, FDR < 0.001) is an invariant in the most diverse dataset.

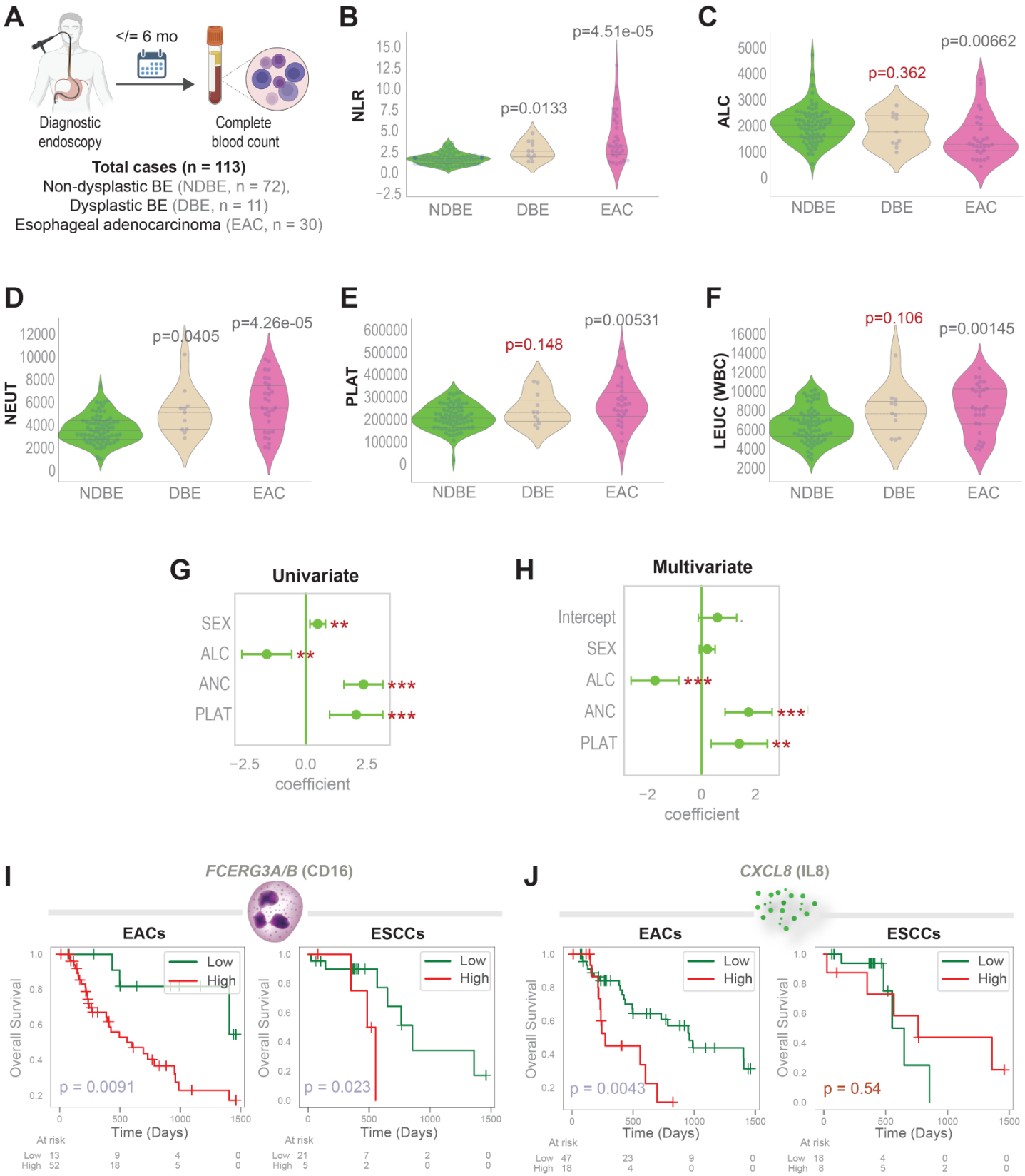
C. A scatterplot for *CXCL8* and *SLC44A4* expression in the E-MTAB-4045 dataset. *CXCL8* high => *SLC44A4* high (S = 2.5, p = 0.0, FDR < 0.001) in an invariant Boolean implication relationship in NE, BE and EAC samples where each sample type is mostly confined to one quadrant (NE, bottom-left; BE, top-left; and EAC, top-right).

D. A schematic to visualize the mathematical model of NE→BE→EAC progression based on MiDReG[10,11] analysis using Boolean Implication relationships. Since *CXCL8* high => *SLC44A4* high, and both *CXCL8* and *SLC44A4* are low in NE, the invariant model suggests that BE stage (*SLC44A4* high, *CXCL8* low) must precede *CXCL8* high and *SLC44A4* high. See also **Figure S1** for model validation in an independent cohort of EACs (**S1A**) and for a head-to-head comparison between ESCC vs EAC (**S1B**).

E. Venn diagram showing the overlaps between the gene clusters from the BE/EAC-maps with the genes reported to be methylated in multiple independent studies (n = 274 / 22178 genes that were tested in total) using Illumina's Infinium HumanMethylation27 BeadChip microarray. Only significant *p values*, as determined using hypergeometric analyses, are displayed.

F-G. Human EAC immune microenvironment is visualized as bubble plots of ROC-AUC values (radius of circles are based on the ROC-AUC) demonstrating the direction of gene regulation (Up, red; Down, blue) for the classification of samples (gene signatures in columns; dataset and sample comparison in rows). P values based on Welch's T-test (of composite score of gene expression values) are provided using standard code (‘’ p> 0.1, ‘.’ p <= 0.1, * P <= 0.05, ** P <= 0.01, *** P <= 0.001) next to the ROC-AUC. Panel F displays the classification of NE, non-dysplastic BE (NDBE), dysplastic BE (DBE), EAC, GEJ-AC based on signatures derived from the BE>>EAC map (upregulated gene clusters *IL10RA* and *LILRB3*), NE>>BE map (downregulated *SPINK7*-cluster, and upregulated *SLC44A4*-cluster), CXCL8/IL8 and its three receptors (CXCR1/2 and DARC), and tumor inflammation signatures that predict anti-PD1 Rx (TIS-5, TIS-18) in two independent dataset (E-MTAB-4054, GSE74553). Panel G displays the classification of same samples in F based on neutrophil signatures (NETosis, Degranulation, Inflammation, Neutrotime) in E-MTAB-4054, and GSE74553. Violin plots for selected neutrophil signatures across NE, NDBE, DBE, EAC, GEJ-AC samples are displayed in **Figure S2**. See also **Figure S3** for extended data on the relationship between neutrophil and EAC signatures with the TIS signature.

H-I. Murine BE-EAC immune microenvironment was analyzed using the same set of signatures and visualized as bubble plots exactly as in F-G. IM, intestinal metaplasia. NCD, normal chow diet; HFD, high fat diet; Ov Exp, overexpression. Violin plots for selected signatures are displayed in **Figure S4.**



A. Schematic summarizing the cross-sectional study design and cohort composition that is analyzed in panels B-H.

B-F. Violin plots display the neutrophil:lymphocyte ratio (NLR, B), or absolute lymphocyte (ALC, C), neutrophil (NEUT, D), platelet (PLAT, E) and leukocyte (LEUK, F) counts in various patients within each diagnostic group shown in panel A. *P* values indicate comparison of each subgroup against the non-dysplastic BE (NDBE) group, as determined by Welch's t-test.

G-H. Univariate (G) and multivariate (H) analyses models the risk of BE to EAC progression as a linear combination of sex and the indicated hematologic parameters. Coefficient of each variable (at the center) with 95% confidence intervals (as error bars) and the *p* values were illustrated in the bar plot. The *p* value for each term tests the null hypothesis that the coefficient is equal to zero (no effect). $^{**}p \leq 0.01$; $^{***}p \leq 0.001$.

I-J. Schematic (left) indicates the chemoattractant *CXCL8* (J) and CD16 (I; a heterodimer that is encoded by the genes *FCERG3A/B*), the marker of TAN abundance that is assessed here. Kaplan-Meier (KM) plots display the overall survival of patients with tumors stratified based on the high vs low composite scores of 2 genes (*FCERG3A, FCERG3B)* and the high vs low expression values of *CXCL8.* High and low values are obtained using StepMiner threshold +/- noise margin (See supplementary methods for detail). *p* values were determined by logrank analysis. See also **Figure S5** for KM plots on these and other signatures, analyzed in EACs, ESCCs and gastric adenocarcinomas (GCs).

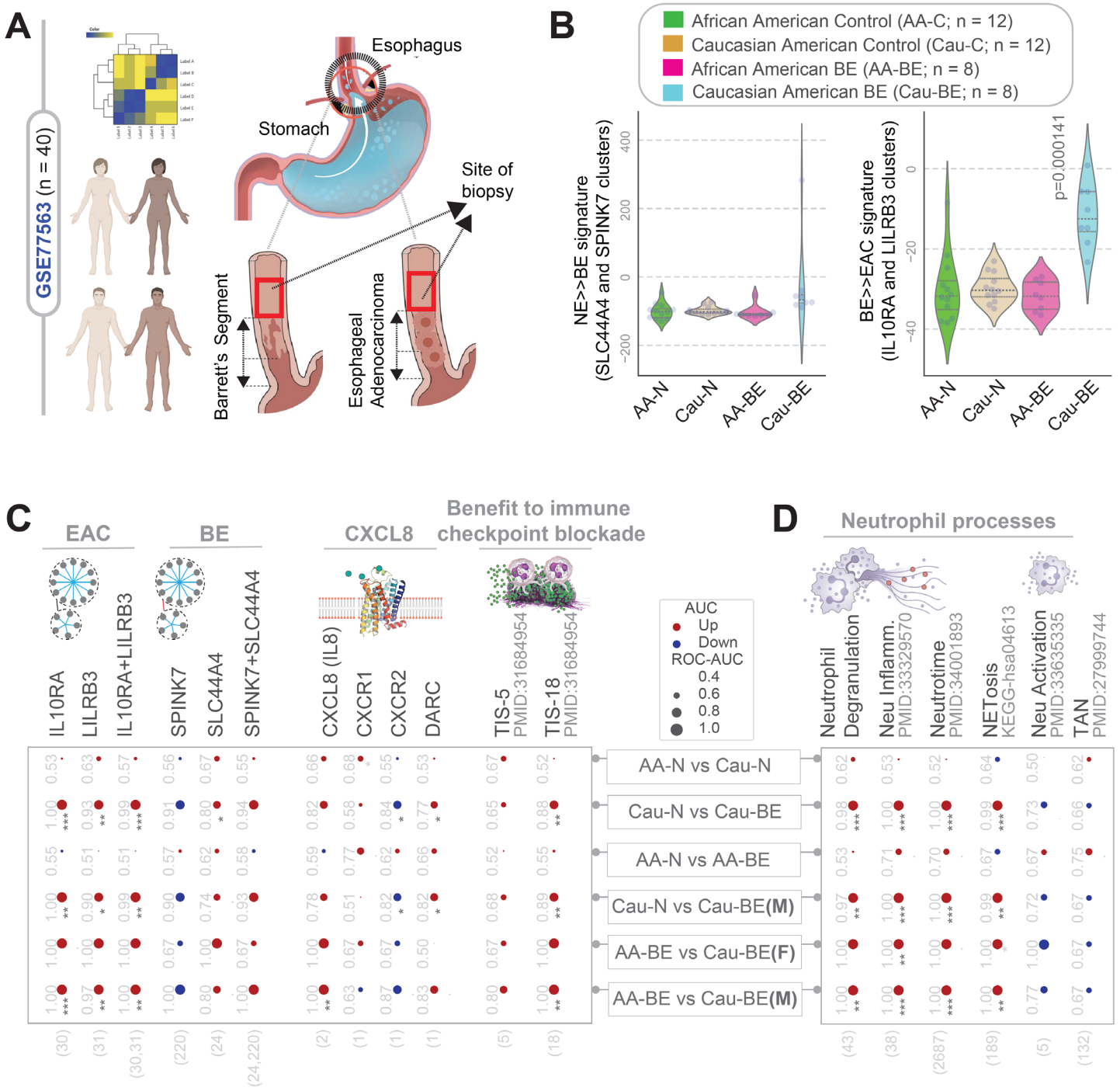
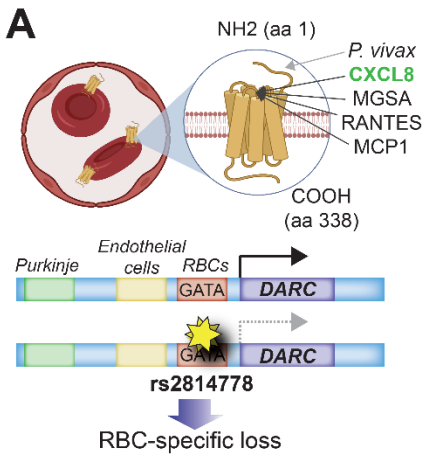


Figure 7. Caucasians (Cau), but not African Americans (AAs) mount IL8- and neutrophil-centric inflammation.

GSE77563. Microarray studies were conducted on histologically normal squamous mucosa arising within the tubular esophagus above the gastrointestinal junction from self-identified AA or Cau subjects, who had no history of BE or EAC (normal controls), or those diagnosed with BE and/or EAC (AA-BE or Cau-BE).

B. Violin plots showing the composite scores of upregulated gene clusters (*Left*, BE signatures; *Right*, EAC signatures) in normal esophagus from control AA or Cau subjects (AA-N and Cau-N) and esophagus from AA/Cau subjects diagnosed with BE/EACs (AA-BE and Cau-BE). *P* values indicate comparison of each sample type against the normal samples, as determined by Welch's t-test.

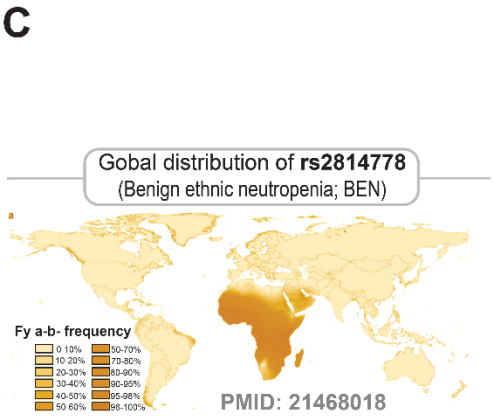
C-D. Human EAC immune microenvironment is visualized as bubble plots of ROC-AUC values (radius of circles are based on the ROC-AUC) demonstrating the direction of gene regulation (Up, red; Down, blue) for the classification of samples (gene signatures in columns; dataset and sample comparison in rows). P values based on Welch's T-test (of composite score of gene expression values) are provided using standard code (‘’ p> 0.1, ‘.’ p <= 0.1, * P <= 0.05, ** P <= 0.01, *** P <= 0.001) next to the ROC-AUC. Panel C displays the classification of AA vs Cau samples from control (AA/Cau-N) or BE/EAC subjects (AA/Cau-BE) in males (M) or female subjects (F) based on signatures derived from the BE>>EAC map (upregulated gene clusters *IL10RA* and *LILRB3*), NE>>BE map (downregulated *SPINK7*-cluster, and upregulated *SLC44A4*-cluster), CXCL8/IL8 and its three receptors (CXCR1/2 and DARC), and tumor inflammation signatures that predict anti-PD1 Rx (TIS-5, TIS-18) in GSE77563. Panel D displays the classification of same samples in C based on neutrophil signatures (NETosis, Degranulation, Inflammation, Neutrotime) in GSE77563. Violin plots for selected neutrophil signatures in AA-BE vs Cau-BE samples are displayed in **Figure S6**.

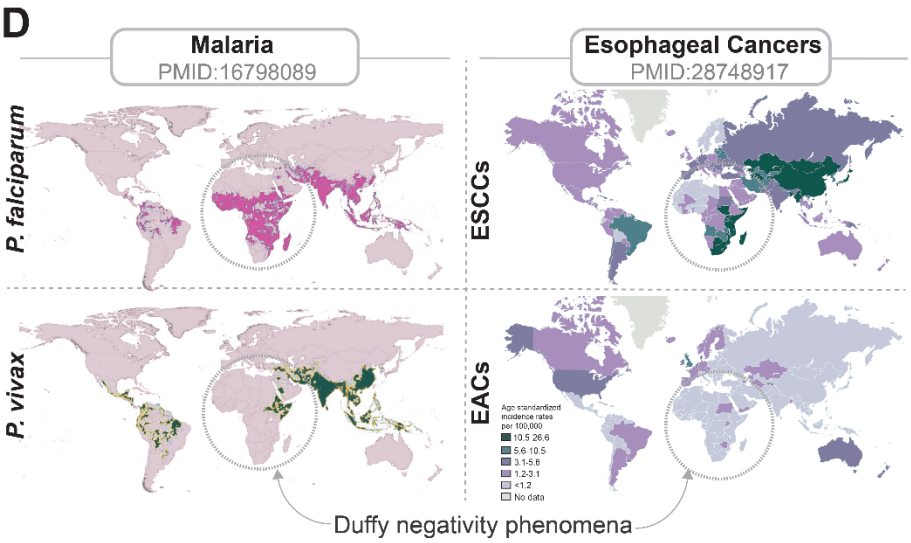


B

	rsID	Gene	Chr Position	NP Count	P Count	Binomial Test	Diff (P-NP)
Risk Alleles	**rs2814778** *	*DARC/ACKR1*	chr1:159204893:T:C	1	0	0.00000	-1
	rs246221	*ABCC1*	chr16:16044465:T:C	24	14	0.001397823	-10
	rs4148350	*ABCC1*	chr16:16076620:G:T	2	5	0.227479554	3
	rs299293	*HMMR*	chr5:163482208:C:T	16	9	0.012693315	-7
	rs299313	*HMMR*	chr5:163486945:G:A	16	11	0.10880832	-5
	rs299314	*HMMR*	chr5:163487994:T:C	16	11	0.10880832	-5
				75	**50**		**-25**
Protective Alleles	**rs445**	*CDK6*	chr7:92779056:C:T	7	11	0.213844802	4
	rs2380606 #	*SLCO5A1*	chr8:69828661:T:C	29	29	1	0
	rs35272691 #	*PSMD3-CSF3*	chr17:40001588:T:C	14	29	8.43E-07	15
	rs7882966 #	*CPXCR1*	chrX:88727152:T:C	1	6	0.024075999	5
				51	**75**		**24**

* *SNP causative of BEN in AA, Arab and Middle east population, Yemeni and Ethiopian Jews*
\# *SNP protective of neutropenia in Hispanic/Latino ethnicities*





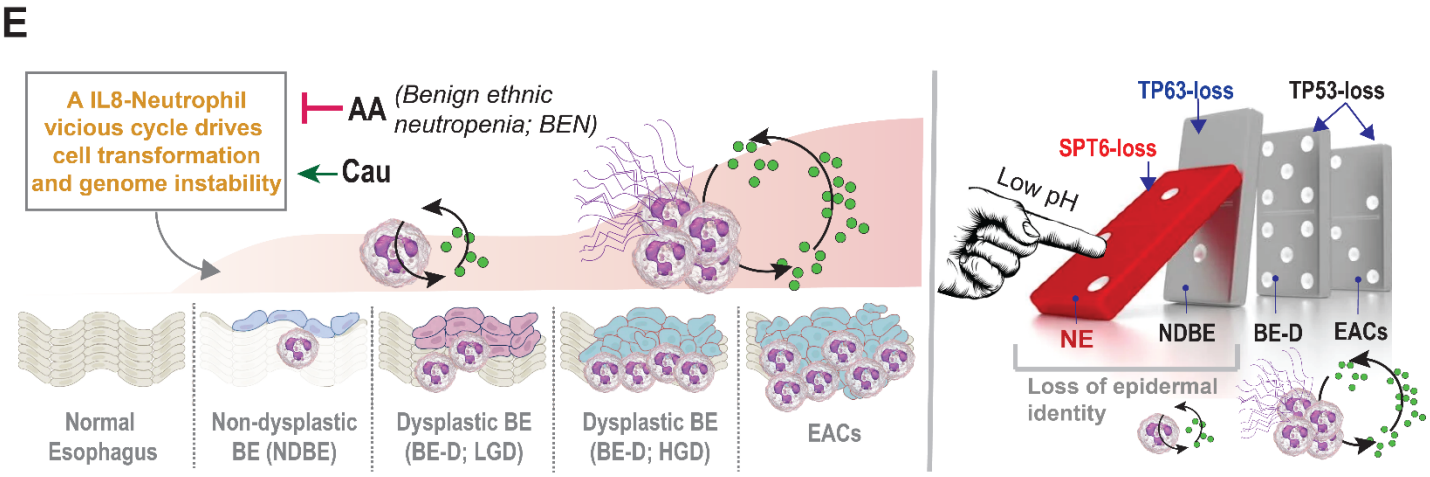


Figure 8. Ethnic neutropenia may reduce the risk of transformation in BE.

B. Table showing the frequency of neutropenia risk alleles (top) or protective alleles (bottom) among patients with BE (Seattle BE program; n = 80) who progress (Prog, P; n = 40) or not (Non-prog, NP; n = 40) to EAC. Statistical significance was determined for each SNP using Fisher's test with binomial probability distribution. See also **Supplementary Information 6**.

C. Global prevalence of the African *Duffy-null* polymorphism that causes benign ethnic neutropenia.

D. The prevalence of malaria (left) and the age-adjusted incidence of ECs (right) are displayed side by side. Black interrupted circles highlight how inability of *P. vivax* to attach to RBCs in *Duffy*-null zones (see C) protect Africans from that specific form of malaria (bottom-left) but not from *P. falciparum* (top-left). Circles on highlight similar differences in ESCC vs EAC in the African population.

E. Summary and working model: *Left:* A vicious IL8↔neutrophilic storm, perhaps biphasic, may be critical for driving the metaplasia-dysplasia cascade during NE→BE→EAC progression. Due to the AA *Duffy-null* polymorphism which manifests as benign ethnic neutropenia, some races/ethnicities (like AA and Hispanic/Latino) are protected from both the biphasic IL8/neutrophil-centric inflammatory storm *and* the metaplasia-dysplasia cascade. *Right:* As for epithelium intrinsic events, we identified a molecular mechanism that permits NE→BE transition. This mechanism is triggered by acid exposure (at least in part), suppresses the expression of SPT6, which in turn stalls TP63 function and expression, and a resultant loss in keratinocyte cell fate and gain in metaplastic features. These epithelium-intrinsic mechanisms are likely to be fuelled by the vicious IL8↔neutrophilic storm pinpointed by our computational models.